

Capfit³gen3



**Una caja de herramientas
para la conservación y promoción
del uso de la biodiversidad agrícola**



Funded by the
Horizon 2020
Framework
Programme of the
European Union



UNIVERSITY OF
BIRMINGHAM



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Capfit gen3

**Una caja de herramientas
para la conservación y promoción
del uso de la biodiversidad agrícola**

Autor de las herramientas

Mauricio Parra Quijano

Autores del documento

Mauricio Parra Quijano, José María Iriondo,
María Elena Torres, Francisco López,
Jade Phillips y Shelagh Kell



Funded by the
Horizon 2020
Framework
Programme of the
European Union



UNIVERSITY OF
BIRMINGHAM



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Catalogación en la publicación Universidad Nacional de Colombia

Parra Quijano, Mauricio, 1973-

CAPFITOGEN 3 : una caja de herramientas para la conservación y promoción del uso de la biodiversidad agrícola / autor de las herramientas, Mauricio Parra Quijano ; autores del documento, Mauricio Parra Quijano [y otros cinco] ; [autor de las figuras y fotografías, Mauricio Parra Quijano] ; [Ana María Díaz, traducción y edición]. -- Primera edición. -- Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias Agrarias, 2022. 1 recurso en línea (311 páginas) : ilustraciones (principalmente a color), diagramas, figuras, fotografías, mapas

Incluye referencias bibliográficas al final de cada capítulo

ISBN 978-958-505-034-1 (en línea)

1. Conservación de los recursos genéticos 2. Agricultura de conservación 3. Germoplasma 4. Agrobiodiversidad 5. Desarrollo de programas informáticos 6. Análisis espacial I. Iriondo, José María, 1963- II. Torres, María Elena, 1972- III. Francisco López, 1976- IV. Phillips, Jade, 1990- V. Kell, Shelagh, 1962- VI. Díaz, Ana María, traductor, editor VII. Título

CDD-23 333.953416 / 2022

CAPFITOGEN3: Una caja de herramientas para la conservación y promoción del uso de la biodiversidad agrícola

© Universidad Nacional de Colombia

Sede Bogotá - Facultad de Ciencias Agrarias

© Autores varios

2021

ISBN: 978-958-505-034-1

Dolly Montoya

Rectora

Universidad Nacional de Colombia

Autor de las herramientas:

Mauricio Parra Quijano

Profesor Asociado

Facultad de Ciencias Agrarias - sede Bogotá

Universidad Nacional de Colombia

Autores del documento:

Mauricio Parra Quijano

José María Iriondo

María Elena Torres

Francisco López

Jade Phillips

Shelagh Kell

Diseño y diagramación:

Laura Londoño M.

Autor de las figuras y fotografías:

Mauricio Parra Quijano

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales.

Todos los derechos reservados

Bogotá, D. C., Colombia

Tabla de Contenido

	Pag.
Agradecimientos	7
1. Antecedentes y diseño de herramientas	11
2. Herramientas CAPFITOGEN v3: Características y funcionamiento	19
3. Herramienta TesTable	41
4. Herramienta GEOQUAL	51
5. Herramienta SelecVar	63
6. Herramienta ELC mapas	77
7. Herramienta ECOGEO	93
8. Herramienta Representa	109
9. Herramienta DIV mapas	123
10. Herramienta ColNucleo	149
11. Herramienta FIGS_R	161
12. Herramienta rLayer	177
13. Herramienta Complementa	185
14. Herramienta Bfuture	205
15. Herramienta Modela	217
16. Herramienta Mcompare	263
17. Herramienta Tzones	273
18. Errores frecuentes	289
19. Anexos	297





Agradecimientos

Agradecimientos

El desarrollo de CAPFITOGEN3 fue apoyado por el proyecto “Networking, Partnerships and Tools to Enhance *in situ* Conservation of European Plant Genetic Resources” (short name Farmer’s Pride), financiado por el Programa Marco Horizonte 2020 de la Unión Europea.

Las herramientas CAPFITOGEN y su evolución, CAPFITOGEN3, acumulan un trabajo continuo desde 2012, cuando las primeras dos herramientas se pensaron y diseñaron. Obviamente no salieron de una noche a la mañana siguiente, sino que provienen de desarrollos hechos desde 2005 cuando se obtuvo el primer mapa ELC y posteriormente se fueron desarrollando otras aplicaciones ecogeográficas útiles para la conservación y uso de los recursos fitogenéticos para la alimentación y la agricultura (RFAA). Desde 2012 han sido muchos los aciertos y errores, los obstáculos y dificultades, pero también las personas que aparecieron en el momento indicado para que el sueño de CAPFITOGEN se materializara. Hago referencia a un sueño, porque estas herramientas fueron literalmente eso, un sueño que tuve cuando terminé mi tesis doctoral y pensé que algunos adelantos metodológicos no podían quedar sólo en manos de algunos muy pocos investigadores del futuro que citaran mis publicaciones. Asumí, a partir de ese sueño, la premisa que el esfuerzo de hacer ciencia sólo se compensa cuando el avance llega en realidad a la gente para mejorar o facilitar su vida. En este caso CAPFITOGEN pudo llegar a muchos técnicos e investigadores que en su rutina hacen conservación y uso de la agrobiodiversidad, y logró en muchos casos facilitar y hasta permitir que pudieran hacer muchos análisis y tareas que antes no hubiera sido posible.

Mis primeros agradecimientos a Fernando Latorre (España), quien representó en su día a España ante el TIRFAA, por su decidido apoyo al Programa CAPFITOGEN y el desarrollo de las herramientas. Agradecimientos también por su apoyo y comentarios a otros investigadores del CRF-INIA, en particular Lucía de la Rosa y Rosa García.

Gracias a los “fans” de CAPFITOGEN. Los fans son investigadores y personas que trabajan por la conservación de la agrobiodiversidad que no sólo les gustó la idea o que usaron en algún momento alguna de las aplicaciones CAPFI-

TOGEN, sino que han sido difusores de su potencial entre la comunidad RFAA y muchas veces se han convertido en buscadores de oportunidades para mayor difusión de la tecnología o financiación de mejoras. A las personas que me acompañan en la autoría de este libro se les debe en gran parte lo que son hoy las herramientas. Txema, Elena, Francisco, Shelagh y Nigel son piezas clave sin las cuales el rompecabezas hubiera sido imposible de completar. También otros investigadores y recurrentes usuarios se han hecho fans de CAPFITOGEN como César Tapia, Hannes Gaisberger, Joana Brehm, Lorena Marinoni, Sandibel Vera, Rosalinda González, Jade Phillips, Heli Fitzgerald, Riina Jalonen y muchos más.

Técnicamente, muchas gracias a todos los que contribuyen en R, particularmente aquellos que han desarrollado paquetes que luego he podido integrar en las herramientas. Resaltar el trabajo de Robert J. Hijmans y otros generosos desarrolladores de paquetes de R para análisis espaciales. Robert también ha permitido el uso de datos worldclim para poder ser distribuidos de manera facilitada a los usuarios CAPFITOGEN. De algunas personas, las cuales admiro por su capacidad de generar nuevos métodos y aplicaciones, me he robado más de una idea que he transformado en herramienta. Entre ellas, de nuevo Robert Hijmans, Marteen van Zonneveld, Evert Thomas, Stephanie Greene, o Michael Mackay.

Quisiera agradecer a Nigel Maxted por su guía y consejos en la producción de esta nueva versión de las herramientas. Gracias a Ana María Díaz por su apoyo en la traducción y edición del manual de usuario de CAPFITOGEN3 (version en Inglés), y Laura Londoño por la hermosa diagramación y/o maquetación.

Finalmente debo agradecer a los más de 300 estudiantes, tesisistas, técnicos curadores y directivas de múltiples centros de educación e investigación donde se hace conservación y/o uso de recursos fitogenéticos por apostar por CAPFITOGEN, al asistir a los talleres de entrenamiento. Sus experiencias como usuarios aportaron valiosos y sugerencias comentarios para la continua mejora de las herramientas.



Taller Regional, Binzhou, China, Octubre 2018.



1 | Antecedentes y Diseño de Herramientas

1.1 ¿Qué son las herramientas CAPFITOGEN3?

El futuro de la agricultura descansa en gran medida en la conservación eficiente y la utilización sostenible de los recursos fitogenéticos para la alimentación y la agricultura (RFAA), lo cual requiere el desarrollo de estrategias con relación costo beneficio óptima para enfrentar los impactos ambientales del cambio climático y otros retos futuros. La caja de herramientas CAPFITOGEN fue desarrollada para apoyar la comunidad global que conserva y usa sosteniblemente los RFAA, ofreciendo herramientas de software diseñadas para realizar análisis espaciales y de diversidad ecogeográfica para facilitar una conservación más eficiente y efectiva y la planificación del uso sostenible. Las herramientas CAPFITOGEN han ofrecido apoyo a técnicos y científicos que trabajan en RFAA alrededor del mundo durante los últimos nueve años, alcanzando cifras interesantes como dieciséis talleres de entrenamiento en cuatro continentes, más de 400 personas capacitadas y su uso ha sido citado en al menos 71 publicaciones en el periodo de 2014 a 2021. CAPFITOGEN3 es una nueva iteración de la caja de herramientas CAPFITOGEN y está compuesta de 15 aplicaciones que pueden ser usadas directamente en un servidor accesible a través de un portal *on line* o descargable y utilizable en una computadora local.

1.2. Historia del desarrollo de las herramientas CAPFITOGEN

Bajo el auspicio del Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura (TIRFAA) y la Agencia Española de Cooperación Internacional para el Desarrollo (AECID), entre los años 2008 y 2010 se desarrollaron dos talleres para la implementación del TIRFAA en el grupo de los países de América Latina y el Caribe (GRULAC) en Cartagena de Indias (Colombia, julio-agosto de 2008) y La Antigua (Guatemala, agosto de 2010). Estas actividades indicaron lo importante que podría ser contribuir a la implementación de los objetivos del TIRFAA dentro de la comunidad de países GRULAC a través de talleres de formación y transferencia de conocimiento. Adicionalmente el éxito de estas actividades son un reflejo de la estrecha relación y cooperación entre Programas Nacionales de países GRULAC y España.

Tomando como base estos antecedentes y las necesidades de la región, fue implementado el Programa para el Fortalecimiento de las Capacidades en Programas Nacionales de Recursos Fitogenéticos de América Latina. Este programa se enfocó en el desarrollo de tecnología apropiada para países con abundante agrobiodiversidad y limitados recursos económicos y su efectiva transferencia. Su función fue la de desarrollar la tecnología, transferirla y proveer formación adecuada al personal técnico de los países de América Latina que son parte del Tratado. La tecnología desarrollada fue inicialmente pensada para apoyar y facilitar labores relacionadas con la conservación *ex situ* de los recursos fitogenéticos, y no crear nuevas obligaciones o deberes adicionales al personal encargado de la recolección, conservación, caracterización o evaluación y promoción de la utilización. Así mismo la tecnología que se diseñó debería ser sencilla y entendible por parte del personal técnico, evitando procesos complejos o de aplicación no práctica en los procesos rutinarios asociados a la conservación *ex situ*. Posteriormente, y dada la necesidad expresada por los técnicos latinoamericanos durante los primeros talleres de transferencia de la tecnología desarrollada, se optó por el diseño de herramientas adicionales cuyo objetivo es facilitar o hacer más eficientes procesos de conservación *in situ*.

En el periodo 2012-2018 se llegaron a realizar 15 talleres en 14 países diferentes, ocho de ellos en países latinoamericanos, tres en Europa, uno en África y uno en Asia. Fueron capacitados alrededor de 320 técnicos de cuatro continentes a través de cuatro talleres regionales y 11 talleres nacionales. De esta forma el Programa para el Fortalecimiento de las Capacidades en Programas Nacionales de Recursos Fitogenéticos de América Latina se convirtió en un programa no para GRULAC sino una iniciativa de alcance global. El programa en el tiempo se convirtió en un modelo de transferencia tecnológica para la comunidad de técnicos y científicos de los recursos fitogenéticos para la alimentación y la agricultura, con una estrategia de diseño de herramientas donde el usuario o beneficiario comprende la utilidad de la aplicación que se le transfiere y realiza el proceso por su propia cuenta en base a las necesidades reales y actuales de su escenario de trabajo. En este modelo se garantiza a los beneficiarios el acceso directo a los científicos y desarrolladores tecnológicos para que éstos sirvan de acompañantes y ayuden en la resolución de sus dudas o dificultades de aplicación. A la vez, los científicos se nutren directamente de las experiencias y problemáticas de los técnicos que directamente manejan la agrobiodiversidad, lo cual incide en futuros desarrollos tecnológicos aún más cercanos a las necesidades reales en los programas nacionales.

1.2.1 Evolución de las herramientas: CAPFITOGEN v3 modo local y en servidor

A finales de 2019 se inicia un proceso de repotenciación de las herramientas y se decide la apertura de una nueva modalidad de uso, a través de su instalación en un servidor y un portal web, adicionalmente a la versión de instalación local y uso *off line*. Se trata de una revisión a fondo de todos los procedimientos, actualización de funciones y scripts R, lo cual da como resultado la tercera versión de las herramientas CAPFITOGEN. Este proceso fue posible gracias al decidido apoyo del proyecto “Networking, Partnerships and Tools to Enhance in situ Conservation of European Plant Genetic Resources” (Farmer’s Pride en corto), financiado por el Programa Marco Horizonte 2020 de la Unión Europea, y la interacción de la Universidad de Birmingham (UK) y la Universidad Nacional de Colombia.

1.2.2 Estrategia básica del diseño de las herramientas CAPFITOGEN

Los métodos para recolectar, conservar y caracterizar la agrobiodiversidad con estándares científicos han venido usualmente desde regiones y centros donde los recursos económicos, la infraestructura o la capacitación del personal no representan mayores limitaciones. Esto ha generado que en países en desarrollo estas metodologías o no se puedan aplicar, o si se aplican, no se puedan extender a todo el germoplasma conservado. Una situación que contrasta con el hecho de que es en los países en desarrollo donde se concentra la mayor riqueza genética agrícola vegetal.

Este escenario ha conducido a que algunos grupos de investigación en el mundo hayan dirigido sus esfuerzos a explorar metodologías de menor coste y complejidad, que se adapten mucho mejor a las condiciones de Programas Nacionales para la conservación y uso de la agrobiodiversidad de países en desarrollo. Las alternativas metodológicas incluyen el aprovechamiento de la información ambiental de los sitios de colecta (ecogeográfica), por ejemplo, para hacer estimaciones de la variabilidad genética que contendría el germoplasma o estimar con mayor éxito la probabilidad de encontrar genes de interés. Así mismo, se plantea el uso de los sistemas de información geográfica como medio de obtención y aprovechamiento de dichos datos ecogeográficos. Dado que la mayor parte de la información

ecogeográfica y los programas informáticos para realizar los análisis son de acceso gratuito, la inversión se reduce a un ordenador/computador con una configuración de tipo comercial y capacitación del personal. Se trata entonces de tecnología compatibles con escenarios de recursos limitados, condición que ocurre a menudo en Programas Nacionales de países en desarrollo.

Una vez aceptada la premisa de que la tecnología simple pero eficiente tiene un mayor impacto y es más adoptada por los técnicos que hacen el trabajo cotidiano de conservación y uso de los Recursos Fitogenéticos para la Alimentación y la Agricultura (RFAA), se consideró que la entrega de este tipo de tecnología debía ser también en forma de herramientas sencillas. La sencillez tendría que llegar al punto que los futuros usuarios deberían ser capaces de formarse en el uso de las herramientas durante un taller corto de entre dos y cuatro días de duración.

1.3. Contenido de la caja de herramientas CAPFITOGEN3

Las herramientas a través de las cuales se presentó la tecnología apropiada a la comunidad RFAA se denominaron CAPFITOGEN desde 2012. Su diseño tuvo las siguientes premisas:

- Su base debe ser un avance científico publicado en revistas internacionales. La herramienta puede abarcar total o parcialmente la metodología de un artículo, preferiblemente revisado por pares.
- La metodología se selecciona por su aplicabilidad en procesos que actualmente se realizan en programas nacionales, bancos de germoplasma y proyectos de conservación *in situ*. El método incluido en una herramienta debe ser lo suficientemente simple como para que la mayor parte de técnicos en RFAA puedan comprenderlo y lo encuentren válido a la hora de su aplicación en labores rutinarias.
- Finalmente, el método seleccionado debe ser implementado en el software R (R Core Team, 2020). Si bien algunas metodologías tal y como han sido publicadas no tienen un desarrollo en R, si esto fuera posible, son métodos susceptibles de convertirse en herramientas CAPFITOGEN.
- Las herramientas CAPFITOGEN ya plenamente desarrolladas en forma de *script* (líneas de programación) de R, debe presentarse de una manera sencilla o facilitada al usuario, de tal manera que el mismo no tenga que aprender a usar R ni su lenguaje de programación para hacer uso de ellas.
- Finalmente, las herramientas CAPFITOGEN deben ser diseñadas de tal modo que su entrega y funcionamiento no implique ningún cargo económico a los potenciales usuarios.

Las versiones de herramientas CAPFITOGEN 1 y 2 se entregaron dentro de una interfaz amigable construida en html y java script, con el uso de un servidor virtual. Para ello, el usuario debía descargar un instalador de las herramientas, que realizaba para el usuario la instalación de R y java, y posteriormente abría el navegador de internet, a través del cual se accedía a los formularios y se podía introducir los parámetros que configuran la ejecución de los procesos (Parra Quijano, *et al.*, 2014).

Para la nueva versión de las herramientas CAPFITOGEN, se realizó una revisión completa de todos los scripts de R, actualizándose muchas funciones y procesos, adecuándolos para las dos nuevas formas de utilización:

1. A través de la descarga del conjunto de scripts R de herramientas y de lista de parámetros desde , que se visualizan, configuran y ejecutan en el software RStudio Desktop (<https://rstudio.com/products/rstudio/download/>),

distribuido bajo licencia y términos estipulados por la versión 3 de GNU Affero General Public License (<http://www.gnu.org/licenses/agpl-3.0.txt>).

- Mediante CAPFITOGEN on server, la versión 3 de las herramientas dispuesta en servidor, donde el usuario sólo debe entrar en un sitio web, registrarse como usuario y posteriormente acceder a los formularios de las herramientas, subir las tablas e información necesaria para cada proceso y solicitar la ejecución del script de R en el servidor. El acceso a CAPFITOGEN on server se hace a través del vínculo <http://onservercapfitogen.net/>.

La lista de las herramientas hasta ahora desarrolladas es:

TesTable: Herramienta auxiliar que comprueba el formato de tablas que serán usadas en otras herramientas. Capítulo 3.
GEOQUAL: Realiza una evaluación de la calidad de la georreferenciación de los datos de pasaporte útil para el descarte de presencias o sitios de recolección con deficiencias en la asignación de coordenadas. Capítulo 4.

SeleccVar: Hace posible la selección de variables ecogeográficas de importancia en la formación de grupos de adaptación, aspecto de mucha utilidad para la utilización de herramientas como ELC mapas o modela. Capítulo 5.

ELC mapas: Permite obtener mapas de caracterización ecogeográfica del terreno (ELC) que reflejan escenarios adaptativos para una especie y país o región determinada. Capítulo 6.

Representa: Permite hacer un análisis de representatividad ecogeográfica de una especie dentro de una colección de germoplasma, detectando sesgos o cualquier vacío o faltante geográfico y ecogeográfico en la colección objetivo. Capítulo 7.

ECOGEO: Permite hacer una caracterización ecogeográfica de los sitios de recolección del germoplasma. El usuario tiene más de 170 variables ecogeográficas (bioclimáticas, edáficas y geofísicas) disponibles. Capítulo 8.

DIV mapas: Permite la obtención de mapas que mostrarán zonas de alta variabilidad ecogeográfica, fenotípica y/o genotípica, basándose en la determinación de distancias ecogeográficas, fenotípicas o genotípicas en vecindades. Capítulo 9.

ColNucleo: Establece sub-colecciones de tipo nuclear/núcleo basadas en mapas de caracterización ecogeográfica del terreno (mapas ELC) a partir de estrategias de afijación proporcionales. Estas sub-colecciones nucleares-núcleo son representativas de la colección original desde un punto de vista ecogeográfico. Capítulo 10.

FIGS_R: Aplica filtros para seleccionar el germoplasma más adecuado con relación a estreses abióticos de interés para mejoradores de cultivos y así generar sub-colecciones de germoplasma de tipo FIGS. Capítulo 11.

rLayer: Esta herramienta permite obtener conjuntos de capas ecogeográficas que no corresponden a los límites de un país sino a la extensión de la distribución de las recolecciones/presencias del usuario, las cuales quedan automáticamente disponibles para las demás aplicaciones. Capítulo 12.

Complementa: Hace posible análisis de complementariedad entre celdas o áreas protegidas y además determina el grado de cobertura de redes de áreas actuales en términos de conservación in situ de la agrobiodiversidad. Capítulo 13.

Bfuture: Permite descargar desde internet y adaptar información bioclimática proyectada a futuro (cambio climático) útil para obtener distribuciones potenciales de especies en escenarios de futuro en herramienta Modela. Capítulo 14.

Modela: Permite la obtención de modelos de distribución de especies de forma facilitada, sólo con la introducción por parte del usuario de los datos de presencia. Capítulo 15.

Mcompare: Herramienta complementaria para comparar proyecciones de distribución de especies en escenarios de presente y futuro, determinando cuatro posibles situaciones relativas a su conservación. Capítulo 16.

Criteria: Facilita la toma de decisiones sobre donde enfocar esfuerzos de conservación in situ a través de análisis multicriterio, donde se puede combinar información de índole agronómica, ecológica, socio-cultural, económica, etc. Herramienta en desarrollo. Capítulo 17.

1.4. Variables ecogeográficas

Buena parte de las herramientas CAPFITOGEN requieren dos insumos básicos para su funcionamiento: la información que aporta el usuario y un set de variables ambientales en forma de capas SIG, necesario para realizar los análisis ecogeográficos. En versiones anteriores se llegó a consolidar un conjunto de 103 variables, incluyendo 67 variables bioclimáticas, todas ellas provenientes de la base de datos Worldclim v 1.4 (<https://www.worldclim.org/data/v1.4/worldclim14.html>), 31 variables edáficas provenientes de la base de datos armonizada mundial HWSD v 1.21 (<https://iiasa.ac.at/web/home/research/researchPrograms/water/HWSD.html>) y cinco variables geofísicas, todas ellas derivadas del DEM de SRTM v 3 (<http://srtm.csi.cgiar.org/>). Para CAPFITOGEN v3 se realizó también una actualización de las capas ecogeográficas, actualizando a Worldclim v 2.1 las capas bioclimáticas, incluyendo una nueva variable mensual (presión de vapor), añadiendo 35 nuevas variables edáficas provenientes de datos originales de SoilGrids (<https://soilgrids.org>) ensamblados para ser comparables con las variables HWSD (top y sub soil) y finalmente 26 nuevas variables geofísicas (radiación solar y velocidad de viento, mensuales y promedio anual), provenientes de Worldclim v2.1 (<https://www.worldclim.org/data/worldclim21.html#>).



Taller Nacional, Sancti Spiritus, Cuba, Septiembre 2013.



2 | **Herramientas CAPFITOGEN v3: características y funcionamiento**

2.1. Modo local

En la línea de la primera y segunda versión de las herramientas, donde el usuario se descarga un instalador, el cual permite la utilización local (en un PC de escritorio o en un laptop) sin necesidad de estar conectado a internet, para CAPFITOGEN v3 se decide mantener esta vía de utilización local y off line. Sin embargo, el formato de interfaz amigable diseñado para ser visualizado en un navegador de internet conectado a un servidor virtual, creado en html y java para las versiones 1 y 2, originó problemas en algunos usuarios y demandó la creación de algunos parches y soluciones provisionales.

Por esta razón para CAPFITOGEN v3 modo local se opta por enfrentar al usuario a los scripts de R y a proporcionar la lista de parámetros requeridos por cada herramienta, a través de un script adicional de R. A pesar de esto, el usuario no necesita conocer el software R ni su programación para poder ejecutar los scripts. Y para facilitar ese contacto con los scripts R de cada herramienta y particularmente los scripts de parámetros, CAPFITOGEN v3 modo local se apoya en un software asociado a R denominado RStudio (<http://rstudio.com/>). Por tanto, para que un usuario utilice CAPFITOGEN v3 se requiere que primero descargue desde la página web de CAPFITOGEN (<http://capfitogen.net>) los scripts R de las herramientas y de los parámetros de herramientas y la estructura de carpetas requerida. Posteriormente debe descargar e instalar R 3.6.3 (<https://cran.r-project.org/bin/windows/base/old/3.6.3/>) y RStudio desktop (en <https://rstudio.com/products/rstudio/download/#download> o directamente desde <https://download1.rstudio.org/desktop/windows/RStudio-1.3.1073.exe>).

2.1.1 Cómo instalar CAPFITOGEN v3 modo local

Los siguientes pasos son necesarios para la instalación de CAPFITOGEN v3 modo local:

- a) Descargar el software R para Windows. Si el usuario desea realizar análisis con Modela (modelos de distribución basados en el paquete biomod2), debe descargar la versión R 3.1.2 desde <https://cran.r-project.org/bin/windows/base/old/3.1.2/>. Si el usuario no piensa utilizar esta herramienta, para mayor eficiencia puede descargar una versión más reciente de R como la 3.6.3 desde <https://cran.r-project.org/bin/windows/base/old/3.6.3/>. Si bien las herramientas CAPFITOGEN v3 modo local no se ha probado en macOS, hay versión R para este sistema operativo, y en principio todos o casi todos los scripts CAPFITOGEN v3 serían funcionales en estas versiones de R. Lo mismo podría decirse para el sistema operativo Linux.
- b) En el proceso de instalación de R, no olvidar apuntar la ruta de instalación, ya que en el próximo paso será necesario acceder a la carpeta donde se realiza la instalación de este software.
- c) Descargar las librerías de R necesarias para que funcionen las herramientas. Si se descargó la versión R 3.1.2, debe descargar las librerías desde: <http://t.ly/GOOF>. Si descargó la versión R 3.6.3, las librerías se descargan desde: <http://t.ly/tqu1>. Se descargará un archivo comprimido (.zip), el cual debe ser descomprimido. El resultado será una carpeta llamada “library”.
- d) Reemplazar la carpeta “library” original de la instalación de R (que se encontrará en la ruta y carpeta donde se instaló R en su computadora), por la carpeta “library” producto de la descompresión del archivo .zip descargado en el paso anterior. Se recomienda primero eliminar la carpeta “library” de la instalación original, y luego copiar o mover la carpeta “library” fruto de la descompresión.

- e) Descargar e instalar el software RStudio Desktop (versión gratuita) desde <https://rstudio.com/products/rstudio/download/#download>. Como referencia, descargar la versión 1.3.1073 o anterior (<http://t.ly/IKTV>).
- f) Descargar el cuerpo principal de carpetas y archivos necesarios para el funcionamiento correcto de CAPFITOGEN v3 modo local desde <http://t.ly/F9eq>. Obtendrá un archivo denominado “CAPFITOGEN3.zip” de alrededor de 1.9 Gb. Al descomprimir este archivo obtendrá una carpeta denominada “CAPFITOGEN3” con todos los archivos y carpetas auxiliares necesarias para usar las herramientas. Puede ubicar esta carpeta en cualquier sitio de su disco duro, sin embargo se recomienda que la ruta que lleve a ella no contenga espacios (por ejemplo, evite rutas como “C:/Mis documentos/CAPFITOGEN3”).
- g) Es el momento de descargar las capas de información ecogeográfica necesarias para el análisis o estudio, y que la mayoría scripts de herramientas va a necesitar. Entre en este link que lleva a una carpeta compartida en Google Drive: <http://t.ly/89eu>. La estructura de carpetas y archivos .tif de capas SIG de información ecogeográfica es la siguiente:

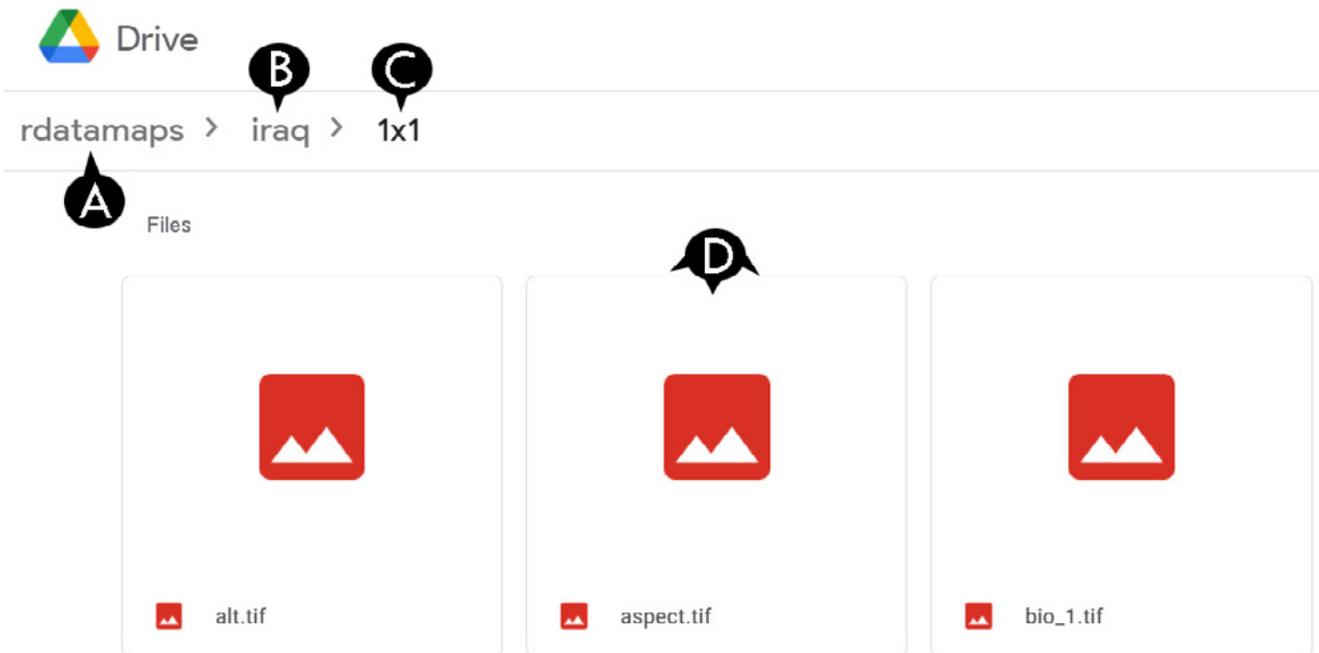


Figura 1. Niveles en los que se encuentra las capas GIS. A) Nivel superior, B) nivel país, C) nivel resolución y D) 177 capas. Google Drive permite descargar la información en cualquiera de estos niveles.

En la estructura de carpetas y archivos CAPFITOGEN v3 (descargada en el apartado f), la ruta que debe llevar a las variables ecogeográfica descargadas, es similar a la de la carpeta compartida en Google Drive. Las capas descargadas deben localizarse en la carpeta rdatamaps (que está dentro de la carpeta CAPFITOGEN3), dentro de ella en la carpeta del país correspondiente (con el nombre tal y como aparece en Google Drive), y dentro de ella en una carpeta que indica la resolución de celdas (opciones: 1x1, 5x5, 10x10 o 20x20). Por ejemplo, si la carpeta principal (CAPFITOGEN3) está localizada en la raíz del disco C, y se quiere hacer un análisis para Argentina con capas con celdas de tamaño 5x5 km, y sólo se va a usar la capa de altitud, entonces la ruta debe ser: **C:/CAPFITOGEN3/radatamaps/argentina/5x5/alt.tif**

- h) Descargar los scripts de las herramientas CAPFITOGEN v3 desde <http://t.ly/h7FQ>. Se descarga el archivo “Scripts.zip”, que al descomprimirlo genera dos carpetas: “Parameters” y “Tools”. En la carpeta “Parameters” se encuentran los scripts que permite ingresar los parámetros en Rstudio de cada herramienta, mientras que en la carpeta “Tools” están los scripts de las herramientas, los que permiten obtener los resultados. La localización de estas carpetas y de los scripts en el disco duro es totalmente irrelevante.

2.1.2 Cómo utilizar CAPFITOGEN v3 modo local

Una vez el usuario tenga claro la funcionalidad de cada herramienta (descritas en detalle en los capítulos siguientes) y haya decidido que herramienta va a usar, debe buscar el script de parámetros de esa herramienta y hacer doble click sobre el archivo “Parametros_NombreHerramienta_2020.R”. Se debe abrir ese script en el programa RStudio. Si no lo hace, debe configurar el sistema para que RStudio este asociado a la extensión .R y pueda abrir estos archivos. Si a pesar de configurar el sistema, por alguna razón no se abre Rstudio al hacer doble click sobre archivos .R, entonces se debe abrir primero Rstudio y luego hacer click sobre la pestaña “File”, luego seleccionar “Open File” y busca los archivos .R que contienen los scripts y la lista de parámetros.

Una vez se haya logrado abrir RStudio, se apreciará la siguiente distribución:

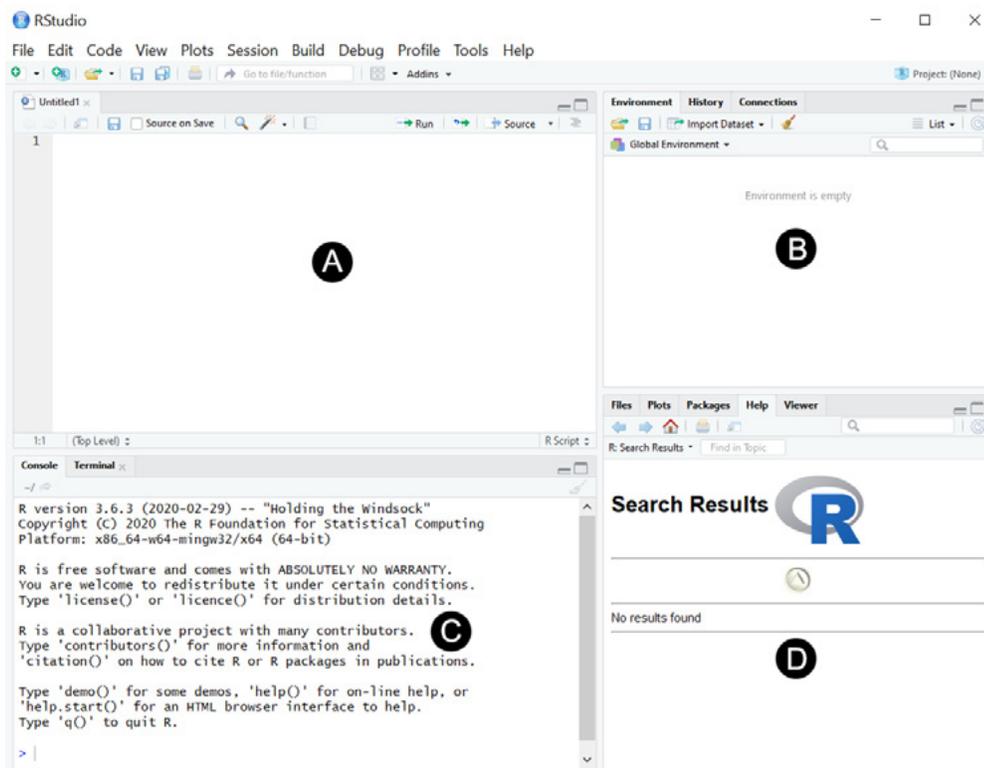


Figura 2. Ventanas del software RStudio desktop v 1.3.1073. A) Ventana de escritura o edición. B) Ventana de objetos producidos. C) Ventana de avance del proceso. D) Ventana de ayuda sobre los paquetes de R y sus funciones y visualización de gráficos.

Cuando se abre el script de parámetros de una herramienta, como por ejemplo el de Complementa, en la ventana de escritura/edición se puede observar:



Figura 3. Apertura de un script de parámetros en la ventana de escritura/edición. A) Pestaña con el nombre del script. B) Título del script. C) Nombre del primer elemento (parámetro). D) La parte a la izquierda de los símbolos <= (que significa =) es donde el usuario configura el parámetro de acuerdo a sus condiciones o necesidades. E) La primera línea de script debajo de la línea del parámetro indica el tipo de parámetro (texto, lógico o numérico). Esta, como otras líneas que inician con un # o varios ### significa que la línea sólo es informativa. F) Las líneas “# Nota:” (con uno o varios #), le dan pistas y ayudan al usuario a configurar correctamente el parámetro. G) Algunos scripts tienen secciones, que se titulan con este tipo de líneas.

Los tipos de parámetros, de texto, numéricos o lógicos, se visualizan en la Figura 4.



Figura 4. Tipos de parámetros. A) subtítulo. B) Parámetro tipo numérico “porcol”, donde el número (en este caso 10) se muestra en color azul. C) Línea informativa que indica que se trata de un parámetro numérico. D) Parámetro tipo texto “estratcol”, donde el texto (en este caso “P”) se muestra en color verde. E) Otras dos opciones permitidas para el parámetro “estratcol”: “C” o “L”. Si se desea usar una de ellas, primer se debe anular la primera (“P”) poniendo al inicio de la línea un símbolo # (#estratcol<—“P”, lo cual torna toda la línea al color verde) y activar la línea/opción deseada eliminando el símbolo # del inicio (al activar la línea el color del nombre del parámetro se torna negro). Tener cuidado con dejar más de una línea activada para un parámetro, ya que sólo se tendrá en cuenta la última línea activada. F) Línea informativa que indica que se trata de un parámetro de texto. G) Parámetro tipo lógico (sólo tiene dos respuestas válidas, TRUE o FALSE), donde la opción, como en el caso del numérico, también se muestra en azul. H) Línea informativa que indica que se trata de un parámetro lógico.

En algunos casos, frente a un parámetro se indica si este debe ser cumplimentado o ajustado por parte del usuario o no, dependiendo de la configuración de un parámetro anterior. En la Figura 5 se muestra un fragmento del script de la herramienta rLayer, donde varios parámetros se deben configurar de acuerdo a lo que se haya especificado en el parámetro “cropway”. En algunos casos un parámetro “aplica” (debe ser configurado) dependiendo lo que se haya indicado en un parámetro o en varios parámetros previos. En caso de que un parámetro no deba ser ajustado por el usuario ya que no se cumple una condición previa en otro parámetro, el usuario no debe hacer nada, ni eliminarlo ni modificarlo.

El usuario debe tener también ciertas precauciones a la hora de indicar la ruta que lleva a ciertos elementos dentro de la estructura de su disco duro. Cada parte de una ruta se debe separar con el símbolo slash (/) y no backslash (\), símbolo que usualmente aparece en rutas en Windows. Así mismo, si el usuario indica una ruta, debe estar seguro que el elemento o la carpeta que le está indicando a R, realmente existe. Por ejemplo, si el usuario quiere que sus resultados se guarden en la ruta C:/CAPFITOGEN3/Resultados, primero debe asegurarse que la carpeta Resultados ya está creada dentro de la carpeta CAPFITOGEN3 en su disco C.

```

14 cropway<-"polygon"
15 #cropway<-"square"
16 #cropway<-"buffer"
17 #Parámetro de tipo texto (un texto que va entre comillas "")
18 ##### Nota1: Este parámetro le indica a rlayer como se va a proceder a cortar las capas mundiales (world) para ade
19 ##### Nota2: Si cropway="polygon", se utilizará un archivo vectorial tipo shapefile que proporcionará el usuario q
20 ##### Nota3: Si cropway="square", se utilizarán los sitios de recolección (coordenadas) incluidos en la tabla de p.
21 ##### Nota4: Si cropway="buffer", se utilizarán los sitios de recolección (coordenadas) incluidos en la tabla de p.
22
23 buffer<-30 #Sólo aplica si cropway="buffer"
24 #Parámetro tipo número (el cual en Rstudio aparecerá con color azul)
25 ##### Nota: Parámetro numérico que expresa kilómetros (km) y con el que se indica el radio que se usará para gener.
26
27 shapefile<-"albania" #Sólo aplica si cropway="polygon"
28 #Parámetro de tipo texto (un texto que va entre comillas)
29 ##### Nota1: En este parámetro se debe indicar el nombre del shapefile, que debe estar en el sistema de coord lat-lon.
30 ##### Nota2: El shapefile (los 4-7 archivos que lo componen, entre los que debe estar los imprescindibles .shp, .dl
31
32 pasaporte<-"Pasaporte/PasaporteOriginalEvaluadoGEOQUAL.txt" #Sólo aplica si cropway="square" o cropway="buffer"
33 #Parámetro de tipo texto (un texto que va entre comillas "")
34 # Nota1: este archivo de texto debe estar en la carpeta Pasaporte, que a su vez es una carpeta que está dentro de
35 # Nota2: esta tabla es de igual que estructura a otras tablas de pasaporte. Puede haber pasado ya por la herramie

```

Figura 5. Parámetros que dependen de otros usando como ejemplo el script de parámetros de rLayer. A) Parámetro llamado “cropway” de referencia para otros. B) Parámetro “buffer” que debe ser ajustado sí previamente “cropway” se ha configurado como “buffer”. C) Parámetro llamado “shapefile”, en el cual se debe indicar el nombre de un archivo tipo shapefile (con extensiones .shp, .shx, .dbf, entre otras) que debe proveer el usuario y que sólo aplica si “cropway” está configurado como “polygon”. D) Parámetro “pasaporte”, el cual es necesario ajustar de acuerdo a las condiciones del usuario sólo si “cropway” es igual a “square” o “buffer”.

Sólo ajuste los parámetros que aplican al caso o utilización que quiere hacer de la herramienta, los demás los puede dejar tal y como aparecen en el script. Una vez el usuario ha terminado de ajustar el script de parámetros de acuerdo a sus condiciones y datos disponibles, debe seleccionar todas las líneas del script y ejecutarlo, como se muestra en la Fig. 6.

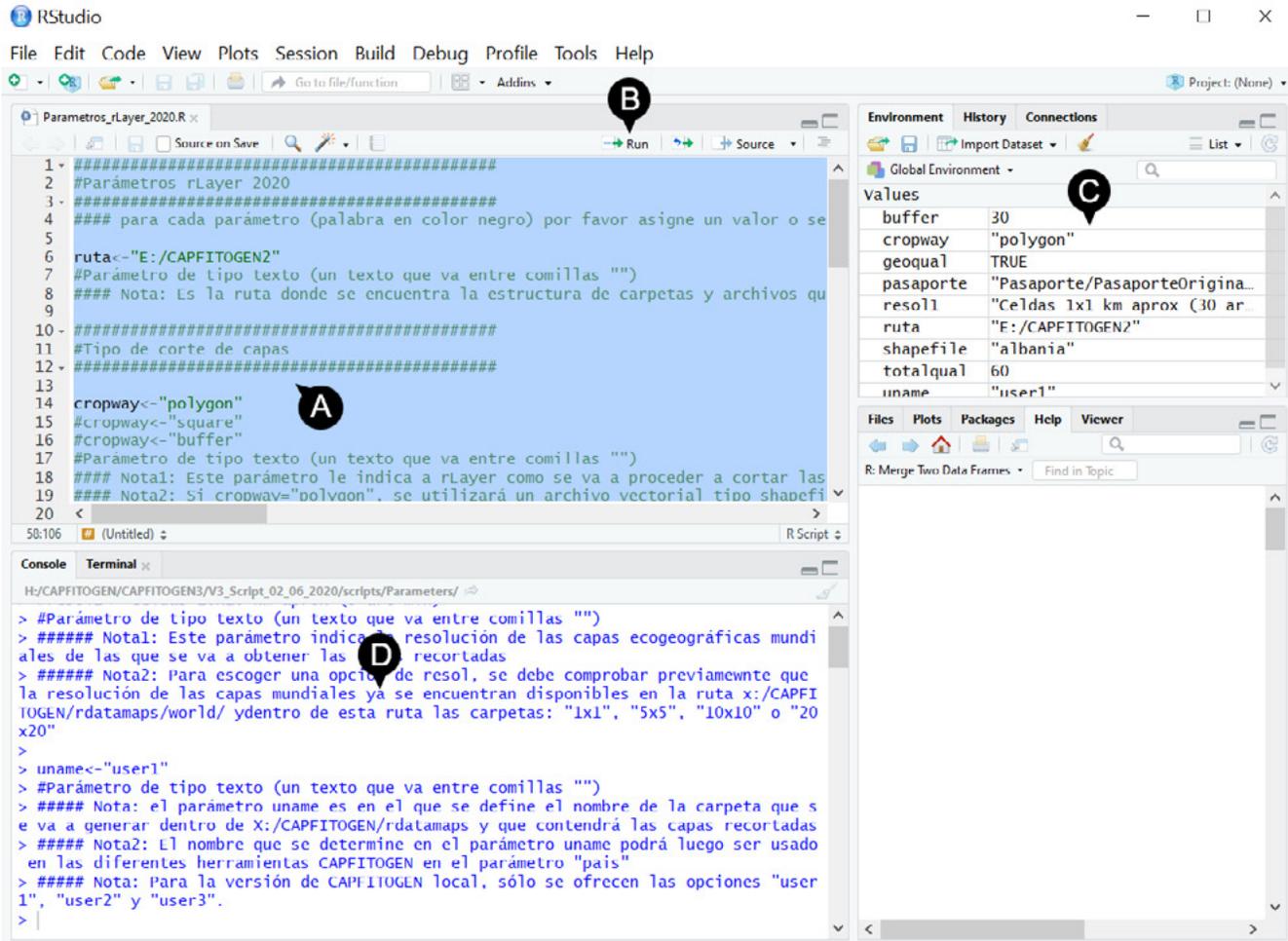


Figura 6. Ejecución del script de parámetros de una herramienta, en este caso rLayer. A) El primer paso, luego de configurar el script es seleccionarlo completamente, de manera que la ventana superior izquierda se resalta con un fondo color azul. B) Ejecute el script haciendo click en el pequeño botón que dice “Run”. C) Luego de ejecutar el script, aparecerá en la ventana superior derecha una lista de objetos que se acaban de crear y que serán los que requiere el script de la herramienta. D) También aparecerá en la ventana inferior izquierda de avance del proceso un texto en color azul que indica las acciones ejecutadas. En caso de que se produzca algún error, en este cuadro aparecería un mensaje en texto de color rojo.

Antes de ejecutar el script de la herramienta correspondiente, debe revisar en el cuadro superior derecho si los elementos que creo son los correctos. Si se percata de que cometió un error en su configuración, puede corregir el problema de manera sencilla, haciendo la corrección en el script de parámetros y repitiendo los pasos para ejecutarlo nuevamente.

Si desea guardar la configuración del script de parámetros para futuros procedimientos, simplemente vaya a la pestaña “File” y seleccione la opción “Save as...”. Así podrá guardar el script con la configuración deseada y luego lo podrá ejecutar cuando desee.

Luego de abrir el script de la herramienta, aparecerá en el cuadro superior izquierdo y su ejecución se realiza de manera idéntica que el script de parámetros (ver Figura 6). El script de herramienta suele ser bastante largo, por lo que se debe tener mucho cuidado al momento de seleccionarlo, pues no puede quedar ninguna parte sin ser resaltada en color azul (color de fondo), desde la primera hasta la última línea. Luego se debe hacer click en el botón “Run”, y gradualmente se va mostrando en texto azul en el cuadro inferior izquierdo todo lo que se va ejecutando. Al terminar la ejecución, aparecerá en el cuadro de la parte superior derecha los elementos generados y podrá buscar los resultados en la carpeta designada para ello.

El usuario debe tener ciertas precauciones al ejecutar el script de la herramienta que corresponde al script de parámetros que previamente configuró y ejecutó. Múltiples errores se producirán de no haber concordancia entre los dos scripts. Por otra parte, el script de la herramienta NO debe ser modificado a menos que el usuario sea un experto en el lenguaje R y quiera introducir alguna variación sobre el script original. Aun contando con algo de experticia, una modificación del script original, voluntaria o sin intención por parte del usuario, puede originar problemas desde productos inesperados o no deseados hasta bloquear la ejecución de las funciones e impedir la generación de cualquier producto.

2.1.3 Errores en CAPFITOGEN v3 modo local

Como se indicaba anteriormente, los errores que puedan producirse en la ejecución, tanto del script de parámetros como del script de herramientas, aparecerán en la ventana de avance del proceso (inferior izquierda), en texto de color rojo.

Los avisos en rojo pueden ser debidos a varias causas, no necesariamente errores, pero si todos los errores aparecerán de este color. Como lo muestra la Fig. 7, la carga de los paquetes de R por primera vez, origina una serie de texto en color rojo, que en la mayoría de veces no ofrece información de mayor importancia al usuario. También un texto en rojo puede deberse a que R quiere avisar de algo al usuario, pero no se puede considerar como un error. Estos avisos aparecen precedidos de la frase “Warning message:” y a continuación indica el asunto que se debe informar. En ocasiones, cuando los mensajes de advertencia (warning) son numerosos (pues provienen de procesos repetitivos), se hace necesario escribir en la última línea de la parte inferior del texto en la ventana de avance del proceso lo siguiente: warnings()

Luego se da click en “enter” y entonces aparecerán los mensajes de advertencia. Y finalmente están los mensajes de error, como los ejemplos que muestra la Fig. 7.

También puede aparecer texto en color negro, que puede indicar una acción que se ha realizado, pero que en ningún caso correspondería a un error. Varios de los errores más frecuentes se encuentran documentados en la parte final de este manual de usuario, junto con una explicación del por qué puede producirse y alternativas de solución.

```

Source
Console Terminal x
~/
> library(dismo)
> library(rgdal)
rgdal: version: 1.4-8, (SVN revision 845)
Geospatial Data Abstraction Library extensions to R successfully loaded
Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
Path to GDAL shared files: C:/Users/hmparraq/Documents/R/win-library/3.6/rgdal/gdal
GDAL binary built with GEOS: TRUE
Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
Path to PROJ.4 shared files: C:/Users/hmparraq/Documents/R/win-library/3.6/rgdal/proj
Linking to sp version: 1.4-1
>
> #introducción tabla de lista de países y resoluciones de extracción a elegir y traducción
> #Rversion
> vvv<-R.Version()
> vvv<-as.numeric(vvv$year)
>
> if(vvv<=2019){
+   resol<-read.delim("resol.txt")
+ }
> if(vvv>2019){
+   load("resol.RData")
+ }
Error in readChar(con, 5L, useBytes = TRUE) : cannot open the connection
In addition: Warning message:
In readChar(con, 5L, useBytes = TRUE) :
  cannot open compressed file 'resol.RData', probable reason 'No such file or directory'
> resol<-subset(resol,resolucion==paste(resol1))
Error in subset(resol, resolucion == paste(resol1)) :
  object 'resol' not found
> resol<-as.character(resol[1,2])
> setwd(paste(ruta,"/rdatamaps",sep=""))
> dir.create(as.vector(paste(uname)))
Warning message:
In dir.create(as.vector(paste(uname))) : 'user2' already exists
> setwd(paste(ruta,"/rdatamaps/",uname,sep=""))
> dir.create(as.vector(paste(resol)))
Warning message:
In dir.create(as.vector(paste(resol))) : '1x1' already exists
> setwd(paste(ruta))

```

Figura 7. Textos en color rojo que pueden aparecer en la ventana de avance del proceso. A) Aquí se indica que se solicitó a R cargar algunos paquetes (dismo y rgdal). B) R muestra en color rojo, que los paquetes se han cargado por primera vez, lo cual NO es un error. Una nueva carga no generará texto de color rojo. C) En este punto se muestra un verdadero mensaje de error, el cual siempre inicia con la palabra “Error”. En el mensaje luego se lee “...cannot open the connection...” y más adelante dice “cannot open compressed file ‘resol.Rdata’”. Esto significa que R no puede encontrar el archivo resol.Rdata donde debería estar dentro de la estructura de archivos y carpetas de CAPFITOGEN3, lo cual probablemente se originó por una mala configuración de un parámetro. D) Este segundo mensaje de error, avisa de que no se encuentra el objeto ‘resol’. Este segundo error es consecuencia del primero, ya que no se logró introducir una tabla y ello no permitió la creación dentro de R de un objeto. Por tanto, el error que hay que atender primero es el del mensaje C). E) En algunas ocasiones el texto en rojo indica un aviso de advertencia sobre algo que podría ser irregular o que quien diseñó la función en R considera que el usuario debe saber. No corresponde a un error.

2.2. Modo on server (on line)

La tercera versión de las herramientas CAPFITOGEN presenta esta importante novedad a los usuarios. Un modo de usar las herramientas, en la cual el usuario no debe descargar ningún archivo de instalación, ningún software particular, y tampoco debe contar con una computadora con especificaciones especiales para obtener los productos deseados. Simplemente el usuario accede mediante su navegador de internet a una dirección web donde debe registrarse como usuario y posteriormente usar las herramientas y producir los resultados que esperan. El software R y los scripts estarán todos instalados en un servidor y se mostrarán y configurarán a través de formularios amigables. El usuario, luego de configurar el funcionamiento de la herramienta, dará inicio al análisis y esperará que los resultados se guarden en su zona de usuario, un espacio privado donde se almacena tanto las tablas particulares del usuario necesarias para los análisis, como los resultados de la aplicación de las diferentes herramientas.

2.2.1 Cómo registrarse en CAPFITOGEN v3 modo on server

El procedimiento para registrarse como usuario de CAPFITOGEN v3 es muy sencillo y similar al registro que suele realizar un usuario en casi cualquier plataforma, con una validación vía correo electrónico. Los pasos son:

1. Diríjase a <http://onserver.capfitogen.net/>. Visualizará una página de acceso como lo muestra la Fig 8. Para iniciar su registro, haga click en donde se indica “Regístrate aquí” (Fig 8E). Esto abrirá un formulario donde el usuario debe ingresar los siguientes datos: Nombre, apellido, correo electrónico, una contraseña sugerida por el usuario y un espacio adicional para confirmar esa contraseña. Finalmente aparece una casilla de verificación para aceptar los términos y condiciones y el botón de finalizar el registro (“Registrarse”).
2. El usuario recibirá un mensaje a su correo electrónico de parte de no-reply@verificationemail.com, el cual no debe responder, y donde aparece un código que debe memorizar.
3. La página de registro ha dado paso a la página de validación del código que el usuario recibió a su correo electrónico (ver Fig. 9). Aquí el usuario debe indicar el correo electrónico que registró y el código que recibió. Posteriormente debe hacer click en el botón “Validar”.

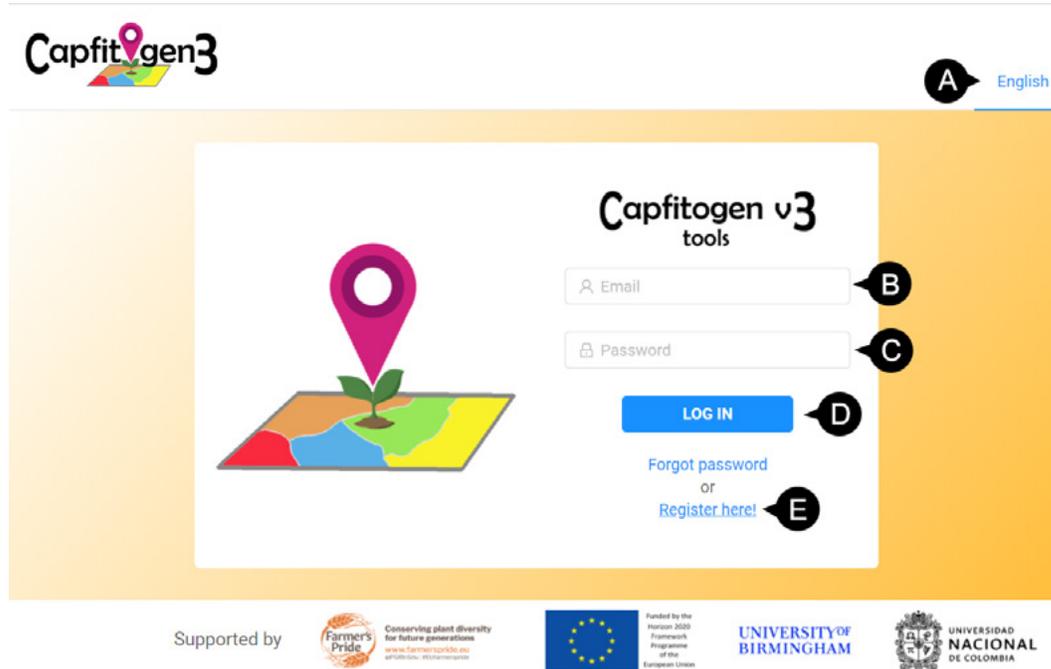


Figura 8. Portal de entrada a CAPFITOGEN3 modo on server. A) Configuración de idioma. En esta parte puede cambiar de idioma entre Español e Inglés. B) Espacio para introducir el correo electrónico con el cual el usuario se ha registrado. C) Espacio para ingresar la contraseña del usuario registrado. D) Botón para iniciar sesión como usuario registrado. E) Vínculo para iniciar el proceso de registro de usuario.

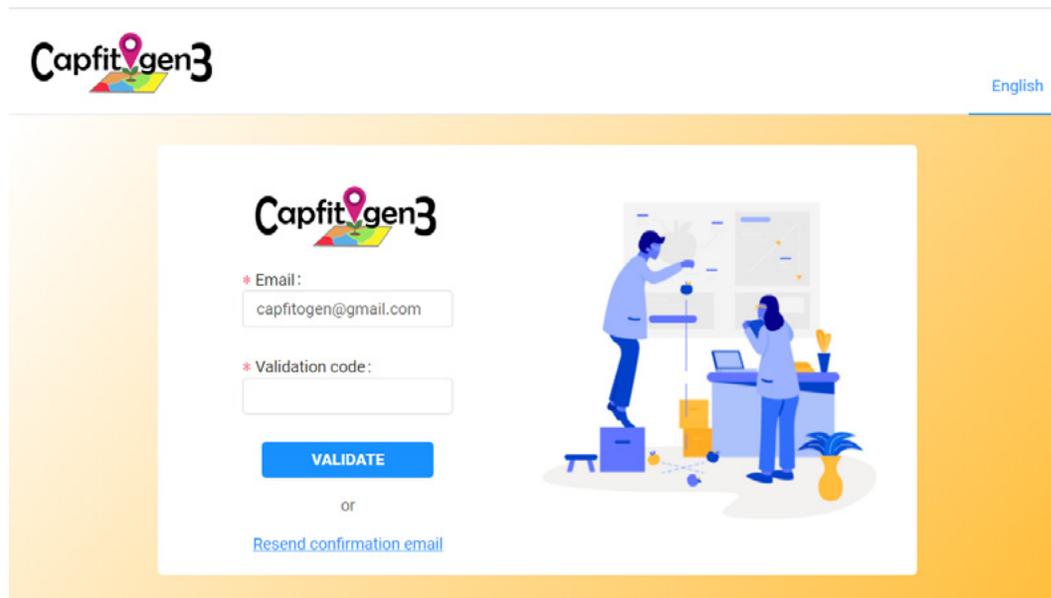


Figura 9. Página de validación del registro. En esta página se ingresa el código recibido vía correo electrónico.

- Una vez se ha hecho la validación exitosa del código, el usuario ya puede iniciar sesión con el correo electrónico y con la contraseña que definió en el proceso de registro, para lo cual se despliega automáticamente la página de acceso a CAPFITOGEN3 *on server*. Si esto no ocurriese, simplemente puede de nuevo ingresar la dirección en su navegador y tendrá a disposición la página de acceso.

2.2.2 Cómo utilizar CAPFITOGEN v3 modo *on server*

En el momento de iniciar sesión, se abrirá la vista principal de las herramientas, donde se despliegan diferentes zonas o accesos a zonas, las cuales se indican en la Fig. 10.

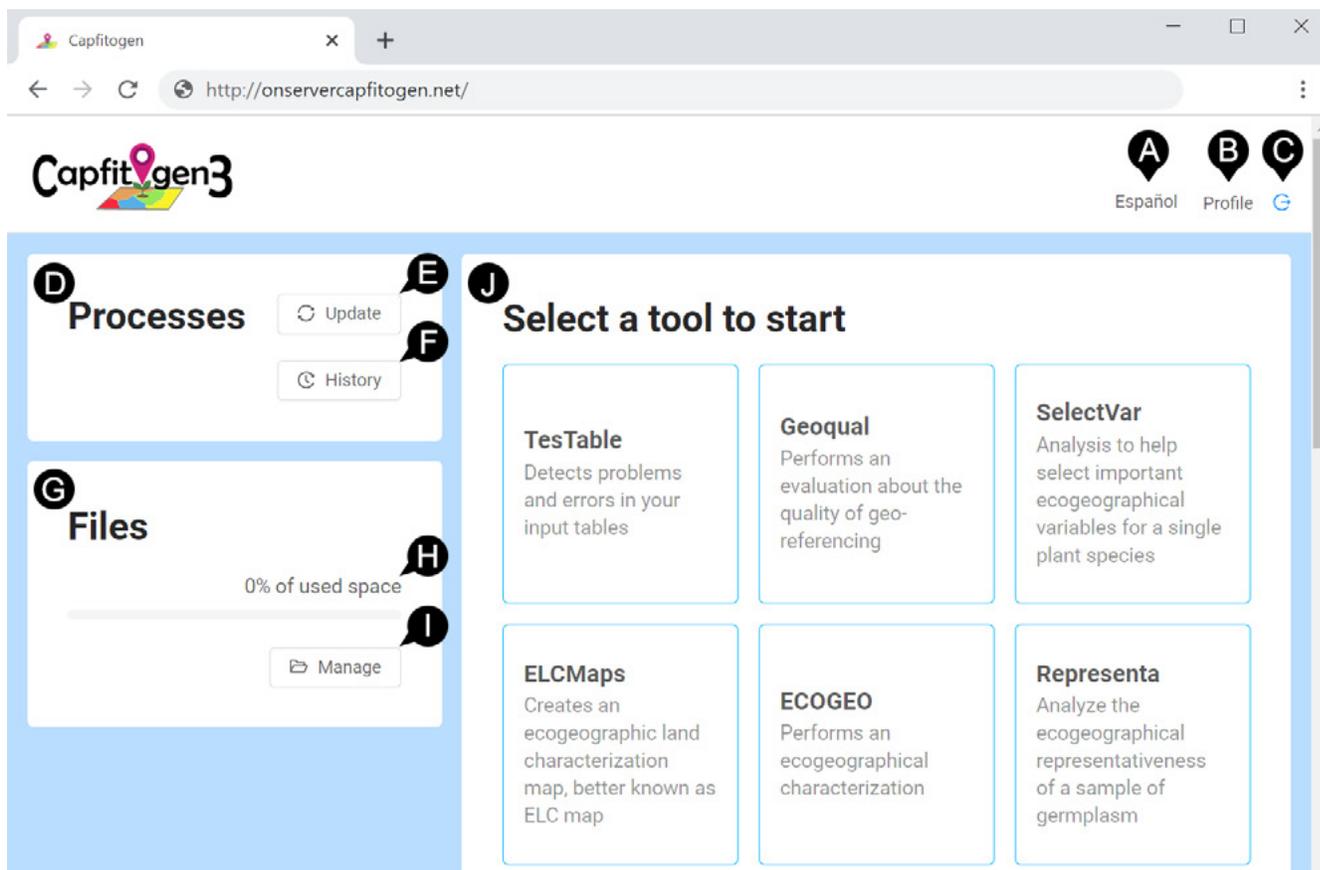


Figura 10. Vista principal de las herramientas CAPFITOGEN 3 modo *on server*. A) Selector de idioma. B) Acceso a la información del perfil de usuario. C) Botón de cierre de sesión. D) Zona de procesos. E) Botón para actualizar el estado de los procesos. F) Botón para acceder al historial de los procesos ejecutados. G) Zona de archivos de usuario y resultado. H) Indicador del porcentaje del espacio asignado a un usuario que ya tiene en uso. I) Botón para acceder a los archivos almacenados en el espacio asignado al usuario. J) Panel de selección de las herramientas.

Básicamente el usuario tiene cuatro secciones donde realizar acciones:

- Panel de selección de herramienta: Aquí el usuario busca la herramienta con la que desea trabajar, y al hacer

click sobre el nombre de la herramienta de interés, ingresa a un formulario en blanco de la misma (Fig 10-J).

2. Zona de procesos: Esta zona está compuesta del panel que muestra el avance de ejecución de alguna herramienta (Fig 10-D), y del histórico al cual se accede a través del botón indicado en la Fig10-F. En la Fig. 11 se detalla las partes de la zona de procesos. A través del uso de colores de semáforo (verde, rojo o amarillo) se indica si el proceso finalizó o si aún se está ejecutando.

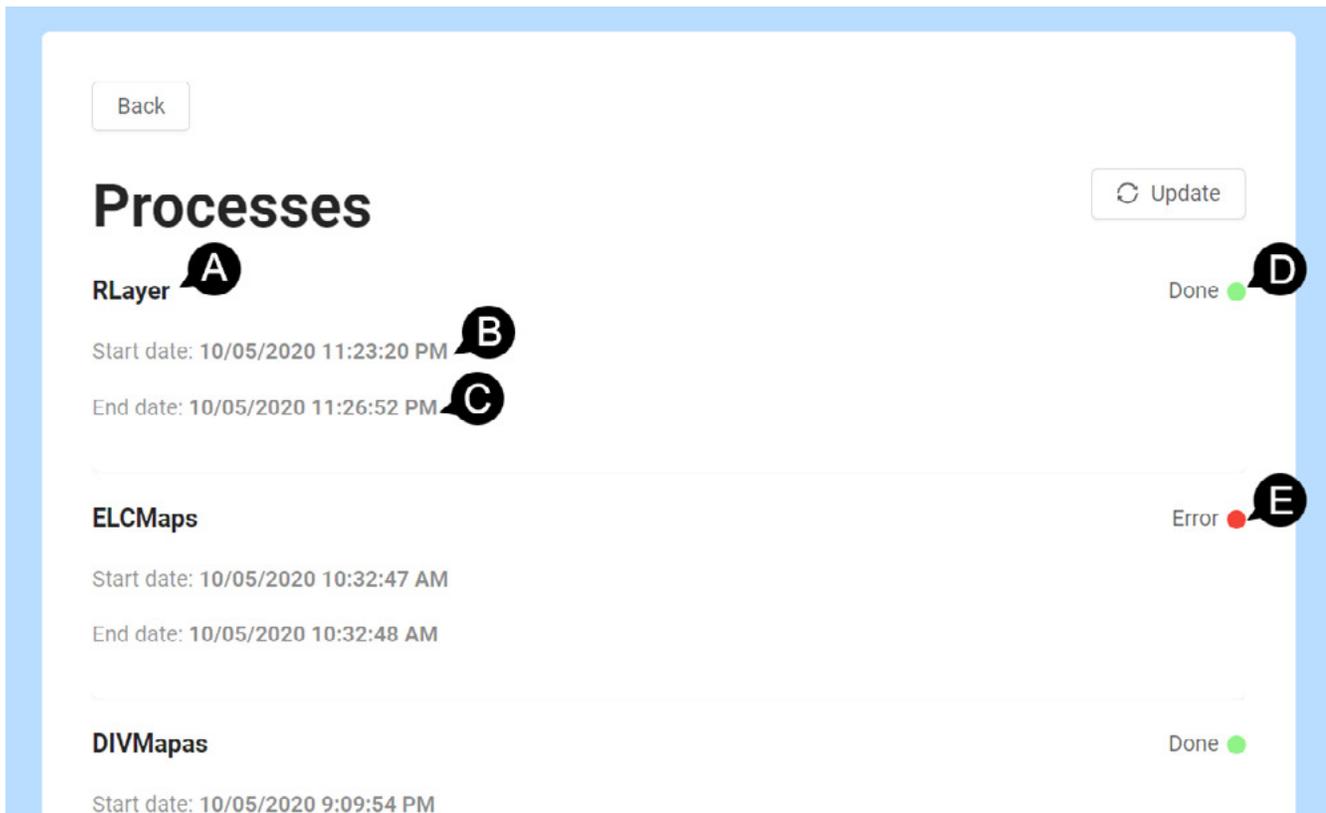


Figura 11. Zona de procesos. A) Nombre de la herramienta que se ejecutó. B) Fecha y hora de inicio del análisis de la herramienta, en este caso Rlayer. C) Fecha y hora de finalización del análisis. D) Estado del proceso. En este caso se reporta con un círculo de color verde que el análisis se ha ejecutado con éxito y el texto (done o finalizado) confirma que el proceso ya fue realizado. E) En este caso, el reporte es de error (color rojo) y por tanto no finalizó correctamente. También puede aparecer un reporte en color amarillo, que indica que el proceso está aún pendiente de finalizar.

3. Zona de archivos de usuario: Se trata del espacio que se asigna a cada usuario para el almacenamiento de sus archivos de trabajo (tablas de pasaporte, tablas de fuentes externas, tablas de caracterización, etc.) o de los resultados de la ejecución de las herramientas. La Fig. 12 muestra el interior de la zona de archivos de usuario al cual se accede a través del botón que se señala en la parte I de la Fig. 10. En esta zona, los resultados de los análisis realizados por las herramientas se van almacenando en carpetas con el nombre de la herramienta de donde provienen. En subcarpetas se van separando los resultados para la misma herramienta, identificándolos por el día y la hora en que fueron obtenidos (lo cual sirve para nombrar la subcarpeta). Dentro de las subcarpetas se encuentran los archivos de resultado, los cuales pueden ser descargados y/o borrados,

con el fin de liberar espacio de la zona de usuarios. También en esta zona se puede encontrar una carpeta cuyo nombre es “uploads” donde se van a almacenar todos los archivos que el usuario suba para realizar los procesos de las herramientas. Estos archivos pueden subirse mediante el botón “Cargar archivos a su carpeta de uploads” (Fig. 12-B) o a través de botones en cada herramienta habilitados para que el usuario suba sus archivos.

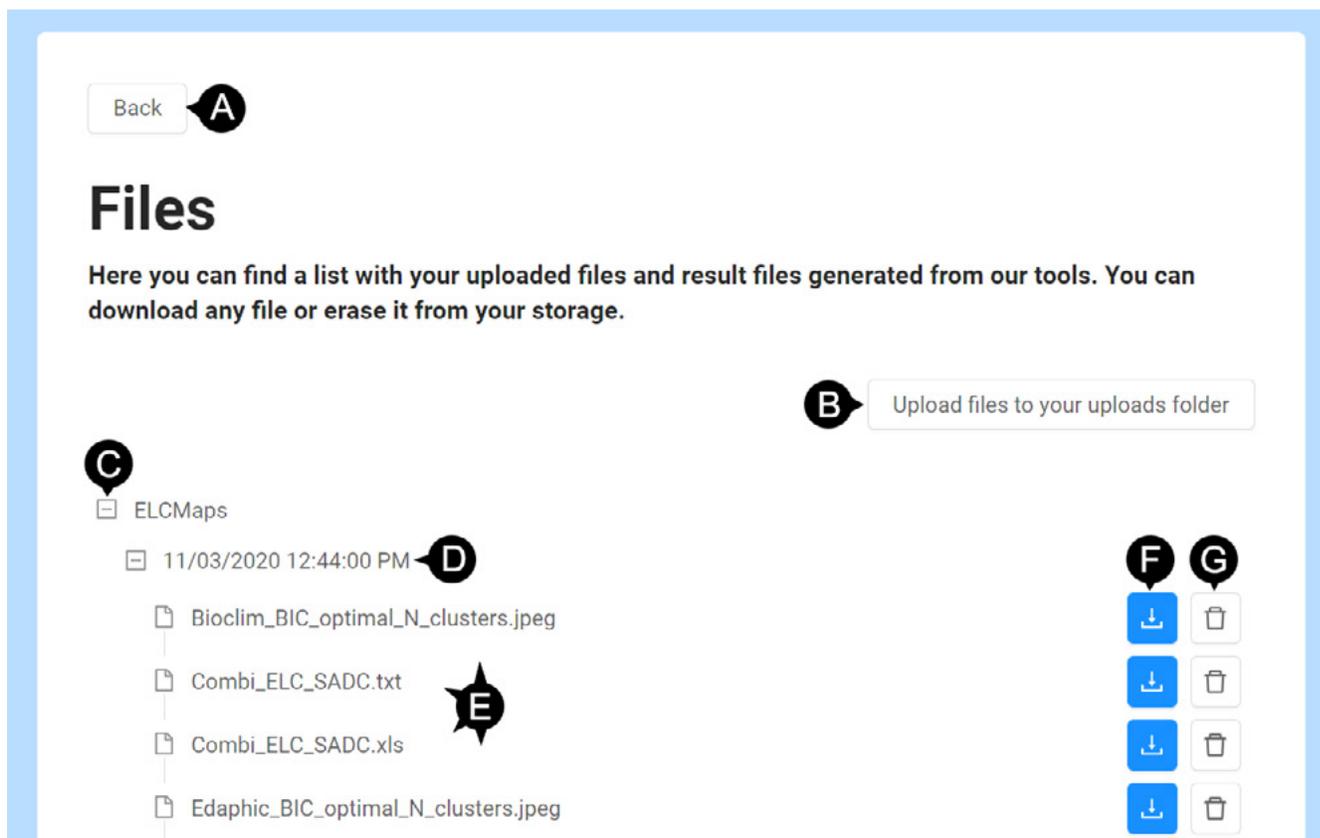


Figura 12. Zona de archivos de usuario y resultados. A) Botón para retroceder y regresar a la vista principal. B) Botón para que el usuario pueda subir archivos necesarios para la ejecución de las herramientas. Si el usuario sube los archivos con antelación, cuando se enfrente al formulario de la herramienta no tendrá que subir de nuevo los archivos sino simplemente seleccionarlos de una lista desplegable. C) Carpeta con el nombre de la herramienta de donde se produjo los resultados. D) Subcarpeta con la información del día y hora de ejecución de la herramienta que produjo los resultados. E) Archivos resultado de la ejecución de la herramienta. F) Botón para descargar el archivo de resultado correspondiente. G) Botón para borrar el archivo correspondiente de la zona de archivos de usuario y resultados.

4. Zona de perfil del usuario. Accesible a través del link que se visualiza en la Fig. 10-B, en esta zona se puede modificar los datos solicitados en el proceso de registro (ver Fig. 13).

Back

Mauricio Parra Quijano
hmparraq@unal.edu.co

change password

To change your personal information please use the form below

* Name: Mauricio

* Last name: Parra Quijano

Title: Teacher

Institute or company: Universidad Nacional de Colombia

What you will use the tools for: Agrobiodiversity analysis

Secondary email: hmparraq@unal.edu.co

Receive notifications from our newsletter

Update

Figura 13. Zona del perfil del usuario y de cambio de contraseña.

La utilización de los formularios para la configuración de los parámetros de cada herramienta, los cuales son accesibles a través del panel de selección de herramientas, es bastante similar a la forma como funcionaban los formularios de la versión 2.0 de CAPFITOGEN, aunque algunas notorias facilidades aparecen en el modo on server. Es importante resaltar que en CAPFITOGEN 3 modo on server los parámetros “ruta” y “resultados” desaparecen ya que no son necesarios. También los parámetros donde se introducían los nombres de archivos tipo tabla o mapa, ahora se presentan como botones para seleccionar los archivos y subirlos al servidor, o el usuario los puede haber subido con anterioridad y ya en el formulario simplemente seleccionar el archivo apropiado de una lista de archivos subidos y almacenados en la zona de archivos de usuario y resultados, dentro de la carpeta “uploads”.

A continuación se muestran las partes de los formularios y los diferentes tipos de parámetros que se pueden encontrar en el modo on server.

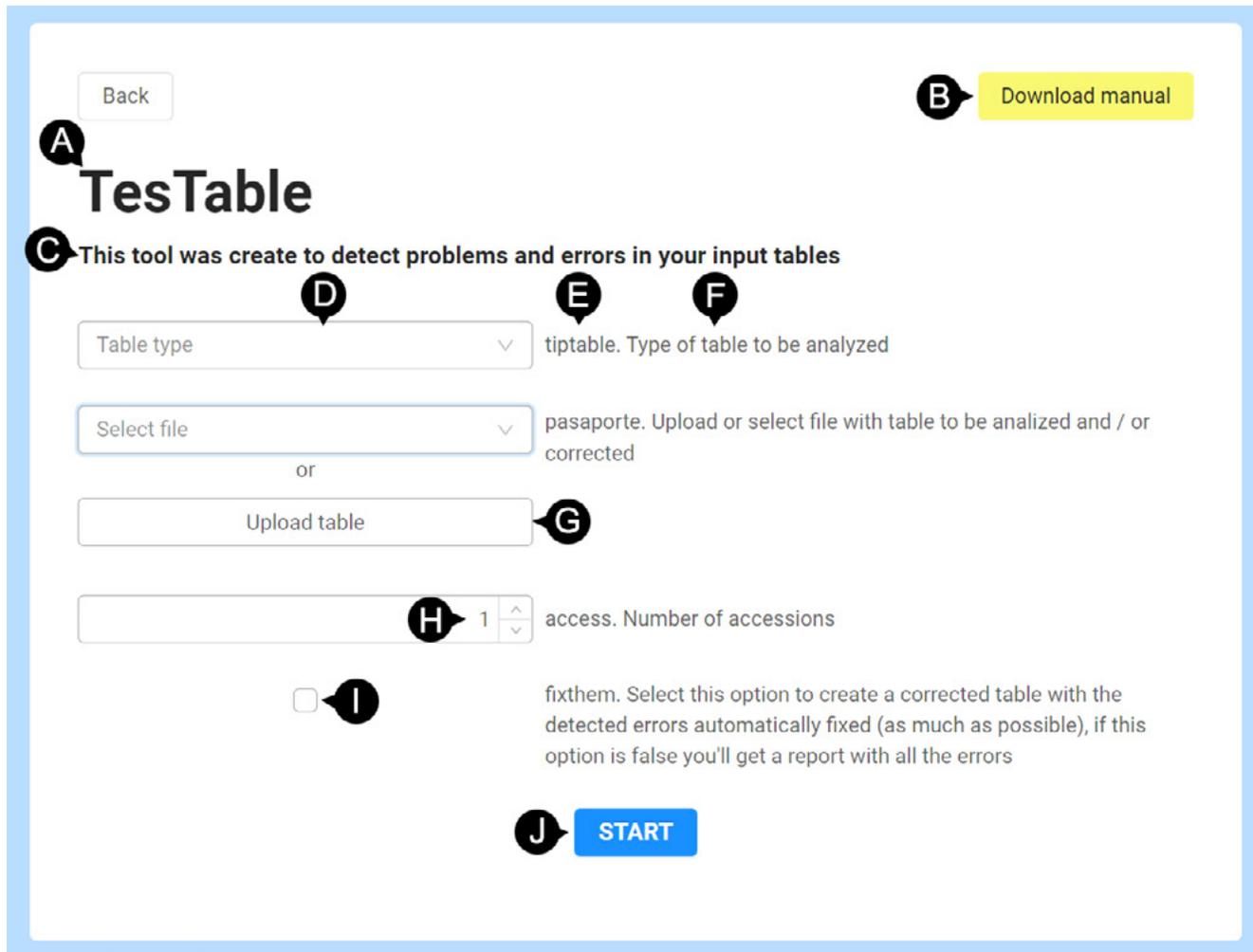


Figura 14. Ejemplo de formulario de herramienta (TesTable en este caso). A) Nombre de la herramienta. B) Botón para descargar el manual de usuario (particularmente el capítulo para la herramienta). C) Breve descripción de la función de la herramienta. D) Al lado izquierdo se sitúan las cajas donde se configuran los parámetros. E) Del lado derecho, lo primero que se encuentra es el nombre del parámetro. F) Seguido del nombre del parámetro, una breve descripción del mismo. G) Para el parámetro “pasaporte” y algunos otros donde el usuario aporta un archivo, aparecen dos opciones, una lista desplegable (para seleccionar el archivo previamente subido y que se conserva en la carpeta “uploads”) o un botón que permite buscar el archivo en el equipo local y luego subirlo al servidor. H) Parámetro de tipo numérico, donde se incrementa o reduce el dígito que configura el parámetro. I) Parámetro tipo checkbox (caja de verificación), donde al hacer click aparece un símbolo de check (✓), lo cual es equivalente a responder la pregunta del parámetro como “sí” o TRUE. Dejar sin marcar el checkbox significa decir “no” o FALSE. J) Botón para iniciar el proceso o ejecución de la herramienta, una vez se han configurado todos los parámetros.

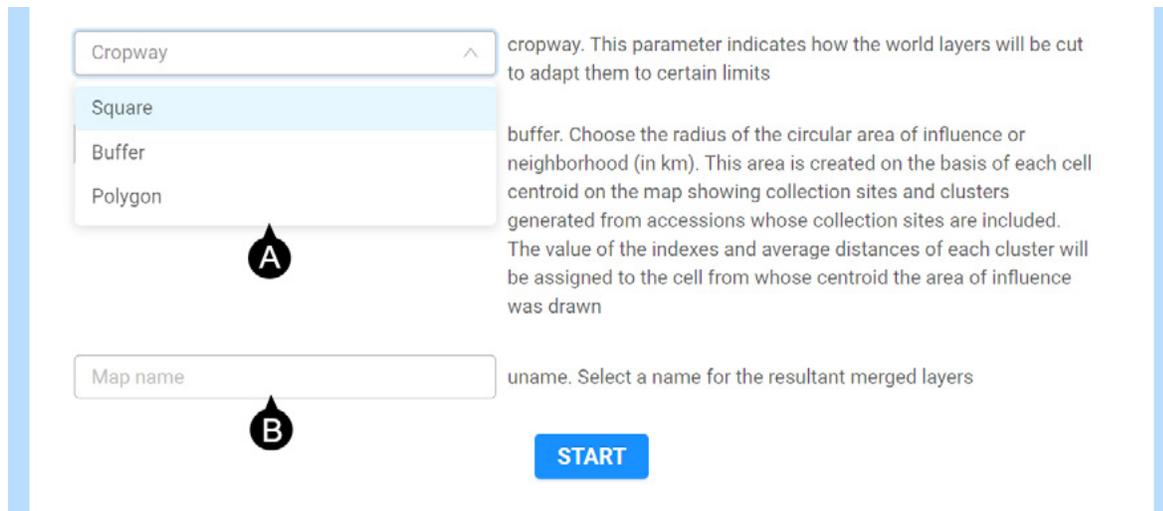


Figura 15. Parámetros tipo lista desplegable de opciones y texto. A) Parámetro de lista desplegable, donde sólo una opción puede ser elegible. B) Parámetro de texto, donde el usuario debe introducir una cadena de texto, usualmente corta y sin espacios.

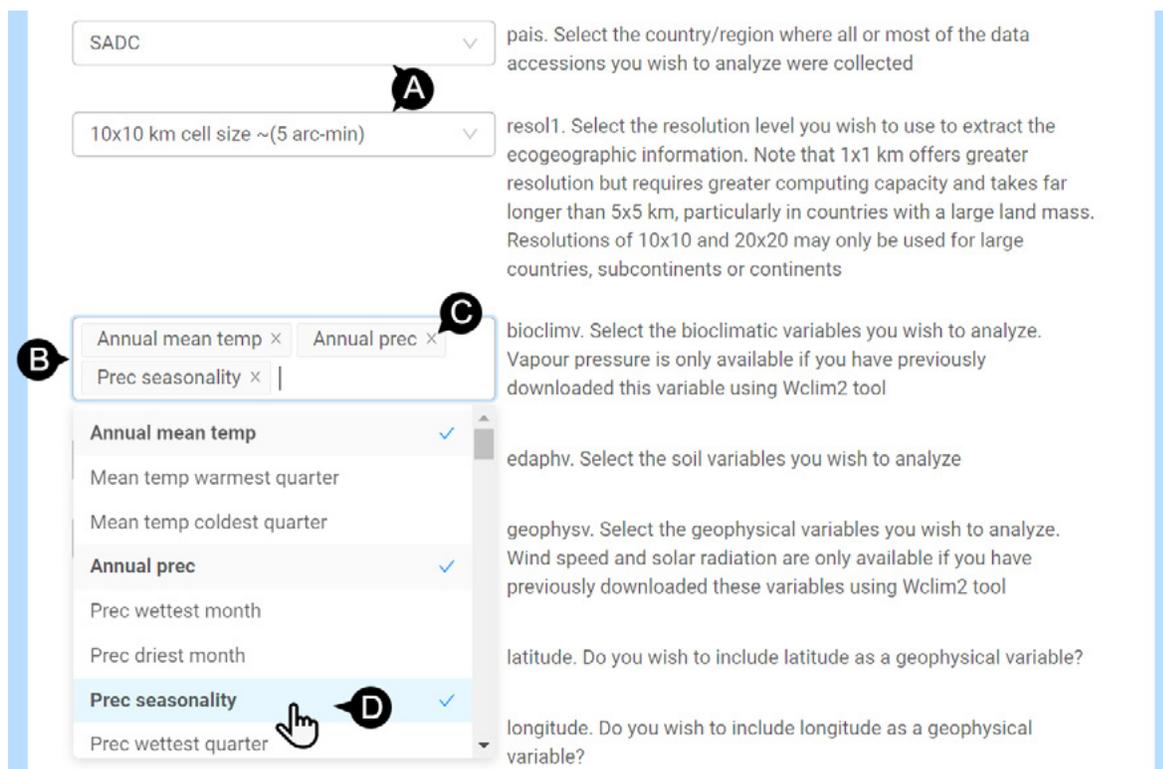


Figura 16. Otros tipos de parámetros en CAPFITOGEN 3 modo on server. A) Parámetros de despliegue de lista de opciones, donde sólo una es elegible. B) Parámetro de despliegue de opciones y selección múltiple. C) Las opciones elegidas aparecen dentro de la caja y pueden eliminarse de ella, haciendo click sobre la X. D) Sobre la lista desplegable, el usuario a través de clicks va seleccionando las variables que desea incluir.

Luego de configurar correctamente los parámetros de la herramienta y hacer click en el botón para iniciar el proceso (Fig. 14-J), aparecerá una ventana al usuario avisando del inicio de la ejecución del proceso en R en el servidor, y de la disponibilidad posterior de los resultados (Fig. 17)

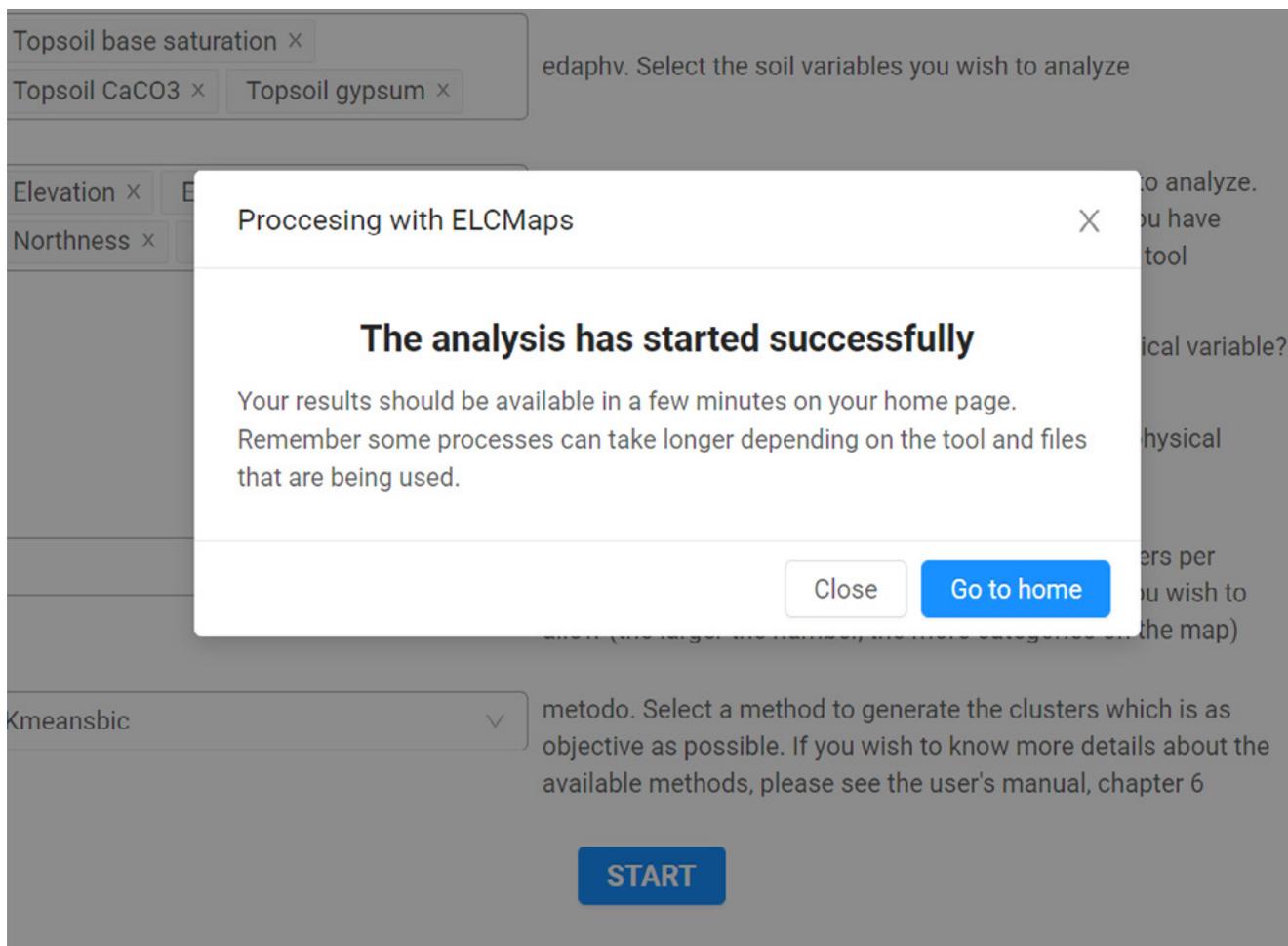


Figura 17. Aviso de inicio de proceso exitoso. El mensaje también indica que los resultados estarán disponibles desde la vista principal (particularmente en la zona de archivos de usuario y resultados) y que su disponibilidad puede tardar en la medida de la complejidad del proceso que realiza la herramienta y el tamaño de los archivos que serán procesados.

El usuario puede regresar a la vista principal, donde dependiendo del tiempo que pueda tomar el proceso de la herramienta, observaría un círculo de color amarillo (con la etiqueta “en progreso”) en la zona de procesos, lo cual indica que aún no se han producido los resultados. En este momento el usuario puede incluso salir del sistema y regresar posteriormente para comprobar si los resultados ya están disponibles (aparecerá en frente del proceso un círculo de color verde). También puede esperar allí y de tanto en tanto hacer click sobre el botón “Actualizar” de la zona de procesos, hasta que el círculo torne a verde y la etiqueta indique “Finalizado”, entonces puede ingresar a la zona de archivos de usuario y resultados para descargar sus resultados (ver Fig. 18).

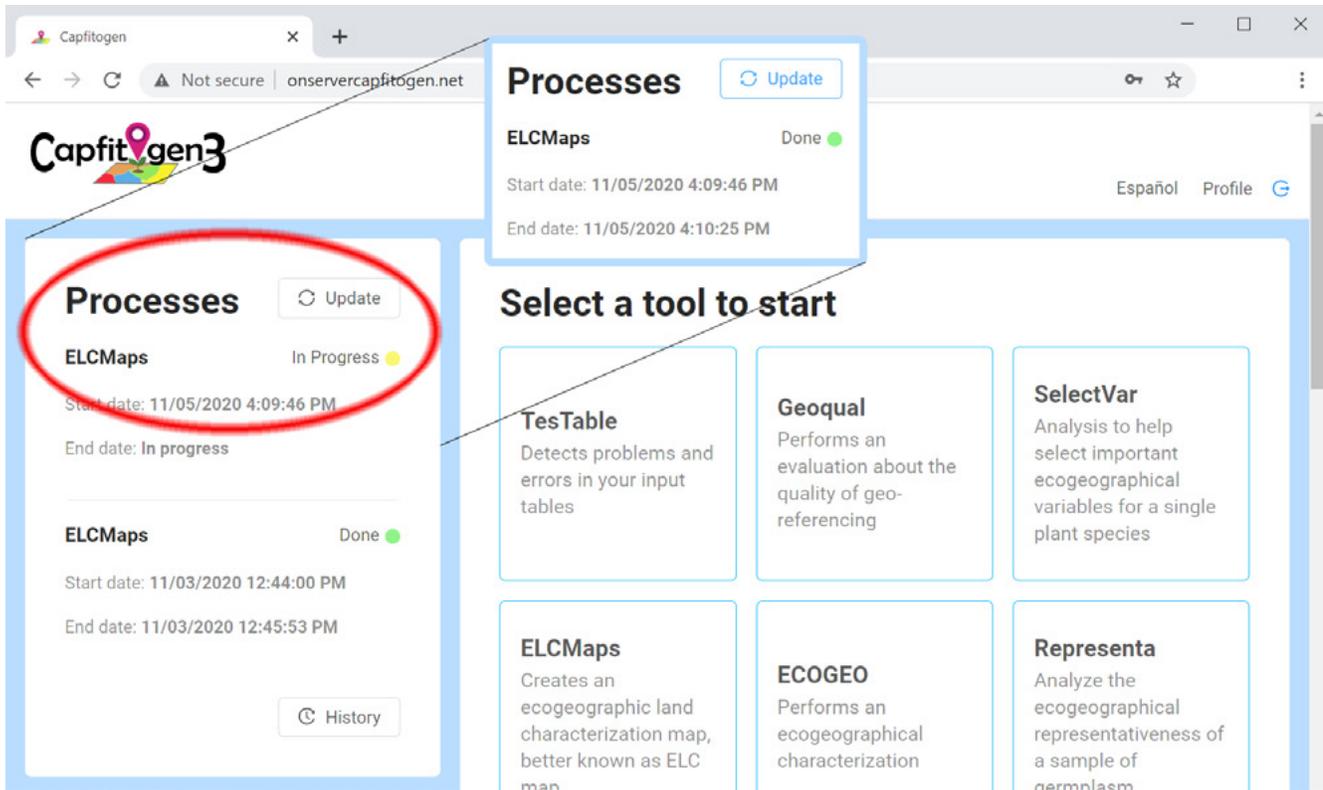


Figura 18. Vista principal luego de regresar del formulario de una herramienta e iniciar su proceso. En la zona de procesos (resaltada en un círculo rojo) se puede ver el último proceso ejecutado en la parte superior, y el o los siguientes en la línea temporal hacia abajo. En este caso se nota el círculo amarillo y la leyenda “En progreso” indicando que aun no se han producido los resultados. Luego de un tiempo prudencial, al hacer click en el botón “Actualizar”, si el proceso ha terminado, se observará un cambio como lo ilustra la imagen hacia un color verde (finalización exitosa) y la leyenda “Finalizado”.

Finalmente, cuando el proceso no termina exitosamente y se produce un error, lo primero que se evidencia es que en la zona de procesos el estado que aparece es de “Error”, junto con un círculo rojo (Fig. 19-A). Para conocer la naturaleza del error, es necesario entrar a la zona de archivos de usuario y resultados, donde dentro de la carpeta y subcarpeta de la herramienta utilizada y de la fecha y hora de ejecución, respectivamente, se encontrará un archivo de nombre “ERROR.txt” (Fig. 19-B). Para que el usuario pueda enterarse del error ocurrido es necesario descargar el archivo de texto y posteriormente visualizarlo con cualquier software de visualización o edición de texto. El texto del error puede compararse con los errores tipificados en el capítulo 18 “Errores frecuentes”, con la idea de poder solventar el problema.

The screenshot displays two panels from the CAPFITOGEN3 interface. The left panel, titled "Processes", lists three tasks: "Geoqual" (marked with a red "Error" icon and labeled "A"), "TestTable" (marked "Done" with a green dot), and "ELCMaps" (marked "Done" with a green dot). Each task entry includes start and end dates. An "Update" button is visible at the top right of this panel. The right panel, titled "Files", shows a directory structure: "ELCMaps" > "Geoqual" > "11/05/2020 7:10:48 PM". Inside this folder, there are three files: "ERROR.txt" (marked with a red "B" icon), "Error_process_info.txt", and "Parameters_Geoqual.txt". The "ERROR.txt" file has a blue download button (marked with a red "C" icon) and a trash icon next to it. A text box at the top right of the "Files" panel says "Upload files to your uploads folder".

Figura 19. Notificación de errores en la ejecución de una herramienta CAPFITOGEN 3 modo on server. A) Aviso de error en la zona de procesos, con una leyenda de Error y un círculo color rojo. B) en la zona de archivos de usuario y resultados, dentro de la carpeta de herramienta y subcarpeta de fecha de ejecución, aparece un archivo llamado "ERROR.txt". C) Para ver el mensaje de error producido por R, el archivo "ERROR.txt" se puede descargar haciendo click en este botón de color azul.



Taller Nacional, Tirana, Albania, Noviembre 2014.



3 | Herramienta TesTable

3.1. Los formatos de tablas que aceptan las herramientas CAPFITOGEN3

Ejemplos y moldes de los diferentes tipos de tablas usadas en CAPFITOGEN3 se encuentran disponibles desde <https://drive.google.com/drive/folders/1xCnILZgzW0uDeClDvcxbADv9H583xzpn?usp=sharing>. De acuerdo a la herramienta que se vaya a utilizar, al usuario se le podrá requerir el aporte de las siguientes tipos de tablas de información:

3.1.1 Tabla de datos de pasaporte

Es una tabla que en esencia contiene los campos indicados en la lista de descriptores multicultivo de FAO-Bioversity (formato MCPD en adelante) versión 2 publicada en 2012. Se añaden a esta lista cinco campos referentes a la descripción de la localidad en términos de niveles administrativos (ADM1, ADM2, ADM3 y ADM4) y al nombre del país (no en código ISO) (NAMECTY). A pesar que puedan existir otros formatos de datos de pasaporte usados por diferentes países o centros de investigación, la lista de descriptores MCPD (tanto la versión 1 como la 2) ha sido la referencia para muchos bancos de germoplasma y sistemas de información (como Genesys - <https://www.genesys-pgr.org/>) a la hora de almacenar sus datos de pasaporte, usualmente con algunas modificaciones y ajustes menores.

La tabla de datos de pasaporte regular para CAPFITOGEN contiene 45 campos (40 de la lista de descriptores MCPD de 2012 más 5 añadidos). Sin embargo pueden aparecer más campos si por ejemplo, el usuario desea incluir un campo sobre la disponibilidad del germoplasma (campo AVAILAB) muy útil para herramientas como ColNucleo o FIGS_R; o si ha pasado por la herramienta GEOQUAL (que determina la calidad de la georreferenciación) la tabla deberá contener 5 campos adicionales (denominados SUITQUAL, LOCALQUAL, COORQUAL, TOTALQUAL y TOTALQUAL100). Todos los campos adicionales debidos a la inclusión de información sobre disponibilidad o a la determinación de la calidad de la georreferenciación se encontrarán después de los 45 campos regulares (en la parte derecha de la tabla).

3.1.2 Tabla de datos de fuentes externas

Corresponde a una tabla de 15 campos, algunos de ellos son iguales en cuanto a nombre y contenido a los de la tabla de datos de pasaporte. Esta tabla sirve para introducir en la herramienta Representa la información sobre ocurrencia de poblaciones provenientes de fuentes diferentes a la colección de germoplasma o inventario de recursos fitogenéticos objetivo (información que estaría consignada en la tabla de datos de pasaporte). En ningún caso en esta tabla deben aparecer campos adicionales.

3.1.3 Tabla de datos fenotípicos

Tabla que incluye una columna de identificación de entrada (ACCENUMB) que debe corresponder al mismo campo de identificación de la tabla de datos de pasaporte y dos o más campos para las variables fenotípicas. Los tipos de

variables que se pueden incluir en esta tabla son cuantitativas y cualitativas expresadas numéricamente.

3.1.4 Tabla de datos de naturaleza de variable fenotípicas

Dado que la tabla de datos fenotípicos admite variables cuantitativas y cualitativas, en esta tabla se debe especificar la naturaleza de cada variable. Esta tabla sólo contiene tres campos:

ID: Un número identificador que puede ser consecutivo

NOMVAR: El nombre de cada variable tal y como aparecen en la tabla de datos fenotípicos (en el mismo orden)

NATVAR: En este campo se indica la naturaleza de cada variable. Las opciones pueden ser: “Cuantitativo”, “Nominal”, “Ordinal”, “Binaria simetrico” y “Binario asimetrico” (sin tildes o acentos).

3.1.5 Tabla de datos genotípicos

Tabla que incluye una columna de identificación de entrada (ACCENUMB) que debe corresponder al mismo campo de identificación de la tabla de datos de pasaporte y dos o más campos para las variables genotípicas. Las variables genotípicas corresponden a datos de caracterización por marcadores de ADN de tipo dominante o codominantes pero expresados como presencia –ausencia de banda. Por tanto, ésta tabla sólo admite variables binarias no simétricas con valores 0 (ausencia de banda) o 1 (presencia de banda).

Una característica importante en todas las tablas que usan las herramientas CAPFITOGEN es que no admiten celdas en blanco. En caso de no contarse con información sobre una entrada y una variable o campo en particular, no se debe dejar en blanco la celda, sino rellenarse con el valor NA (que indica “no aplica”).

CAPFITOGEN requiere que las tablas sean preparadas en formato de texto delimitado por tabulaciones (archivos con la extensión .txt). Para que los usuarios puedan ajustar su información a los formatos requeridos por las herramientas CAPFITOGEN con mayor facilidad, se encuentra disponible en la carpeta “Formatos_Formats” una serie de archivos Excel que contienen para cada tipo de tabla los campos requeridos. Algunas de las tablas Excel contienen algunos pocos casos (entradas ficticias) que sirven como ejemplo para mostrar a los usuarios la forma correcta en que debe ser introducida la información. En caso de existir dudas sobre el significado o la forma correcta de introducción de información en alguno de los campos de cualquier tabla Excel, se debe ubicar el puntero del ratón sobre el encabezado de la columna y de ese modo se abrirá (en español o en inglés) una ventana de comentario con un breve texto sobre lo que debe contener el campo en cuestión y las indicaciones pertinentes sobre cómo llenarlo (ver Fig. 20).

	X	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	LATITUDE	DECLONGITU	LONGITU		OREFMETH	ELEVATION	COLLDATE	BREDCODE	BREDNAME	SAMPST
2	4124--N	NA	002094			NA	19780301	NA	NA	
3	452556S	NA	0684645W			120	1999----	NA	NA	NA
4	00----S	NA	079----W			NA	NA	NA	NA	
5	103936N	NA	0671051W			14	NA	NA	NA	
6	252712N	NA	1112853W			0	201002--	NA	NA	NA
7	NA	NA	NA			NA	194812--	NA	NA	
8	412403N	2.1628	0020945E			NA	19780301	NA	NA	
9	NA	-68.7717	NA			120	1999----	NA	NA	NA
10	NA	-79	NA			NA	NA	NA	NA	
11	NA	-67.202	NA			14	NA	NA	NA	
12	NA	-111.485	NA			0	201002--	NA	NA	NA
13	NA	NA	NA			NA	194812--	NA	NA	
14										

Figura 20. Captura de imagen de la tabla modelo de pasaporte que muestra los encabezados de campo. A) Campo de alta importancia (color verde). B) Campos que no tienen importancia para los análisis con herramientas CAPFITOGEN, sin color. C) Campos que pueden tener utilidad para algunas herramientas CAPFITOGEN, en color amarillo. D) Al situar el puntero del ratón sobre el encabezado de las columnas en la tablas Excel que se proporcionan al usuario en el modo local.

Las tablas Excel para datos fenotípicos, de naturaleza de variables fenotípicas y datos genotípicos incluyen una hoja de cálculo adicional en la que se explica en detalle cómo llenar la tabla (Fig. 21).

	A	B	C	D	E	F	G
1	ACCENUMB	D1	D2	D3			
2	acc_001	3910	1	0			
3	acc_002	3304	5	0			
4	acc_003	4728	4	0			
5	acc_004	903	5	0			
6	acc_005	3284	5	0			
7	acc_006	914	5	1			
8	acc_007	504	5	0			
9	acc_008	723	2	0			
10	acc_009	4402	3	1			
11	acc_010	3978	2	1			
12	acc_011	267	4	0			
13	acc_012	2942	1	0			
14	acc_013	710	5	1			
15	acc_014	3492	1	1			
16	acc_015	4236	4	0			
17	acc_016	3960	3	1			
18	acc_017	4513	3	1			
19	acc_018	184	3	1			

Figura 21. Captura de imagen del formato de tabla de datos fenotípicos, donde se aprecian las dos hojas de cálculo que contiene. A) En la hoja de cálculo "Morfología" o "Phenotypic", donde se debe adjuntar este tipo de información, aparecen dos tipos de columnas, la primera es la reservada para el código de accesión (ACCENUMB), común con otras tablas, y de color verde, de alta importancia. B) En la misma hoja, otras columnas sin color, también son muy importantes pues allí es donde debe ir la información de caracterización. C) Hoja de cálculo "Observaciones" o "Directions" donde se puede consultar las instrucciones de llenado de la parte A.

Una vez el usuario ha llenado los formatos de tablas en Excel con su propia información, las tablas se deben salvar tanto en formato Excel (.xls o .xlsx) para posibles ediciones futuras y en formato de texto delimitado por tabulaciones (.txt), una opción de exportación que ofrece Excel y que corresponde al formato aceptado por las herramientas CAPFITOGEN.

3.2. ¿Características de TesTable?

TesTable es la herramienta que verifica que las tablas que contienen la información introducida por el usuario cumplan con todas las condiciones para ser procesadas correctamente por parte de las herramientas. TesTable es capaz de analizar todos los tipos de tablas descritos en el numeral 3.1.

En TesTable un usuario introduce sus tablas, indica sus características principales y la herramienta genera un informe donde se pormenoriza todos los posibles errores o desajustes que la tabla pueda contener respecto al formato o estándar requerido. Opcionalmente el usuario puede solicitar corregir automáticamente los errores detectados y que se genere automáticamente una tabla sin errores que puedan generar errores al ser usadas posteriormente en las demás herramientas CAPFITOGEN. Dado que para corregir ciertos errores TesTable no puede inventar o asumir información si no tiene suficientes argumentos (como por ejemplo una coordenada introducida con errores de formato), en la tabla corregida el valor original errado de esa celda será reemplazado por el valor NA. Esto implica que según el tipo de errores detectados, la mejor opción puede ser una corrección manual y no la asignación automática de valores NA.

3.3 Utilización de la herramienta TesTable

Una vez cargados los scripts de parámetros y de la herramienta TesTable en RStudio o dicha herramienta ha sido seleccionada en el modo on server, será necesario definir una serie de parámetros para que la programación R funcione correctamente. Las tablas a ser analizadas y opcionalmente corregidas por TesTable en modo local deben localizarse en la carpeta Pasaporte.

Después de definir todos los parámetros y rutas que requiere TestTable, al hacer clic en el botón “Run” (modo local) o “Inicio” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que usualmente es corto para el caso de TesTable, se producirán los resultados que guardará en la misma carpeta “Pasaporte” o en la zona de archivos de usuario y resultados para el modo on server).

3.3.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

3.3.2 Parámetro: tiptable

Explicación: En este campo se indica el tipo de tabla en formato CAPFITOGEN que va a ser analizada y en su caso, corregida. Se debe escoger entre las siguientes opciones:

- Pasaporte (ver numeral 3.1.1)
- FuentesExternas (ver numeral 3.1.2)
- Fenotipo (ver numeral 3.1.3)
- Genotipo (ver numeral 3.1.5)
- NaturalezaVariables (ver numeral 3.1.4)

3.3.3 Parámetro: pasaporte

Explicación: Para modo local, escriba en este campo el nombre del archivo que contiene la tabla de que se quiere analizar (y/o corregir) en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo si el archivo se llama “tabla” debe escribir “tabla.txt”. Recuerde que este archivo debió ser guardado previamente en la carpeta “Pasaporte” que hace parte del conjunto de carpetas que componen CAPFITOGEN. Este campo debe rellenarse para cualquier tipo de tabla. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

3.3.4 Parámetro: access

Explicación: En este campo se debe indicar el número de entradas o accesiones que contiene la tabla que va a ser analizada. Este campo no debe tenerse en cuenta al analizar tablas tipo “NaturalezaVariables”.

3.3.5 Parámetro: geoqual

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto las cinco columnas adicionales COORQUAL, LOCALQUAL, SUITQUAL, TOTALQUAL y TOTALQUAL100. Sólo aplica para tablas tipo “Pasaporte”.

3.3.6 Parámetro: availab

Explicación: Seleccione esta opción si la tabla a ser analizada contiene una columna adicional (respecto a la tabla de pasaporte estándar CAPFITOGEN) denominada “AVAILAB”. Esta columna adicional indica la disponibilidad de cada entrada a ser seleccionada para integrar subcolecciones (herramientas ColNucleo y FIGS_R). Esta opción sólo aplica para tablas tipo ‘Pasaporte’.

3.3.7 Parámetro: fixthem

Explicación: Seleccione esta opción si desea corregir los errores detectados automáticamente. Si la selecciona, se guardará una tabla del tipo de la analizada en la carpeta “Pasaporte” con todos los errores corregidos que podrá ser usada por la herramienta correspondiente. En cualquier caso se generará un reporte de los problemas encontrados en la misma carpeta.

3.3.8 Parámetro: nmark

Explicación: Este campo sólo debe ser rellenado si la tabla a ser analizada es de tipo “Fenotipo” o “Genotipo”. Aquí se debe indicar el número de variables fenotípicas o genotípicas que debería tener la tabla a ser analizada (genotípica o fenotípica, según corresponda).

3.3.9 Parámetro: phenot

Explicación: Éste campo sólo aplica si la tabla a ser analizada es de tipo “NaturalezaVariables”. Escriba aquí el nombre de la tabla que contiene la información fenotípica a la que hace referencia la tabla “NaturalezaVariables” que desea analizar. Recuerde que ésta, como el resto de tablas, debe estar guardada en la carpeta ‘Pasaporte’. Al igual que para el parámetro del numeral 3.3.3, no olvide incluir la extensión del archivo, en este caso “.txt”.

3.4. Resultados de TesTable

En la carpeta “Pasaporte” (modo local) o en la zona de archivos de usuario y resultado (modo on server) encontrará una o dos tablas según haya elegido corregir automáticamente o no la tabla de entrada.

3.4.1 Tabla “TestTableAnalysis.txt”

Esta tabla siempre se producirá independientemente de la posible corrección que el usuario desee realizar. Contiene un informe sobre los posibles errores que pueda tener la tabla analizada. El informe viene dividido en diferentes apartados, todos con información en Español e Inglés (Fig. 16). Los apartados para una tabla de tipo “pasaporte” son:

3.4.1.1 Dimensión: Se prueba que el número de columnas es ajustada al tipo de tabla y sus características (si tiene los campos GEOQUAL y el campo “AVAILAB” por ejemplo).

3.4.1.2 Nombres columnas: Comprueba que los nombres de las columnas (en particular los de una tabla tipo “pasaporte”) estén correctos.

3.4.1.3 Filas fantasma: Comprueba que al final de la tabla no existan filas sin información. Estas “filas fantasma” se suelen producir cuando la tabla en archivo Excel se ha quedado con esas filas extras, fruto por ejemplo de una posible eliminación de filas. Al guardarse esa tabla Excel con filas fantasma en el formato .txt, éstas continúan apareciendo y pueden generar errores en la aplicación de casi cualquier herramientas.

3.4.1.4 Celdas en blanco: Como se mencionó anteriormente, las tablas de entrada de datos en las herramientas CAPFITOGEN no pueden contener celdas en blanco. En este apartado se indica si existen celdas en blanco y en qué posición de la tabla se hallan.

3.4.1.5 Requerimientos de cada campo: Aquí se verifica para cada campo si los datos cumplen con los estándares que requieren las herramientas CAPFITOGEN para poder ser analizados. Así, por ejemplo, si se trata de un campo que refiere a una fecha (COLLDATE), en caso de que alguna entrada no cumpla con el formato de fecha exigido, será reportada en esta sección. Los campos que se analizan son:

- ACCENUMB
- ORIGCTY
- Campos ADM (descripción de localidad mediante divisiones administrativas)
- Coordenadas en formato decimal (DECLATITUDE/DECLONGITUDE) y sexagesimal (LATITUDE/LONGITUDE)
- GEOREFMETH
- COLLDATE
- SAMPSTAT
- COLLSRC
- AVAILAB De manera similar para los otros tipos de tablas, el informe del archivo “TestTableAnalysis.txt” contendrá un apartado para cada campo de importancia que contenga la tabla objetivo.

3.4.2 Tabla “Pasaporte *Passport corr.txt*”

Aparecerá en caso de haber introducido una tabla de tipo “Pasaporte” y que se haya indicado la opción “fixthem”. Corresponde a la tabla de pasaporte corregida automáticamente por TesTable. Esta tabla sería apta para ser usada en las diferentes herramientas CAPFITOGEN que requieren este tipo de tabla (todas excepto ELCmapas).

3.4.3 Tabla “FuentesExt *ExtSources corr.txt*”

Aparecerá en caso de haber introducido una tabla de tipo “FuentesExternas” y que se haya indicado la opción “fixthem”. Corresponde a la tabla de fuentes externas de datos de presencia de especie corregida automáticamente por TesTable. Esta tabla sería apta para ser usada en la herramienta Representa (capítulo 8).

3.4.4 Tabla “Genotipo Genotype corr.txt”, “Fenotipo Phenotype corr.txt” o “NaturalezaVariables VariableType corr.txt”

Aparecerá en caso de haber introducido una tabla de tipo “Fenotipo”, “Genotipo” o “NaturalezaVariables” y que se haya indicado la opción “fixthem”. Corresponde a la tabla de fenotipo o la tabla de genotipo corregida automáticamente por TesTable. Esta tabla sería apta para ser usada en la herramienta DIVmapas (capítulo 9), donde según el tipo de mapa de diversidad que se desee obtener, se puede requerir este tipo de tablas sobre caracterización fenotípica y/o genotípica.

3.5. Consideración final

Es muy importante recordar que TesTable no mejorará la calidad de todos estos datos en caso de que se haya elegido la opción “fixthem” y se obtenga una tabla corregida. TesTable como mucho eliminará la información con formato incorrecto y la sustituirá con valores NA, excepto para el campo ACCENUMB (pues valores NA generarían errores). Para ACCENUMB, el error sólo puede ser debido a la presencia de posibles duplicados o a valores NA. En estos casos TesTable tomará el valor duplicado o NA y le adicionará el número de fila que ocupa el dato errado, con la esperanza de que el nuevo valor (numérico o alfanumérico) no se encuentre duplicado en este campo.



Taller Regional, Florianópolis, Brasil, Mayo 2014.



4 | Herramienta GEOQUAL

4.1. ¿Qué es la Evaluación de la Calidad de la Georreferenciación en datos de Pasaporte?

Es una metodología que determina el grado de certidumbre contenido en algunos descriptores de pasaporte cuya función es definir inequívocamente el lugar donde el germoplasma fue recolectado. De esta manera, GEOQUAL hace una evaluación de la calidad de los datos de descripción de la localidad y de las coordenadas indicadas como sitio de recolección.

En términos amplios, el concepto de calidad aplicado a datos ha recibido diversas definiciones. En el ámbito geográfico, la definición de calidad como “aptitud de uso” o potencialidad de uso ha sido ampliamente aceptada (Chrisman, 1983). Esto relaciona directamente calidad y posibilidad de uso de los datos. La incertidumbre asociada a todo tipo de datos, es una propiedad de quien obtiene o usa los datos más que de los datos mismos. Por ello calidad e incertidumbre tiene un grado de subjetividad variable, la cual puede reducirse hasta cierto punto mediante el uso de metodologías que realicen evaluaciones sobre bases lo más objetivas posibles. En todo caso, calidad e incertidumbre se asumen como medidas de riesgo entendido y riesgo asumido (Chapman, 2005).

La necesidad de evaluar la calidad de la georreferenciación de la información disponible sobre las presencias o incluso ausencias de entidades biológicas, es un asunto palpable en diferentes áreas como la ecología, análisis espacial y patrones de distribución de especies. Son numerosos los estudios que apuntan a que dicha calidad es un asunto crítico en metodologías como la modelación de la distribución de especies. La certeza de la ocurrencia de una especie en un sitio determinado es determinante para cualquier método que use datos de presencia o ausencia como materia prima. Los siguientes estudios hacen referencia a este aspecto: Foley *et al.*, 2009; Hill *et al.*, 2009; Otegui *et al.*, 2013.

Tener una estimación del grado de incertidumbre en la georreferenciación de sitios de presencia o ausencia de especies se convierte entonces en un aspecto clave antes de realizar cualquier análisis que utilice el aspecto espacial para el estudio de las distribuciones. Muchos análisis de este tipo desembocan en la toma de decisiones en aspectos prácticos de áreas como la conservación de la biodiversidad. Por tanto, introducir información de base confiable alimentando análisis adecuados producirá resultados confiables y decisiones acertadas y oportunas.

4.2. Antecedentes de la herramienta GEOQUAL

La metodología detrás de GEOQUAL acumula cuatro años de desarrollo, desde que surgió la necesidad de un estimador de confianza (o de riesgo, según se vea) de la georreferenciación del sitio de recolección, que usualmente quedan reflejados en los datos de pasaporte. La necesidad en particular surgió a finales de 2009, al momento de preparar los datos de pasaporte del Inventario Nacional de Recursos Fitogenéticos Españoles para ser caracterizados ecogeográficamente. Tener una idea de la calidad de la georreferenciación en datos de pasaporte fue entonces una prioridad a la hora de crear el Sistema de Información Ecogeográfica de los Recursos Fitogenéticos Españoles, denominado entonces como SIERFE.

En aquel sistema, se permitía seleccionar germoplasma a partir de la caracterización ambiental del sitio de recolección a través de un portal de internet. Al desarrollar GEOQUAL e incorporarlo en SIERFE, un estimador de la calidad permitía a los usuarios (solicitantes de germoplasma, como mejoradores, científicos o agricultores) determinar su exigencia en términos de calidad de la georreferenciación a la hora de seleccionar germoplasma por variables ecogeográficas. Esto representó un avance único en aquel momento en el desarrollo de sistemas de información y selección de germoplasma. Más de 45.000 entradas del Inventario Español fueron caracterizadas ecogeográficamente y cada una recibió un valor de calidad en una escala de 0 a 100.

GEOQUAL fue entonces desarrollado específicamente para las características de los datos de pasaporte del Inventario Nacional de Recursos Fitogenéticos Españoles y fue posible utilizando diversos programas, la mayor parte de tipo comercial como ArcGIS de ESRI.

En 2011, en el marco del proyecto *PGR secure* del Séptimo Programa Marco de la Unión Europea (<http://www.pgrsecure.org>), fue necesaria la depuración de cuatro bases de datos que contenían información de ocurrencia de variedades y especies silvestres relacionadas con cuatro géneros de interés agrícola en Europa (*Avena*, *Beta*, *Brassica* y *Medicago*). Más de 33.000 registros recibieron un valor GEOQUAL, lo cual permitió no considerar o mejorar la calidad de alrededor de 4.000 registros. A partir de entonces varios de los investigadores europeos en temas de agrobiodiversidad se interesaron por GEOQUAL, demandando el desarrollo de una herramienta de fácil manejo que permitiera aplicar su concepto sobre diferentes formatos de datos de presencia de especies.

Ya en 2012, cuando se inicia el desarrollo de las herramientas CAPFITOGEN y se decide sobre las aplicaciones prioritarias, GEOQUAL fue seleccionada. Se trataba de abordar el reto de crear una herramienta de evaluación de la calidad de datos de georreferenciación que fuera sencilla, con toda la información necesaria precargada, que no requiriera mayores conocimientos en Sistemas de Información Geográfica (SIG) de parte de los usuarios para poder aplicarla. Adicionalmente debería venir en una solución integrada (usando sólo un software de SIG), que usara como base el formato de descriptores de pasaporte definidos por la FAO y Bioversity International en 2012 y que fuera transferible a técnicos de programas nacionales. La herramienta GEOQUAL que se presenta aquí, es por tanto, la evolución de una idea original transformada en tecnología de fácil adopción, reuniendo factores de adaptabilidad apropiados a las condiciones y necesidades de varios programas nacionales de conservación de recursos fitogenéticos.

4.3. Características de GEOQUAL

La herramienta GEOQUAL se compone de cuatro parámetros, tres de ellos aportan aproximaciones diferentes sobre la calidad de la georreferenciación (COORQUAL, SUITQUAL y LOCALQUAL) y un parámetro final (TOTALQUAL) que resume los tres primeros. Los parámetros base se calculan en rangos que van de cero a veinte, siendo cero calidad nula y 20 máxima calidad. En algunas ocasiones, dependiendo de los datos de pasaporte disponibles, se puede obviar el cálculo de LOCALQUAL, como se explica más adelante. Adicionalmente se genera un parámetro que transforma los valores iniciales de TOTALQUAL (0 a 40 ó 0 a 60) a un rango de evaluación de 0 a 100, para facilidad de uso e interpretación del valor de evaluación (TOTALQUAL100). Es importante tener en cuenta que GEOQUAL opera sobre el formato de descriptores de pasaporte FAO-Bioversity publicado en el año 2012 con la adición de cuatro descriptores

de localidad (ADM1, ADM2, ADM3 y ADM4) que tienen correspondencia con diferentes figuras administrativas según cada país. Sin embargo, si los datos estuvieran en el formato FAO-IPGRI 2001, GEOQUAL podría también operar previo traslado de la información del formato 2001 al 2012 sin tener que añadir información para los nuevos campos que incluye la versión 2012, pero sí considerando la inclusión de los cuatro descriptores ADM.

La herramienta GEOQUAL incluye un modelo de tabla de descriptores de pasaporte basado en los descriptores multicultivo (MCPD) de FAO-Bioversity con la adición de los cuatro descriptores ADM en formato Excel. Este modelo se puede encontrar en la carpeta “Formatos_Formats”, subcarpetas “Español” y “Formatos GEOQUAL”, archivo “Tabla_pasaporte_modelo_FAO_Bioversity 2012 modificada.xls”, donde en color verde se señalan los descriptores indispensables para GEOQUAL y en color amarillo aquellos no indispensables pero si importantes. Los campos sin color en su encabezado no son tenidos en cuenta por GEOQUAL pero si se debe mantener su posición dentro de la tabla (como en el caso de los coloreados) para que GEOQUAL encuentre las variables que analiza en el sitio que espera encontrarlas. Como regla general, al llenar esta tabla, cuando no conozca la información que se le requiere, se debe escribir en el campo las letras NA, que normalmente indican “no aplica” pero que en el caso de GEOQUAL también indica que no hay información disponible.

4.3.1 Descripción de los parámetros base de GEOQUAL

4.3.1.1 Parámetro COORQUAL

Es un parámetro que determina la calidad intrínseca de las coordenadas contenidas en los datos de pasaporte. Se determina a través de cuatro sub--parámetros inicialmente:

- a) ERRORES:** Si las coordenadas en formato sexagesimal o decimal contienen valores fuera de marco posible dentro del sistema de coordenadas lat-long WGS84. Usa los descriptores LATITUDE, LONGITUDE, DECLATITUDE y DECLONGITUDE.
- b) PRECIS:** Aplica para coordenadas en formato sexagesimal que cumplen la codificación del listado de descriptores de pasaporte FAO-Bioversity 2012. Este sub-parámetro determina si las coordenadas fueron obtenidas con una precisión de segundos, minutos o grados. Usa los descriptores
- c) GEORBLE:** Es una valoración de la posibilidad de obtener coordenadas del sitio de recolección a partir de los datos de descripción de localidad disponibles.
- d) INTERTEMP:** Toma valores del descriptor COLLDATE y los interpreta en la medida de la posibilidad de uso de métodos de georreferenciación. Por ejemplo, para recolecciones ocurridas posteriormente al año 2000, se asume una alta probabilidad de uso de GPS, lo cual incrementaría la calidad de las coordenadas.
- e) GEOREFMETH:** Valora el sistema por el cual se han asignado coordenadas al sitio de recolección. GEOREFMETH corresponde a un campo de la tabla de pasaporte FAO/Bioversity 2012. Solamente se considerará este sub-parámetro cuando se dispone de valores para todas las entradas de este campo.

Cada sub-parámetro se evalúa en un rango de cero a tres, donde cero corresponde a calidad mínima y tres a calidad máxima. Las combinaciones de los valores de cada sub-parámetro generan el parámetro COORQUAL en un rango de cero a veinte.

4.3.1.2 Parámetro SUITQUAL

Este parámetro asigna un valor de calidad a las coordenadas de acuerdo a lo apropiado del sitio de recolección para el crecimiento de plantas, y diferenciando de la naturaleza de la entrada (silvestre o cultivada de acuerdo al descriptor SAMPSTAT). La información sobre las características del sitio de la recolección proviene de un mapa de uso de suelo (Global Land Cover 2000 o GLC2000). Se trata del mapa de cobertura global de libre acceso sobre uso de suelo más antiguo y que cuenta con una resolución apropiada (1 km). Las clases originales de este mapa se transforman en función de lo apropiado que resulta cada clase a la presencia de plantas cultivadas o silvestres, en una escala de 0 a 20.

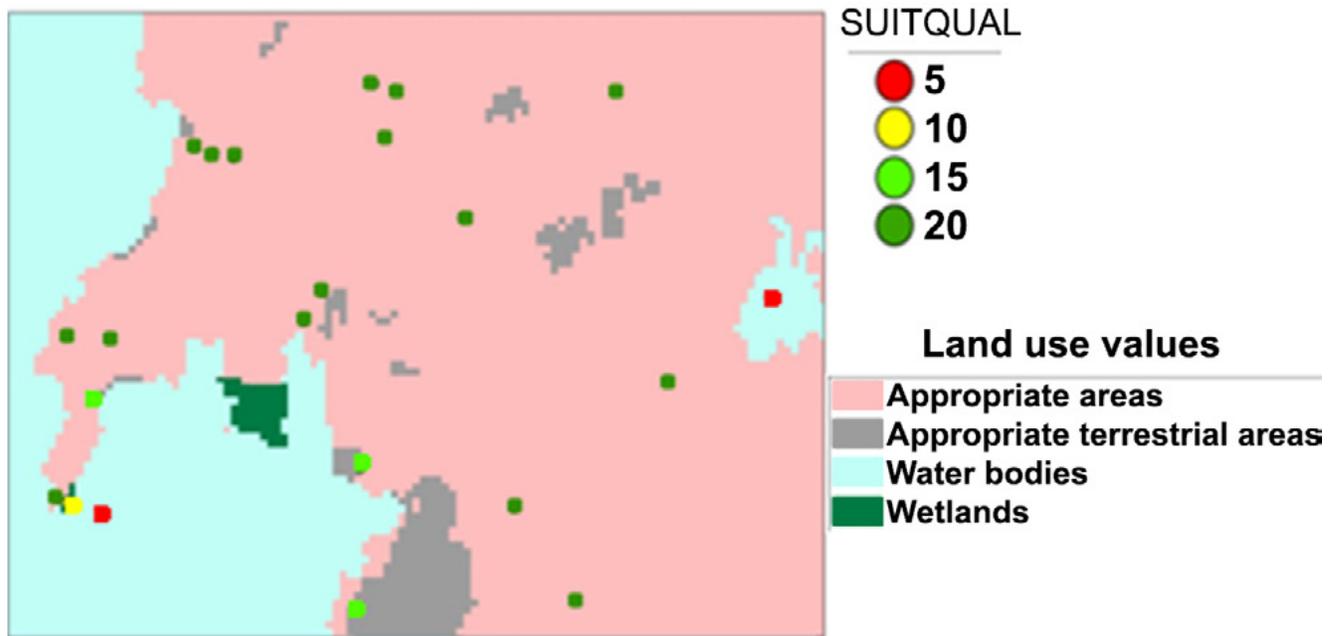


Figura 22. Ejemplo de la obtención de valores SUITQUAL de acuerdo a valores de uso de suelo interpretados.

4.3.1.3 Parámetro LOCALQUAL

LOCALQUAL es el resultado de la comparación de la descripción de la localidad donde se recolectó el germoplasma proveniente de los campos ORIGCTY, ADM1, ADM2, ADM3, ADM4 y COLLSITE, con los campos ISO, NAME1, NAME2, NAME3 y NAME4 de la base de datos “Global Administrative Areas” (GADM) v2.0, extraídos por medio de las coordenadas provistas en DECLATITUDE y DECLONGITUDE (o a través de la transformación a formato decimal de LATITUDE y LONGITUDE). A diferencia del proceso que realiza “Check Coordinates” (comprobar coordenadas) incluido en DIVA-GIS donde la comparación es absoluta (los términos deben coincidir carácter por carácter para adjudicar coincidencia), GEOQUAL utiliza la distancia generalizada de Levenshtein a través de la función “agrep” del paquete base de R, que toma en cuenta el número de inserciones, eliminaciones o cambios de caracteres entre las dos cadenas que se comparan. Así, permitiendo un cierto número de este tipo de cambios, la función “agrep” es capaz de identificar concordancias, aunque ocurran errores de tipografía o por el efecto de aquellos caracteres alfabéticos propios de ciertos idiomas que en ocasiones no se codifican correctamente (caso de la ñ y los acentos en el idioma castellano).

Para más seguridad, LOCALQUAL compara también con los campos VARNAME1, VARNAME2, VARNAME3 y VARNAME4 incluidos en GADM, que son variantes del nombre oficial de la unidad administrativa y que pueden ser usados por los curadores a la hora de registrar el germoplasma en sus bases de datos de pasaporte. Finalmente, LOCALQUAL considera la serie de comparaciones positivas entre los diferentes emparejamientos (ORIGCTY con ISO, ADM1 con NAME1, etc.) para calcular un valor en un rango de cero a veinte.

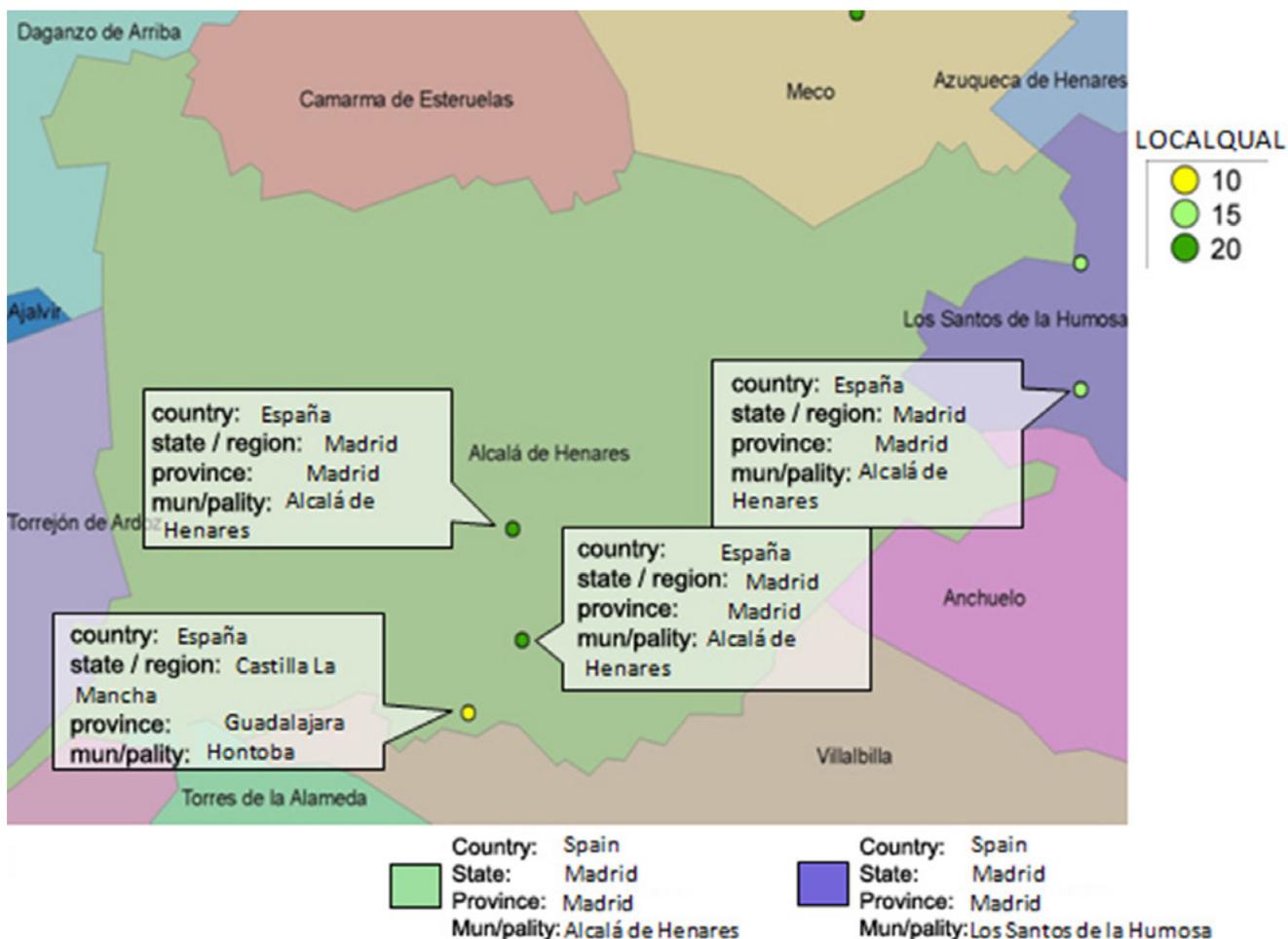


Figura 23. Ejemplo de la obtención de valores LOCALQUAL de acuerdo a la comparación de niveles administrativos de los datos aportados por el usuario y los extraídos por las coordenadas desde GADM.

4.3.2 Descripción de los parámetros TOTALQUAL y TOTALQUAL100

El parámetro final de resumen TOTALQUAL es simplemente la suma de los valores de COORDQUAL, SUITQUAL y LOCALQUAL. Por los rangos posibles de valores de estos tres parámetros, TOTALQUAL puede tomar valores entre 0 y 60. Sin embargo, para facilidad de interpretación y análisis de los resultados de GEOQUAL, se calcula también el parámetro TOTALQUAL100 que es una transformación de TOTALQUAL a un rango de valores de 0 a 100, donde 0 es calidad nula (incluyendo la falta de coordenadas) y 100 representa una calidad teórica óptima.

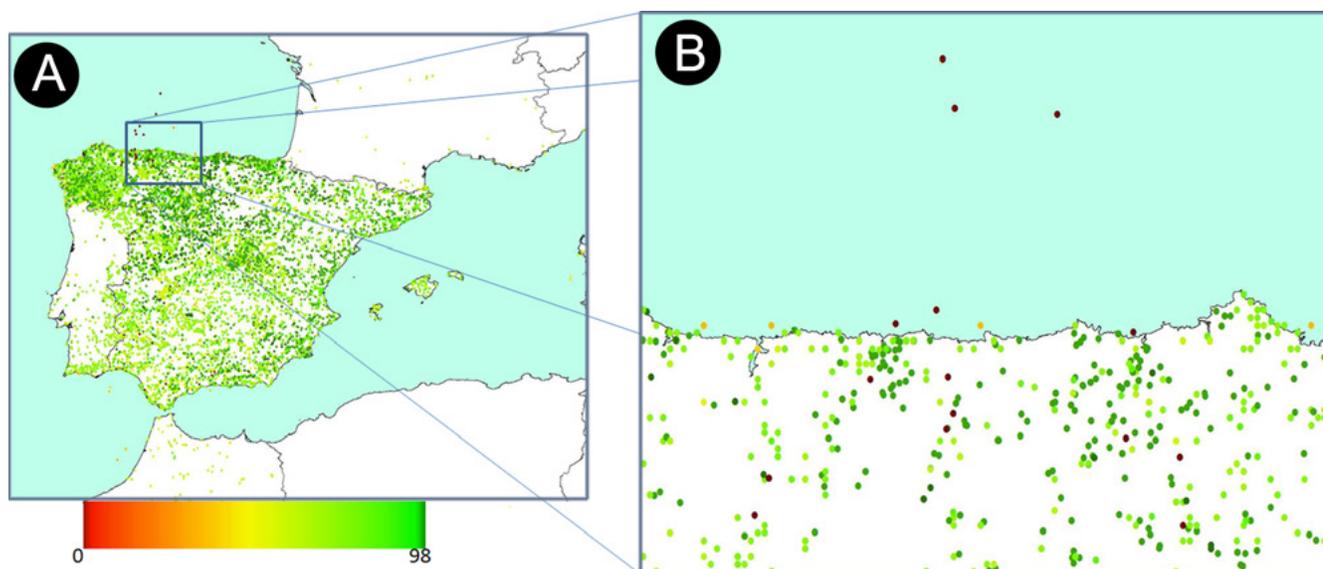


Figura 24. Resultados de la aplicación de GEOQUAL en el Inventario Nacional Español de Recursos Fitogenéticos. Se muestran los valores alcanzados de TOTALQUAL100. A) Cerca de 45,000 accesiones del inventario evaluados por GEOQUAL, con valores de TOTALQUAL100 entre 0 y 98. B) Acercamiento a zona costera donde se puede visualizar como la valor TOTALQUAL descende a medida que el punto se adentra en el mar.

4.3.3 Determinación de umbrales de calidad

GEOQUAL ha sido pensada desde su implementación inicial, como una metodología altamente objetiva, en la que el usuario interviene mínimamente en la obtención del valor definitivo. Sin embargo, toda determinación de calidad conlleva componentes subjetivos y GEOQUAL no es la excepción.

Por ejemplo, hay cierta subjetividad cuando se aplica a ciertas categorías de uso de suelo unos valores de aptitud para el crecimiento de plantas. También la definición de a partir de qué valores se puede considerar alta o baja calidad es una cuestión subjetiva que tiene que ver con el observador más que con la técnica. El umbral sobre el cual se considera que una entrada está correctamente georreferenciada usando valores GEOQUAL debe ser definida por el usuario de los datos en base a sus expectativas y necesidades. Se puede plantear el uso de diferentes umbrales dependiendo del uso que se le dará a los datos, del tipo de estudio al que serán sometidos, y de la precisión y exactitud de la información del resto de las fuentes en un determinado estudio. Es recomendable ver la distribución de valores TOTALQUAL100 del conjunto de entradas y así conocer de antemano que de aplicar un umbral muy exigente (cercano a 100) o poco exigente (por debajo de 50) se seleccionarían muchas o pocas entradas.

4.4. Utilización de la herramienta GEOQUAL

Una vez cargados los scripts de parámetros y de la herramienta GEOQUAL en RStudio o dicha herramienta ha sido seleccionada en el modo on server, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

Después de definir todos los parámetros y rutas que requiere GEOQUAL, al hacer clic en el botón “Run” (modo local) o “Inicio” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar por la introducción de algunos parámetros de resolución, tipo de análisis, tamaño de datos procesados o configuración de hardware del computador, GEOQUAL producirá resultados que guardará donde se le haya indicado (parámetro 4.4.1.5 para el modo local o en la zona de archivos de usuario y resultados para el modo on server).

4.4.1 Parámetros iniciales definidos por el usuario

4.4.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

4.4.1.2 Parámetro: pasaporte

Explicación: Para modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama “tabla” debe escribir: “tabla.txt”. Recuerde que este archivo debe ser guardado previamente en la carpeta “Pasaporte” que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

4.4.1.3 Parámetro: precision

Explicación: Elija utilizar mapas de alta o baja precisión para determinar si las coordenadas de un sitio de recolección caen en el mar y a cuanta distancia lo hace. Alta resolución puede ralentizar un poco el proceso en bases de datos muy grandes (por encima de 15.000 entradas con coordenadas).

4.4.1.4 Parámetro: local

Explicación: Indique si desea usar el parámetro LOCALQUAL para la evaluación de la calidad de la georreferenciación. LOCALQUAL es un parámetro de comparación entre localidad descrita y extraída por SIG. Si sus datos no contienen

ninguna descripción de localidad o la descripción está contenida por completo en el campo COLLSITE, NO es conveniente usar esta opción.

4.4.1.5 Parámetro: resultados (sólo para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

4.5. Resultados de GEOQUAL

En la ruta y carpeta que haya determinado en “resultados” (parámetro 4.4.1.6) en el modo local o en la zona de archivos de usuario y resultados del modo on server, encontrará tres tablas, un mapa de puntos de tipo vectorial (shapefile) y un archivo con la lista de parámetros usados en la ejecución de GEOQUAL (“Parameters_Geoqual.txt”). En el modo on server, también se incluye un reporte del avance de ejecución en un archivo llamado “Error_process_info.txt”.

4.5.1 Tablas

Las tablas producidas por GEOQUAL están en formato de texto delimitado por tabulaciones con dos extensiones (.txt y .xls). Al intentar abrir los archivos .txt, el sistema operativo lo intentará con el software de edición de texto asociado a esta extensión. En el caso del archivo .xls, si el usuario tiene instalado MS Excel®, este software lo intentará abrir, pero como en realidad se trata de un archivo de texto, aparecerá un mensaje de advertencia (ver Fig. 25) el cual puede ser ignorado y finalmente poder visualizar la tabla.

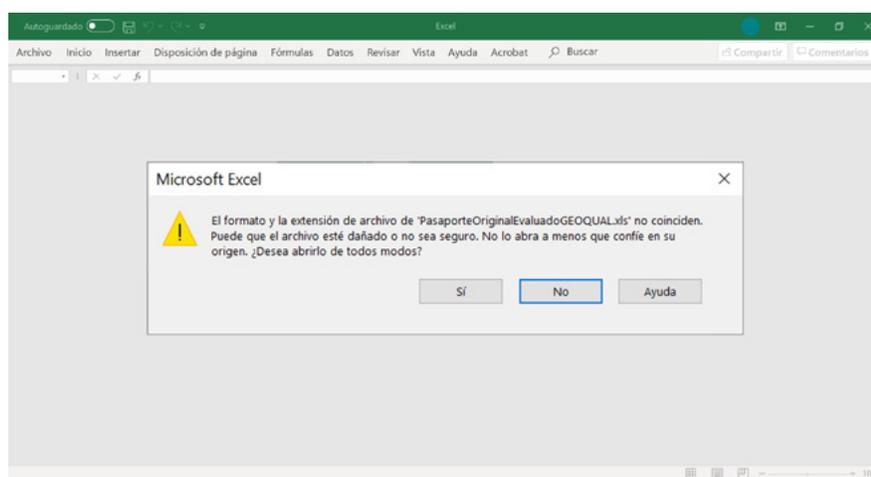


Figura 25. Mensaje de advertencia de MS Excel® sobre el archivo .xls (que en realidad corresponde a un archivo de texto delimitado por tabulaciones) que se va a abrir. En este punto el usuario debe hacer click en la opción “Sí”. Se abrirá correctamente la tabla.

4.5.1.1 “PasaporteOriginalEvaluadoGEOQUAL.txt” y .xls: Es la tabla de pasaporte en el formato sugerido que originalmente se usó para el análisis, con la adición de cinco columnas con los valores obtenidos para los parámetros: SUITQUAL, LOCALQUAL, COORQUAL, TOTALQUAL y TOTALQUAL100.

4.5.1.2 “tabla_de_analisisGEOQUAL.txt” y .xls: Esta tabla contiene también todas las columnas de la tabla de pasaportes originalmente introducidas para su análisis, aunque en este caso sólo se incluyen las entradas con coordenadas. Sin embargo, lo más importante de esta tabla es que en ella se han añadido todas las columnas correspondientes a extracciones, interpretaciones o subparámetros que han sido necesarios para poder calcular los valores de los parámetros de GEOQUAL. La lista de variables adicionales incluidas en esta tabla y su explicación se encuentra en el Anexo 18.4.

4.5.1.3 “StatsSummaryGEOQUAL.txt” y .xls: Es la tabla donde se muestran promedios, moda, valores máximos, mínimos y desviación estándar para SUITQUAL, LOCALQUAL, COORQUAL, TOTALQUAL y TOTALQUAL100 para los datos de pasaporte evaluados.

4.5.2 Mapas

4.5.2.1 Mapa de puntos en formato vectorial tipo “shapefile”. Este mapa viene acompañado de una tabla que incluye los valores de los parámetros de evaluación GEOQUAL, de tal manera que en DIVA-GIS se pueden mostrar los puntos con diferentes grados de color de acuerdo a su puntaje (calidad). Un “shapefile” se compone de hasta 6 archivos de igual nombre pero diferente extensión. En el caso de GEOQUAL, el shapefile sólo está compuesto por tres extensiones (.shp, .shx, y .dbf) y tiene el nombre de ShapefilePuntosGEOQUAL.

4.5.2.2 Mapa de puntos en formato Google Earth. Este mapa corresponde al archivo mapa_puntos_google.kml y si se tiene instalado el programa Google Earth, con sólo hacer doble clic sobre su nombre en el explorador de Windows, el mapa de puntos (en forma de tachuelas o chinchetas) se abrirá en dicho programa, ubicando los sitios de recolección sobre las imágenes de satélite. Al hacer clic sobre las chinchetas se abrirá una pequeña ventana mostrando el valor TOTALQUAL100 de cada entrada.

4.6. Referencias

Chapman, A.D. 2005. Principles of data quality, version 1.0. Report of the Global Biodiversity Information Facility, Copenhagen.

Chrisman, N.R. 1983. The role of quality information in the long-term functioning of a GIS. Proceedings of AUTOCART06, 2: 303-321. Falls Church, VA: ASPRS.

FAO, IPGRI. 2001. Lista de descriptores de pasaporte para cultivos múltiples desarrollada por la FAO y el IPGRI.

FAO, BIOVERSITY. 2015. FAO/Bioversity multi-crop Passport descriptors V.2. URL: <https://bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v21-mcpd-v21/>

Foley, D.H., Wilkerson, R.C., Rueda, L.M. 2009. Importance of the “what,” “when,” and “where” of mosquito collection events. *J Med Entomol.* 2009 Jul;46(4):717-22.

Hill, A.W., Guralnick, R., Flemons, P., Beaman, R., Wieczorek, J., Ranipeta, A., Chavan, V., Remsen, D. 2009. Location, location, location: utilizing pipelines and services to more effectively georeference the world’s biodiversity data. *BMC Bioinformatics.* 2009 Nov 10;10 Suppl 14:S3. doi: 10.1186/1471-2105-10-S14-S3.

Otegui, J., Ariño, A.H., Encinas, M.A., Pando, F. 2013. Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF). *PLoS One.* 2013;8(1), e55144. doi: 10.1371/journal.pone.0055144.

Soberón, J., Peterson, T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Phil. Trans. R. Soc. Lond. B.* 359, 689-698.



Taller Nacional, Madrid, España, Noviembre 2013.



5 | Herramienta SelecVar

5.1. ¿Por qué seleccionar variables de tipo abiótico que determinen los diferentes escenarios adaptativos de una especie?

Esta es una pregunta que ya han respondido la mayoría de las personas que han usado la herramienta ELC mapas (capítulo 6) y que debería plantearse cualquier persona que considere realizar modelos de distribución de especies como los que realiza la herramienta Modela (capítulo 14). Sencillamente porque tratar de determinar los diferentes escenarios adaptativos que ocurren en un territorio para una especie particular usando todas las variables disponibles (en el caso de las herramientas CAPFITOGEN 105 variables) no sólo no tiene sentido biológico sino que también ralentiza los análisis multivariados (que realiza ELC mapas) o modelizaciones (que realiza Modela) y altera sus resultados.

La Fig 20 ilustra el efecto de hacer o no selección de variables ecogeográficas para la obtención de un mapa de caracterización ecogeográfica del territorio (mapa ELC, siglas en inglés) para *Vicia sativa* en España. Es evidente que la división del territorio resulta diferente cuando se usan todas las variables disponibles para el componente bioclimático o un conjunto específico de ellas.

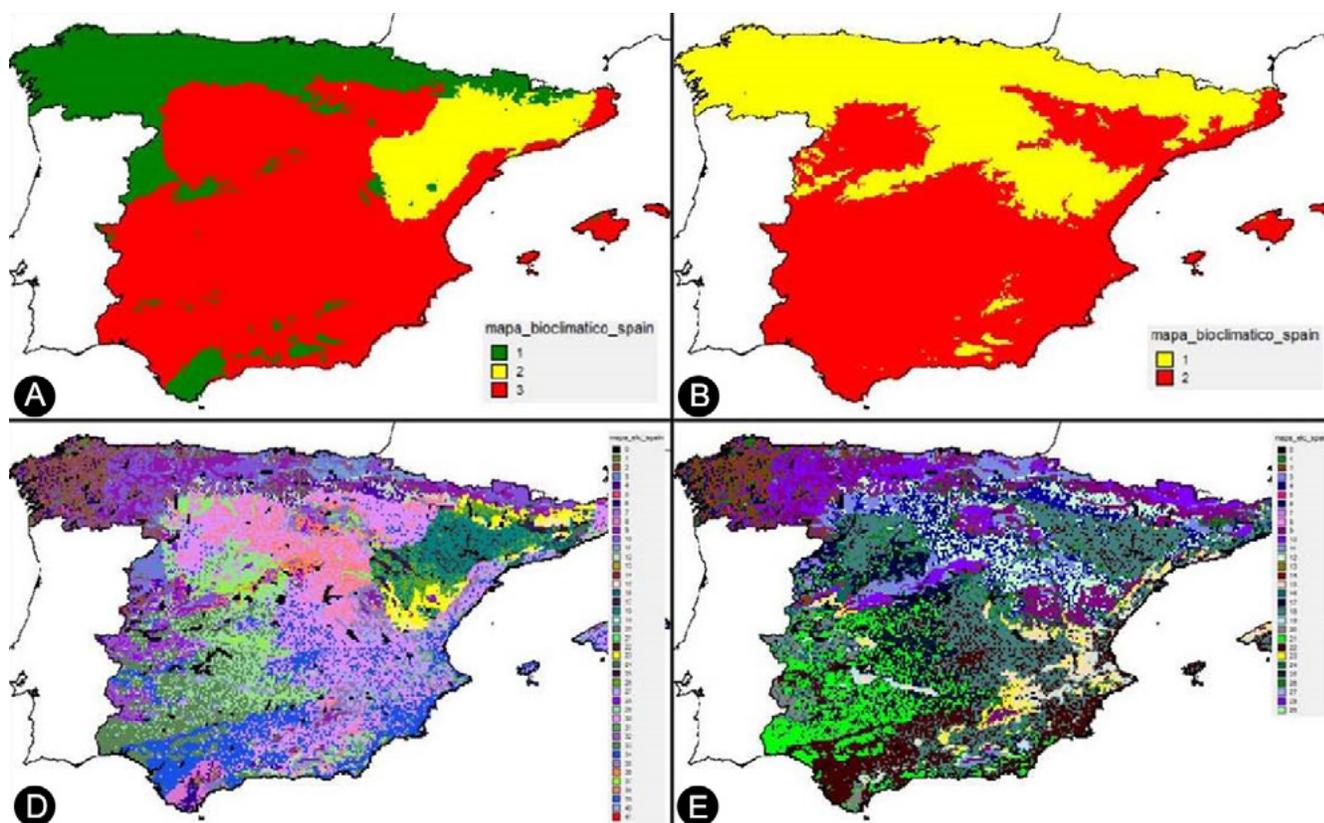


Figura 26. Productos de la herramienta ELC mapas (capítulo 6) con y sin selección previa de variables. A) Mapa de zonas bioclimáticas con selección de variables (3 variables). B) Mapa de zonas bioclimáticas usando todas las variables disponibles (67 variables). C) Mapa ELC con selección de variables bioclimáticas. D) Mapa ELC usando todas las variables bioclimáticas disponibles.

La selección o no de variables ecogeográficas que se introducen en los modelos de distribución de especies también puede afectar seriamente los resultados de este tipo de análisis (Austin y Van Niel, 2010). En un reciente estudio se encontró que usar un conjunto u otro de variables abióticas predictoras en la modelización de la distribución de especies vegetales pueden tener un efecto significativo en los resultados (las distribuciones potenciales), los patrones espaciales y las predicciones (Pliscoff *et al.*, 2014). El efecto de la escogencia de variables predictoras puede hacer variar hasta en un 2000% la predicción de ocupación de área y un 50% los estimadores de riesgo de extinción de una especie vegetal.

5.2. La selección de variables para mapas ELC

Los mapas ELC pueden ser de dos tipos, generalistas y específicos, como se describe en el apartado 6.4. Así, mientras que para los mapas generalistas la selección previa de variables es un proceso simple que puede limitarse a la consulta de expertos en cuanto a las variables que pueden afectar la distribución de la vegetación en sentido amplio, para generar un mapa especialista, es deseable no sólo contar con conocimiento experto sobre las variables más relevantes en términos de adaptación abiótica, sino también realizar una escrupulosa selección de variables desde un punto de vista estadístico. Para ello se busca que el conjunto de variables seleccionadas cumplan, en la medida de lo posible, con las condiciones descritas en los dos siguientes apartados. Estos requerimientos pueden fácilmente trasladarse a las variables que se introduzcan como predictoras en un modelo de distribución de especies.

5.2.1. Heterogeneidad y mínima redundancia

Las variables que alimentan un análisis de agrupamientos (para un mapa ELC) o un modelo de distribución deben ser las mínimas necesarias para explicar la ocurrencia de diferentes escenarios adaptativos a los cuales responde la especie objetivo o la presencia/ausencia de poblaciones respectivamente. Un número reducido de variables permite procesos y análisis más rápidos y eficientes a la vez que facilita la interpretación de los resultados. Es posible descartar variables que no hacen aportes al análisis o variables que aportan la misma información que otras.

Cuando se extrae la información ecogeográfica representada en diferentes variables para un conjunto de sitios de recolección o poblaciones de una especie y una o varias variables presentan el mismo valor para todos los sitios de recolección o poblaciones, estas pueden ser descartadas ya que variables homogéneas no son útiles para crear grupos adaptativos o generar distribuciones potenciales.

A través de análisis de correlaciones bivariadas o BCA (siglas en inglés) o determinación de la correlación por cada par de variables del conjunto inicialmente considerado, es posible determinar que variables aportan información idéntica o muy similar para dejar una sola de ellas como representante de las demás y no introducir redundancia en el proceso de obtención de mapas ELC o modelización de distribuciones.

Otro análisis que permite detectar variables que aportan información similar es el Análisis de Componentes Principales (PCA en sus siglas en inglés). Este análisis originalmente ha sido diseñado para reducir la dimensionalidad de

conjuntos de datos compuestos por numerosas variables. Para ello usa transformaciones ortogonales, consiguiendo crear nuevas variables a partir de la información aportada por las variables originales, pero que a diferencia de éstas últimas, los componentes principales (nuevas variables) no son correlacionadas linealmente.

Dado que el PCA ofrece también información sobre la variabilidad de los datos originales explicada por cada componente (valores propios) y sobre el peso que cada variable original tiene sobre cada componente (vectores propios), este tipo de análisis es usado para exploración de datos.

Para selección de variables, una posible utilización de PCA es considerar un número reducido de componentes principales (que la suma de su representación de la varianza original sea mayor del 70%, por ejemplo) y determinar cuáles son las variables originales con mayor peso en valores absolutos (sin considerar su sentido positivo o negativo). La selección de una o dos variables originales con los más altos pesos en su componente principal, de unos pocos componentes principales, ofrecería en principio un conjunto reducido de variables de alto poder discriminatorio y poca redundancia (correlación lineal) entre sí (similar al método B4 descrito por King y Jackson, 1999).

5.2.2 Contribución de la variable

Se trata de determinar la contribución que hace cada variable considerada en la determinación de grupos adaptativos o en la explicación de la distribución de la especie.

En primer lugar, si lo que se busca es determinar en un territorio los diferentes escenarios de adaptación posibles para una especie objetivo, entonces es de gran utilidad detectar las variables abióticas que más influyan en la conformación de grupos de poblaciones recolectadas o detectadas en sitios con diferentes características adaptativas. Para ello se puede usar diferentes aproximaciones. SelecVar usa la técnica Random forest (clasificación).

Random Forest (RF) es una potente metodología de clasificación, que además de producir clasificaciones (agrupamientos) con alta precisión dentro de ambientes caracterizados por complejas interacciones entre variables, hace también posible la determinación de la importancia de las variables que intervienen en la clasificación (Cutler *et al.*, 2007). Esta técnica de determinación de importancia de variable ha sido implementada en la función “importance” en el paquete de R llamado “randomForest” y como resultado de su aplicación se genera una tabla con tantas filas como variables analizadas y dos columnas con medidas de importancia como el decrecimiento medio de la precisión (mean decrease accuracy) o decrecimiento medio del índice Gini (mean decrease Gini). De acuerdo con Cutler *et al.* (2007), el decrecimiento medio de la precisión para una variable es la diferencia normalizada de la precisión de clasificación del conjunto de datos cuando se incluye la variable en cuestión y cuando los valores de esa variable han sido aleatoriamente permutados. Por tanto las variables de mayor importancia en la clasificación son aquellas de mayor valor de decrecimiento medio de la precisión. En el estudio sobre la importancia de la selección de variables climáticas en la obtención de predicciones fiables de modelos de distribución de especies, se utilizó RF para determinar conjuntos de seis variables con los mayores valores de decrecimiento medio de la precisión.

5.3. Qué hace SelecVar y cómo usar sus análisis

SelecVar está diseñada para ofrecer a los usuarios suficientes argumentos objetivos para realizar una selección de variables, útil para la generación de mapas ELC (ver capítulos herramientas “ELC mapas” o “Tzones”) o para modelos de distribución de especies (ver capítulos herramienta “Modela”). Para ello realiza a petición del usuario análisis de selección de variables por Random Forest (RF), correlación bivariada (BCA) o análisis de componentes principales (PCA). Los dos primeros análisis los realiza secuencialmente (primero RF, luego BCA), para cada componente ecogeográfico (variables bioclimáticas, geofísicas y edáficas) de manera independiente.

La selección de variables para mapas ELC o modelos de distribución de especies, puede ser idealmente un proceso objetivo-subjetivo, donde la parte objetiva la puede aportar la herramienta SelecVar y la parte subjetiva, por ejemplo, una encuesta sobre importancia de las variables entre expertos conocedores de aspectos adaptativos de la especie objetivo. También recopilaciones bibliográficas sobre aspectos adaptativos de la especie pueden ser de utilidad dentro del proceso subjetivo.

La utilización de SelecVar para la determinación final del listado de variables a ser incluidas en ELC mapas o Modela puede variar de acuerdo a las necesidades o punto de vista que tenga el usuario sobre los análisis que realiza. SelecVar para la versión CAPFITOGEN 2, permitía al usuario realizar todas las posibles combinaciones de análisis entre PCA, BCA, RF y otro procedimiento denominado clusvarsel, el cual fue descartado en CAPFITOGEN 3 (ambos modos de utilización) por la alta demanda de recursos computacionales para su ejecución y los numerosos errores que se producían. Así, SelecVar fue modificada de tal manera que en la versión 3 realiza siempre los siguientes pasos:

1. RF: seleccionando por el índice mean decrease accuracy, una fracción del total de las variables que el usuario inicialmente indica a la herramienta para ser evaluadas. El usuario indica el tamaño de la fracción de variables a ser seleccionadas. SelecVar hace una lista de las variables de la más importante a la menos, y selecciona la fracción indicada por el usuario de variables más importante.
2. BCA: el conjunto de variables seleccionadas por RF se someten a una prueba de correlación bivariada. El proceso de descarte de variables por correlación (redundancia) inicia comparando la variable seleccionada de menor importancia por RF respecto al resto de variables de mayor importancia. Si esta variable está correlacionada al menos con una de las de mayor importancia, teniendo como referencia un valor de coeficiente de correlación y un P valor (que proporciona también el usuario), es descartada. Si la variable no se correlaciona con ninguna otra, se selecciona. El proceso continúa con la siguiente variable de menor importancia por RF, la cual se compara con las demás de mayor importancia y se descarta en caso de correlación con cualquier otra. Así se continúa descartando o seleccionando las variables importantes por RF, hasta que el proceso se detiene cuando se alcanza la fracción (que indica el usuario) de variables a ser seleccionadas por correlaciones bivariadas.
3. PCA: Finalmente la totalidad de variables que el usuario ha querido incluir en el análisis, se someten a un análisis de componentes principales. En este análisis la herramienta no realiza descarte de variables, sino que ofrece al usuario información sobre los vectores y valores propios (eigenvectors y eigenvalues), y adicionalmente una tabla donde se indica la posición de cada variable seleccionada por RF-BCA respecto a los vectores propios para los componentes principales que el usuario desee.

5.4. Utilización de la herramienta SelecVar

Una vez cargados los scripts de parámetros y de la herramienta SelecVar en RStudio o dicha herramienta ha sido seleccionada en el modo on server, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

Después de definir todos los parámetros y rutas que requiere SelecVar, al hacer clic en el botón “Run” (modo local) o “Inicio” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar por la introducción de algunos parámetros de resolución, tipo de análisis, tamaño de datos procesados o configuración de hardware del computador, SelecVar producirá resultados que guardará donde se le haya indicado (parámetro “resultados” para el modo local o en la zona de archivos de usuario y resultados para el modo on server).

5.4.1 Parámetros iniciales definidos por el usuario

5.4.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

5.4.1.2 Parámetro: país

Explicación: Seleccione el país/región en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente. También puede utilizar la herramienta rLayer para producir sus propios marcos de trabajo y seleccionarlos aquí.

5.4.1.3 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido

previamente y esté en la zona de archivos de usuario y resultados.

5.4.1.4 Parámetro: **disdup**

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas. La determinación de la distancia depende de condiciones biológicas (flujo de genes) y espaciales (tamaños promedios de las poblaciones). Se trata de un parámetro específico para la especie objetivo y muchas veces será necesario consultar el concepto de un experto.

5.4.1.5 Parámetro: **geoqual**

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

5.4.1.6 Parámetro: **totalqual**

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

5.4.1.7 Parámetro: **buffy**

Explicación: Marque esta opción si desea que la extracción de información ecogeográfica se haga sobre un área circular alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección. Esta opción es muy útil, por ejemplo, cuando la mayoría de los sitios de recolección corresponden a mercados y no directamente las fincas de los agricultores, o cuando existe bastante incertidumbre sobre la real ubicación de los sitios de recolección a pesar de que GEOQUAL pueda otorgar alta calidad de georreferenciación. La explicación de esta característica se encuentra explicada en el capítulo dedicado a la herramienta ECOGEO.

5.4.1.8 Parámetro: **tamp**

Explicación: Solo aplica si buffy ha sido señalado como TRUE (✓ en modo on server). Especifique el radio (en kilómetros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor.

La extracción se hará sobre aquellas celdas cuyo centroide esté dentro del área circular.

5.4.1.9 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución, pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes.

5.4.1.10 Parámetro: bioclimv

Explicación: Liste (modo local) o seleccione (modo on server) las variables bioclimáticas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: "#Lista completa de variables bioclimáticas), cópielos y péguelos separados por un signo de punto y coma (;). También encontrará una línea bloqueada la selección de las 19 variables bioclim, listas para usar, simplemente eliminando el símbolo # inicial. En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro bioclimv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo "Variables names - Nombres de variables.xlsx" descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

5.4.1.11 Parámetro: edaphv

Explicación: Liste (modo local) o seleccione (modo on server) las variables edáficas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: "#Lista completa de variables edáficas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro edaphv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo "Variables names - Nombres de variables.xlsx" descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

5.4.1.12 Parámetro: geophysv

Explicación: Liste (modo local) o seleccione (modo on server) las variables geofísicas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: "#Lista completa de variables geofísicas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable

aparecerá en la caja frente al parámetro geophysv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

5.4.1.13 Parámetro: latitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la latitud (Y) como una variable geofísica más a ser analizada.

5.4.1.14 Parámetro: longitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la longitud (X) como una variable geofísica más a ser analizada.

5.4.1.15 Parámetro: percenRF

Explicación: Indique aquí un valor entre 0 y 1. El valor percenRF indica el porcentaje de variables que serán seleccionadas por Random Forest (clasificación) y seguirán en el proceso de selección. Sin embargo el porcentaje no se expresa con un valor entre 0 y 100 sino entre 0 y 1. Por ejemplo, si se desea seleccionar el 60% de las variables por Random Forest para que luego sigan siendo seleccionadas por correlaciones bivariadas, el valor en percenRF debe ser 0.6. Así, 1-percenRF serán las variables descartadas por ser las menos importantes de acuerdo al índice mean decrease accuracy.

5.4.1.16 Parámetro: percenCorr

Explicación: Indique aquí un valor entre 0 y 1. El valor percenCorr indica el porcentaje de variables que serán seleccionadas por BCA (análisis de correlación bivariada) en el proceso paso a paso descrito anteriormente. Sin embargo, el porcentaje no se expresa con un valor entre 0 y 100 sino entre 0 y 1. Por ejemplo, si se desea seleccionar el 30% de las variables importantes por Random Forest, el valor en percenCorr debe ser 0.3. Así, 1-percenCorr serán las variables descartadas por ser redundantes por sus correlaciones y menos importantes por Random Forest. percenCorr puede ser alterado en base al número mínimo de variables que el usuario defina en el parámetro nminvar (5.4.1.19).

5.4.1.17 Parámetro: CorrValue

Explicación: Indique para este parámetro un valor entre 0 y 1, el cual será utilizado para definir el umbral de lo que será aceptado como una alta correlación de acuerdo al coeficiente de correlación. Así, indicar en CorrValue un valor de 0.5, hará que valores de correlación entre 0.5 a 1 y -0.5 y -1 sean asumidos como alta correlación. Entonces inicialmente un par de variables con un valor de correlación de -0.85 serían redundantes y la de menor importancia sería descartada. Sin embargo, hay que tener en cuenta que SelecVar no sólo tendrá en cuenta el valor de correlación sino también su significancia para el descarte de las variables.

5.4.1.18 Parámetro: pValue

Explicación: Indique un valor de 0.001 o 0.05. pValue define el valor de umbral de significancia para las correlaciones bivariadas. Usualmente este valor fluctúa entre 0.001 y 0.05. La correlación se asumirá como significativa cuando el valor de P sea menor que el indicado en el parámetro pValue.

5.4.1.19 Parámetro: nminvar

Explicación: Indique el número mínimo de variables que desea seleccionar por componente. Tenga en cuenta que nminvar debe ser un número mayor que el número de variables seleccionadas en bioclimv, edaphv o geophysv.

5.4.1.20 Parámetro: ecogeopcaxe

Explicación: Dado que SelecVar realizará un PCA por cada componente ecogeográfico (bioclimático, geofísico y edáfico), en ecogeopcaxe se le indica a la herramienta el número de componentes principales que se retendrán, o el número de componentes para el cual se ofrecen resultados como puede ser los vectores propios. Por ello, ecogeopcaxe no puede ser un número mayor al número de componentes que se generarán por parte del componente con menos variables. Por ejemplo, si el número de componentes principales que se generaron fueron bioclimáticos=5, geofísicos=3 y edáficos=6, ecogeopcaxe debe ser menor o igual a tres (3). Recuerde que el número de componentes principales es igual a 1-número de variables analizadas.

5.4.1.21 Parámetro: resultados (sólo para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

5.4.2 Resultados de SelecVar

Después de definir todos los parámetros y rutas (para modo local) que requiere SelecVar, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, SelecVar guardará los resultados en la ruta y carpeta que se haya determinado en “resultados” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server. Allí encontrará carpetas para los componentes bioclimático, edáfico y geofísico (“BioclimaticVariables_pais”, “EdaphicVariables_pais”, y “GeophysicVariables_pais” respectivamente), el archivo de texto “Parametros.Parameters.SelecVar.txt” donde se especifican los parámetros usados y tres archivos con extensión .xls que corresponden realmente a archivos de texto delimitado por tabulaciones pero que se puede abrir directamente en Excel® (ver Fig. 25), los cuales contienen las listas de variables finalmente seleccionadas. Dentro de las carpetas para cada componente, se encuentran unas nuevas carpetas para cada análisis realizado (RF, BCA y PCA). Para cada componente se generarán las mismas tablas, listas y gráficas resultado, por tanto se describe a continuación lo que el

usuario encontrará dentro de cada carpeta de análisis.

5.4.2.1 Correlaciones bivariadas (carpeta “BivariateCorrelations”) ‘ConfidenceInterval_correlation_bioclim.xls’)

requiere mover toda la primera fila una celda a la derecha (iniciando en la celda B1, quedando vacía la celda A1).

- Tabla ‘Pvalue_correlation_bioclim.xls’: Tabla con los p-valores de probabilidad de cada estimado de correlación. Dependiendo del nivel de significación que el usuario haya definido (usualmente 0.05 o 0.01), se aceptaría o no la hipótesis nula (que la estimación de la correlación es igual a 0, es decir, no hay correlación). Para ver correctamente esta tabla se requiere mover toda la primera fila una celda a la derecha (iniciando en la celda B1, quedando vacía la celda A1).
- Tabla ‘RelevantValues_correlation_bioclim.xls’: Esta tabla combina los resultados de las tablas ‘Estimate_correlation_bioclim.xls’ y ‘Pvalue_correlation_bioclim.xls’ de la siguiente manera: a) los valores cero indican que se cumplen las siguientes dos condiciones: valores estimados de correlación entre -0.5 y 0.5 (baja o nula correlación) y aceptación de la hipótesis nula (p-valores mayores o iguales a 0.05), y b) valores estimados de correlación originales diferentes de 0 indican que se cumplen las siguientes condiciones: valores estimados de correlación menores a -0.5 o mayores a 0.5 (alta correlación) y rechazo de la hipótesis nula (P-valores menores o iguales a 0.05). De esta manera, la tabla de valores relevantes muestra de manera sencilla los casos donde la correlación es alta y la significancia estadística apoya esa determinación.
- Tabla ‘Estimate_correlation_bioclim.txt’: tabla simétrica con los valores estimados de correlación de Pearson entre todas las variables, los cuales van de -1 a 1, siendo 0 el valor absoluto para no correlación, 1 para correlación positiva y -1 para correlación negativa o inversa. Para ver correctamente esta tabla se requiere mover toda la primera fila una celda a la derecha, iniciando en la celda B1, quedando vacía la celda A1.
- Tabla ‘ConfidenceInterval_correlation_bioclim.txt’: Tabla con los intervalos de confianza para los estimados de correlación de Pearson entre todas las variables. Para ver correctamente esta tabla se requiere mover toda la primera fila una celda a la derecha, iniciando en la celda B1, quedando vacía la celda A1.

5.4.2.2 Análisis de componentes principales (carpeta “PCA”)

- Tabla ‘Componente_eigenvalues.xls’: Tabla con los valores propios.
- Tabla ‘Componente_eigenvectors.xls’: Tabla con los vectores propios.
- Tabla ‘Bioclim_IVposition.xls’: Tabla de los vectores propios pero sin sus valores originales, sino la posición que ocupa la variable en cada componente según su valor de vector propio (absoluto).

5.4.2.3 Selección por Random Forest (carpeta “RandomForest”)

- Tabla ‘VariableImportance_RandomForest_bioclim.xls’: En esta tabla aparecen los valores de decrecimiento medio de la precisión (mean decrease accuracy) o decrecimiento medio del índice Gini (mean decrease Gini). Se puede ordenar fácilmente las variables de acuerdo al decrecimiento medio de la precisión (de mayor a menor), de tal manera que las variables más importantes aparecen en la parte superior de la tabla.
- Gráfica ‘VarImportanceDotChart_bio.jpeg’: Esta gráfica (Fig 27) muestra las variables ordenadas de mayor (parte superior) a menor importancia (parte inferior) de acuerdo al decrecimiento medio de la precisión (mean decrease accuracy) o decrecimiento medio del índice Gini (mean decrease Gini). Este mismo tipo de gráfica (para el caso del decrecimiento medio de la precisión) se usa en Cutler et al. (2007) para la determinación de la importancia de las variables.

Adicionalmente, dentro de las carpetas de cada componente, se hayan dos archivos que corresponde a tablas: 'accessions_used_componente_varselection.xls' y 'Componente_extractedValues.xls'. El primero muestra las accesiones que finalmente se usaron para determinar la importancia de las variables. La segunda tabla muestra los valores extraídos para cada accesión del total de las variables que el usuario quiso someter al análisis de SelecVar.

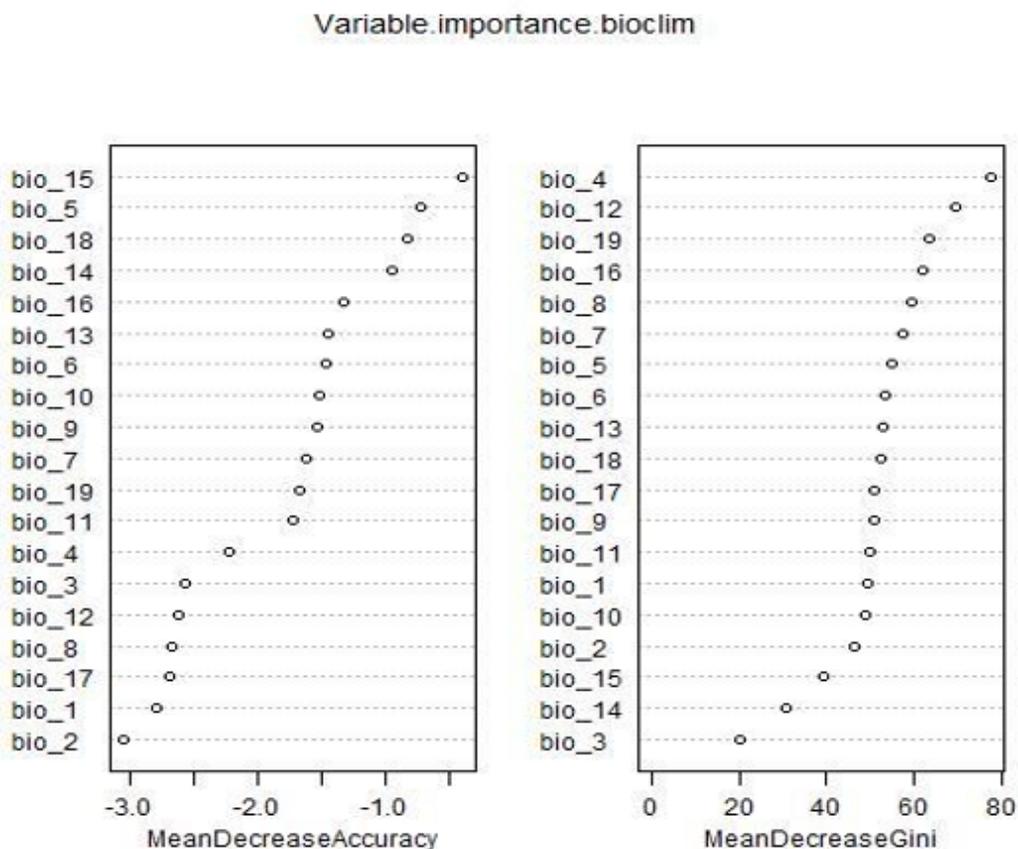


Figura 27. Ejemplo de gráfica del índice mean decrease accuracy y mean decrease Gini contenida en el archivo 'VarImportanceDotChart_bio.jpeg'.

Finalmente, SelecVar ofrece tres tablas: SelectedVariables_bioclim.xls, SelectedVariables_edaphic.xls y SelectedVariables_geophysic.xls, que muestran en una matriz diagonal, con los valores de correlación, las variables finalmente seleccionadas por cada componente, luego de aplicar los parámetros percenRF, percenCorr, CorrValue, pValue y nminvar. Estas variables serían el resultado de SelecVar (selección objetiva).

5.5. Referencias

Austin, M.P., Van Niel, K.P. 2010. Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography* 38(1): 1-8.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. J. 2007. Random forests for classification in ecology. *Ecology*, 88(11): 2783-2792.

Dean, N., Raftery, A.E., Scrucca, L. 2015. Package 'clustvarsel', variable selection for Model-Based clustering. <http://cran.r-project.org/web/packages/clustvarsel/clustvarsel.pdf>

King, J. R., Jackson, D. A. 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics*, 10(1): 67-77.

Plissock, P., Luebert, F., Hilger, H. H., Guisan, A. 2014. Effects of alternative sets of climatic predictors on species distribution models and associated estimates of extinction risk: A test with plants in an arid environment. *Ecological Modelling*, 288, 166-177.

Raftery, A. E., Dean, N. 2006. Variable selection for model-based clustering. *Journal of the American Statistical Association*, 101(473):168-178.



Taller Nacional, Comayagua, Honduras, Noviembre 2013.



6 | Herramienta ELCmapas

6.1. ¿Qué es un mapa de caracterización ecogeográfica del terreno?

Se trata de un tipo de mapa donde podemos visualizar diferentes escenarios ambientales que pueden corresponderse con los diferentes procesos adaptativos de una especie vegetal a lo largo de un determinado territorio. Entre sus múltiples usos, los mapas de caracterización ecogeográfica del terreno (mapas ELC) son útiles para la conservación y uso razonable de la agrobiodiversidad.

La idea de expresar adaptación a través de mapas no es nueva. Desde mediados del siglo pasado vienen desarrollándose mapas de biomas, ecosistemas, regiones ecológicas, etc. Estos mapas representan unidades ambientales generalmente como regiones grandes y homogéneas. En principio los “climas” o “ambientes” (los términos se usaban indistintamente) que representan estos mapas eran usados en estudios de diversos tipos de organismos (plantas, animales, microorganismos). Algunos mapas afinaron más y representaban, por ejemplo, los climas afines a las formaciones vegetales descritas por Leslie Holdridge en 1947, aunque luego se generalizaron como “sistemas de clasificación de zonas de vida”.

Estos mapas fueron de mucha utilidad para biólogos y naturalistas que intentaban comprender la distribución de organismos vivos en relación a la temperatura y la humedad. Aun hoy en día, el sistema de Holdridge tiene aplicación, por ejemplo, en estudios de cambio climático. Sin embargo, la mezcla a veces indistinguible de rasgos bióticos (vegetación) y abióticos (temperatura, precipitación) en estos mapas, la simplificación del componente abiótico en sólo dos factores y la forma de delinear las regiones (grandes, homogéneas y continuas) representaban un impedimento serio para su uso en estudios de adaptación en especies en particular.

La utilización de información sobre adaptación para diseñar una colecta, o tener criterios a la hora de conservar y utilizar de manera eficiente los recursos fitogenéticos tampoco es nada nuevo, aunque son escasos los trabajos publicados sobre la materia donde explícitamente se haga referencia a la adaptación. Existe, como referencia, un primer mapa ecogeográfico cuyo propósito era crear colecciones núcleo/nucleares (core collections) en 1997 (Tohme *et al.*, 1995), donde al final del proceso de selección de entradas además de la dimensión ecogeográfica se consideraron otros criterios adicionales de diferente naturaleza.

Desde entonces se han producido varios desarrollos: los programas SIG se han hecho gradualmente más flexibles y “amigables”, e incluso algunos paquetes estadísticos actualmente incluyen utilidades y herramientas SIG; la información ecogeográfica disponible (en forma de capas SIG) es de mejor calidad y más accesible; los equipos informáticos de alta capacidad de análisis se venden a precios reducidos y el acceso a internet se ha incrementado en países en desarrollo. Estos avances han repercutido en el desarrollo de mapas que representan diferentes escenarios de adaptación para plantas cultivadas y silvestres emparentadas. Es así como se obtuvo un primer mapa de caracterización ecogeográfica del territorio para España en 2005 (Parra Quijano *et al.*, 2008) de tipo generalista (que pudiera ser aplicado a varias especies silvestres emparentadas con cultivadas) aunque sólo fue utilizado para algunas especies del género *Lupinus*. Este mapa, obtenido por técnicas de análisis multivariado y determinación del número de grupos por criterios bayesianos, representó las diferentes unidades ambientales de forma reticulada y como regiones homo-

géneas, pero de dimensiones reducidas y discontinuas. Estas características físicas ya contrastaban con los mapas bioclimáticos tradicionales. Otra diferencia fue la inclusión de variables de tipo geofísico y edáfico (además de las bioclimáticas), con la idea de representar los aspectos abióticos que pueden afectar el desarrollo de una planta desde un punto de vista agronómico.

A mediados de 2008 se desarrolló un nuevo mapa ecogeográfico para España Peninsular e Islas Baleares usando otras fuentes de información ecogeográfica. La metodología de creación de este nuevo mapa fue similar al anterior. En este punto lo que más interesaba a los investigadores era verificar que el mapa realmente reflejara escenarios adaptativos, es decir, realizar una validación. Para ello se evaluó el desempeño del nuevo mapa para ocho especies (cuatro leguminosas y cuatro gramíneas), dos de ellas silvestres emparentadas con cultivadas y las otras seis compuestas por variedades locales, usando su distribución y la variable “peso de la semilla” como variable fenotípica indicadora con valor adaptativo. Los resultados se compararon con dos mapas de referencia, uno con una estructura física similar al mapa ecogeográfico (alta reticulación, discontinuidad, unidades de tamaño reducido), pero construido sin tener en cuenta aspectos relacionados con la adaptación abiótica de plantas (mapa CORINE land cover, un mapa de uso de suelo) y otro con una estructura física diferente (más parecida a la de los mapas tradicionales) pero construido con un fin similar (mapa DMEER o mapa digital de regiones ecológicas europeas). Una comparación sobre un mapa de ecoregiones más reciente puede observarse en la Fig. 28.

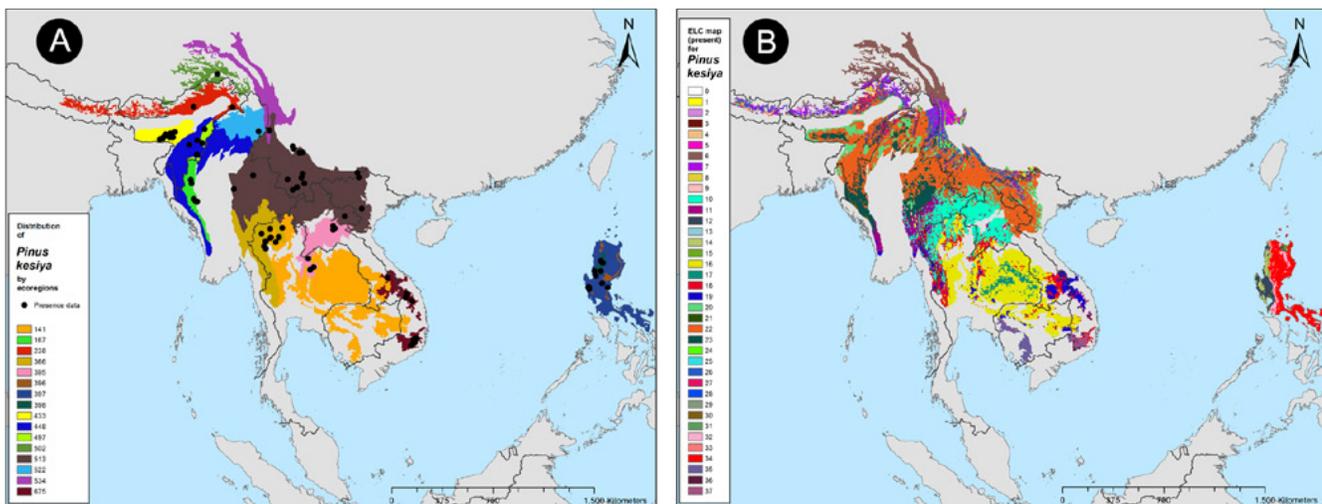


Figura 28. Comparación de un mapa de ecoregiones y un mapa de caracterización ecogeográfica del territorio (ELC) para la especie *Pinus kesiya* en el sudeste asiático. A) Distribución de las ocurrencias conocidas de la especie *Pinus kesiya* (puntos negros) sobre las 16 diferentes ecoregiones terrestres ocupadas del mapa Global Ecoregions2017© (Dinerstein et al., 2017). B) Mapa ELC obtenido con la herramienta ELCmapas, de mucho mayor detalle para la especie, con 38 categorías.

Los resultados del estudio de validación fueron variopintos. El mapa de caracterización ecogeográfica del territorio funcionó en general mejor para especies leguminosas que para gramíneas, aunque la excepción fue *Zea mays* para la cual el desempeño del mapa fue bastante aceptable. Como era de esperar, el mapa reflejó escenarios adaptativos para las dos especies silvestres, pero también produjo resultados satisfactorios en casos de especies compuestas por sólo

variedades locales, como el caso de *Phaseolus vulgaris*. Como conclusión, los mapas de caracterización ecogeográfica del territorio son capaces de reflejar escenarios adaptativos y, por tanto, pueden ser utilizados en muchas actuaciones relacionadas con la colecta, conservación y utilización eficiente de los recursos fitogenéticos. Sin embargo, es recomendable crear mapas específicos para cada especie o grupo de especies relacionados filogenéticamente. Realizar mapas de caracterización ecogeográfica del terreno de tipo generalista puede resultar arriesgado a la hora de sacar conclusiones para un grupo numeroso de especies, particularmente si no se realiza una validación del mapa. También es importante hacer una selección apropiada de las variables ecogeográficas involucradas en la creación del mapa, pero siempre representando los tres aspectos abióticos involucrados en el desarrollo de las plantas: bioclimático, geofísico y edáfico.

6.2. Antecedentes de la herramienta ELCmapas

La herramienta ELCmapas que aborda este instructivo representa el desarrollo del concepto Ecogeographical Land Characterization Maps (ELC) expuesto originalmente en la publicación de Parra Quijano y colaboradores (2012 A)

Este tipo de mapas han tenido usos diversos en colecta, conservación y uso de recursos fitogenéticos como en los trabajos de Parra Quijano y colaboradores (2011 A, 2011 B y 2012 B) particularmente para España, Thormann (2012) a nivel Europa, Taylor y colaboradores (2013) para República Checa, Phillips y colaboradores (2016) para Noruega, Fitzgerald y colaboradores (2019) para los países nórdicos, Contreras-Toledo y colaboradores (2019) para México, Tapia y colaboradores (2019) en Ecuador, o Zair y colaboradores (2020) para países del medio oriente.

El interés que esta metodología había despertado en varios equipos y proyectos de investigación en colecta, conservación y uso de recursos fitogenéticos contrastaba con una observación recurrente entre posibles usuarios: la metodología descrita en dicha publicación es un tanto compleja dado que mezcla técnicas de sistemas de información geográfica (SIG) y análisis multivariado. Además, el desarrollo original implicaba el uso de un programa de análisis estadístico de uso no gratuito. Estos aspectos limitaban de manera importante el desarrollo de mapas ELC por parte de investigadores y técnicos.

6.3. Características de ELCmapas

La herramienta ELCmapas provee una nueva opción para desarrollar mapas de caracterización ecogeográfica del terreno mediante el uso de R, evitando las complicaciones anteriormente descritas. Este software, gratuito y con amplia capacidad de cálculo estadístico y potencia gráfica, permite integrar SIG y análisis multivariado. Así, la nueva herramienta ELCmapas es capaz de producir mapas de caracterización ecogeográfica del terreno sin necesidad de alternar entre diferentes programas, ni descargar y manipular información ecogeográfica. Sin embargo, es importante aclarar que los productos de la herramienta ELCmapas son justamente mapas y tablas que pueden ser visualizadas en programas como DIVA-GIS, Google Earth o Microsoft Excel. Los mapas de caracterización ecogeográfica del terreno de la herramienta ELCmapas son así mismo un componente de otras herramientas como Representa.

ELCmapas pone a disposición del usuario seis diferentes procedimientos para determinar de manera objetiva el número óptimo de grupos. Estos procedimientos son:

- a) **elbow**: Un sistema simple que usa como algoritmo de agrupamiento K-means con determinación de punto de corte basado en la disminución de la suma de cuadrados intragrupo (Ketchen y Shook, 1996). El número de grupos óptimo es el alcanzado cuando la disminución de la sumas de cuadrados intragrupos entre una solución de n y una de $n+1$ grupos es menor del 50%. Este método, también conocido como de “elbow” o codo, es el de más rápido cálculo y es capaz de actuar sobre grandes cantidades de datos sin grandes demoras, por lo cual se aconseja para países de gran tamaño.
- b) **medoides**: Método de agrupamiento de partición alrededor de los medoides (pam). Se utiliza el método de silhouette de interpretación y validación del número de grupos. Este sistema (en principio gráfico, posteriormente adaptado en R por el paquete fpc) permite determinar qué tan apropiadamente han sido los datos agrupados (Kaufman y Rousseeuw, 1987; Rousseeuw, 1987). Es un sistema que consume más recursos computacionales, por lo que aplicado a grandes conjuntos de datos, consumirá más tiempo.
- c) **calinski**: Utilización del criterio Calinski-Harabasz (1974) para seleccionar el número óptimo de grupos a partir de la generación de agrupamientos en cascada por Kmeans. Al igual que medoides, la determinación del número óptimo de grupos y la asignación de grupo a cada celda puede consumir bastante tiempo cuando se trata de áreas de trabajo grandes (se generan errores de memoria al usarse en capas de más de 800,000 celdas, por ejemplo, áreas/países mayores a 800,000 km² a una resolución de 1x1).
- d) **ssi**: “Simple structure index” (Donilcar *et al.*, 1999), para seleccionar el número óptimo de grupos a partir de la generación de agrupamientos en cascada por Kmeans. Al igual que medoides, la determinación del número óptimo de grupos y la asignación de grupo a cada celda puede consumir bastante tiempo cuando se trata de áreas de trabajo grandes (se generan errores de memoria al usarse en capas de más de 800,000 celdas, por ejemplo, áreas/países mayores a 800,000 km² a una resolución de 1x1).
- e) **bic**: “Bayesian Information Criterion”. Este criterio establece el número óptimo de grupos para modelos gaussianos mixtos parametrizados inicializados por un método de agrupamientos jerárquico basado en modelos. Dado que la descripción en detalle de este criterio y el método de agrupamiento basado en modelos resultan inabordables en el contexto de éste manual, se recomienda a los usuarios leer el artículo de Fraley y Raftery (2007) y la implementación del método en el paquete mclust de R (ver <https://cran.r-project.org/web/packages/mclust/index.html>). Este método es muy demandante de memoria, por lo que no se recomienda usar para áreas de trabajo de más de 50,000 km² con resolución de celda 1x1 km o más de 350,000 km² con una resolución de celda 5x5 km.
- f) **kmeansbic**: Se trata de una combinación del procedimiento k means con la determinación de BIC. Para ello, la función en R realiza el proceso de agrupamiento que se usa en el Análisis Discriminante de Componentes Principales (DAPC en sus siglas en inglés). El proceso consiste en la ejecución de sucesivos agrupamientos por k means usando como variables los componentes principales extraídos de las variables originales, donde en cada agrupamiento se va incrementando el número de agrupamientos (k). Para cada k means se calcula como medida estadística de bondad de ajuste el criterio de información bayesiana (BIC), lo cual permite determinar el número óptimo de agrupamientos.

Los métodos de determinación del número de grupos no son del todo objetivos, pues al menos el usuario debe determinar un máximo número de grupos que permitirá. Dado que la información ecogeográfica a niveles de re-

solución de 1 km o incluso de 5 km puede llegar a ser bastante voluminosa a nivel continental o regional, como por ejemplo Latinoamérica, la herramienta ELCmapas está inicialmente pensada para ser utilizada a nivel país. Sin embargo, la distribución de las especies objetivo o la distribución de las recolecciones de germoplasma pueda ir más allá de los límites de un país determinado, por tanto, mapas ELC a niveles regionales (varios países contiguos), subcontinentales, continentales e incluso globales serían necesarios. Por ello la herramienta ELCmapas en su actualización para CAPFITOGEN 3, utilizada desde una computadora de prestaciones medias a altas, o más aún desde un servidor, sería capaz de procesar información y proporcionar resultados a niveles regionales, subcontinentales, continentales o globales.

6.4. La selección de las variables ecogeográficas

Este es un aspecto muy importante a tener en cuenta antes de usar la herramienta ELCmapas. Por esta razón fue desarrollada la herramienta SelecVar (capítulo 5). El cambio, adición o eliminación de una sola variable de un sólo componente (bioclimático, geofísico o edáfico) cambiará sensiblemente la configuración final del mapa y su correlación con los escenarios adaptativos de la especie.

La técnica de mapas de caracterización ecogeográfica del terreno originalmente no contemplaba hacer mayor discriminación de variables dado que se pensaba en mapas de corte generalistas. Sin embargo, a medida que se reconoció que su capacidad de discriminar correctamente escenarios adaptativos se incrementaba al enfocarse en una especie en particular o un grupo de especies muy afines (en términos genéticos), surgió la idea de seleccionar las variables ecogeográficas de cada componente que más influyen en la adaptación abiótica de la especie o especies, y que por tanto determinan su distribución.

El proceso de selección es entonces un aspecto clave a la hora de obtener mapas más precisos en términos adaptativos. El listado de variables potencialmente seleccionables puede crearse a partir de:

- a) Búsquedas bibliográficas: No resulta difícil encontrar en literatura de tipo técnica y/o científica, como artículos, libros o documentos, algunas referencias sobre los factores ambientales que influyen, determinan o limitan la distribución de una especie. Frecuentemente se debe buscar la correspondencia entre las referencias sobre esos factores con las variables disponibles en forma de capas SIG para realizar el mapa.
- b) Conocimiento experto: La consulta de personas expertas en la especie o grupo de especies, sobre qué variables ecogeográficas pueden ser consideradas factores clave en su adaptación y distribución suele ofrecer información muy valiosa a la hora de seleccionar variables. Esta consulta introduce subjetividad en el proceso, algo a lo que no se debe temer en absoluto. En el proceso de creación de mapas de caracterización ecogeográfica del terreno, el conocimiento experto en estas fases previas puede ser la diferencia entre un mapa correctamente validado o un mapa con poco sentido en términos de adaptación de la especie objetivo. El aporte del conocimiento experto puede ser más decisivo en la medida que se consulte el mayor número de expertos posible y luego se determinen consensos. Un buen ejemplo de mapa de caracterización ecogeográfica del terreno en el que se ha utilizado conocimiento experto es el trabajo de Parra Quijano y colaboradores (2012 C). En este estudio el mapa fue utilizado para la determinación de la localización idónea de reservas genéticas para varias especies del género *Beta* en Europa.

Alternativamente a la elaboración de este listado potencial de variables de tipo subjetivo, se puede determinar por cada componente (bioclimático, geofísico y edáfico) qué variables son de mayor importancia a la hora de clasificar los escenarios o cuales pueden estar aportando información redundante. Para ello se puede realizar un análisis de clasificación (como Random Forest), de correlaciones bivariadas o un análisis de colinealidad. En caso de encontrarse valores altos de correlación entre dos variables, que hagan parte de un mismo componente, se debería descartar una de ellas. Adicionalmente un análisis de componentes principales (en caso de que todas las variables sean cuantitativas) puede ayudar a determinar las relaciones de las variables y facilitar la selección final. Estos análisis están a disposición del usuario de la tecnología CAPFITOGEN en la herramienta SelecVar (ver Capítulo 5). No es recomendable utilizar más de cinco variables por componente pues la configuración de las zonas (celdas adyacentes con el mismo valor) del mapa resultante puede llegar a ser difícil de comprender. Así mismo, utilizar latitud y longitud (parámetros 6.5.1.7 y 6.5.1.8) ofrece mapas con categorías más agrupadas geográficamente (menos reticuladas). El efecto contrario (reticulación) se obtiene al utilizar variables como “orientación” del componente geofísico.

Una vez determinada la lista final de variables, éstas se seleccionan en los parámetros bioclimv, geophysv y edaphv.

6.5. Utilización de la herramienta ELCmapas

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta ELCmapas, se deben especificar una serie de parámetros por parte del usuario.

6.5.1 Parámetros iniciales definidos por el usuario

6.5.1.1 Parámetro: ruta (sólo aplica para el modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

6.5.1.2 Parámetro: país

Explicación: Seleccione el país para el cual quiere construir el mapa ELC. En la lista desplegable dispone de más de 160 países, algunas coberturas a nivel subcontinental, a nivel global y coberturas personalizadas (provenientes de la herramienta rLayer).

6.5.1.3 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la generación del mapa. Note que 1x1 km ofrece mayor resolución pero exige mucha más capacidad de cómputo y tarda mucho más que 5x5 km, especialmente en países de gran extensión. Mapas de alta resolución (i.e. 1x1 km) compuestos de más de 100,000 celdas (países de

más de 100,000 km²) pueden generar problemas de procesamiento para algunos métodos de agrupamiento y determinación de número óptimo de grupos como “cluster”.

6.5.1.4 Parámetro: bioclimv

Explicación: Liste (modo local) o seleccione (modo on server) las variables bioclimáticas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables bioclimáticas), cópielos y péguelos separados por un signo de punto y coma (;). También encontrará una línea bloqueada la selección de las 19 variables bioclim, listas para usar, simplemente eliminando el símbolo # inicial. En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro bioclimv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

6.5.1.5 Parámetro: geophysv

Explicación: Liste (modo local) o seleccione (modo on server) las variables geofísicas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables geofísicas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro geophysv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

6.5.1.6 Parámetro: latitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la latitud (Y) como una variable geofísica más a ser analizada.

6.5.1.7 Parámetro: longitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la longitud (X) como una variable geofísica más a ser analizada.

6.5.1.8 Parámetro: edaphv

Explicación: Liste (modo local) o seleccione (modo on server) las variables edáficas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece

en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables edáficas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro edaphv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITO-GEN3 modo local.

6.5.1.9 Parámetro: maxg

Explicación: Indique el número máximo de agrupamientos por componente (bioclimático, geofísico y edáfico) que desea permitir (a mayor número, mapas con más categorías). Se recomienda un valor menor o igual a cinco, de lo contrario se pueden llegar a formar mapas ELC de más de 125 categorías.

6.5.1.10 Parámetro: metodo

Explicación: Seleccione el método para generar los agrupamientos con determinación objetiva del número óptimo de grupos. Encontrará disponibles los seis métodos descritos en este capítulo por los términos: “kmeansbic”, “medoides”, “elbow”, “calinski”, “ssi” y “bic”

6.5.1.11 Parámetro: iterat

Explicación: Sólo aplica si para el parámetro “metodo” se ha seleccionado la opción “calinski” o “ssi”. En este campo indique el número de permutaciones usadas para generar agrupamientos K-means sobre las cuales se calcula los criterios “calinski” o “ssi”.

6.5.1.12 Parámetro: resultados (sólo aplica para el modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis (modo local). Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

6.6. Resultados de ELCmapas

Después de definir todos los parámetros y rutas (para modo local) que requiere ELCmapas, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, ELCmapas guardará los resultados en la ruta y carpeta que se haya determinado en “resultados” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server. Allí encontrará cinco mapas y tres tablas.

6.6.1 Los mapas

El mapa ELC final del país o la región determinada en el parámetro del apartado 6.5.1.2 está contenido en el archivo mapa_elc_pais.tif (geotif) o en versión grd el cual se puede ver y editar en DIVA-GIS (al abrir el archivo mapa_elc_DIVA_pais.grd), además de otros formatos gráficos o para Google earth (png o kml respectivamente). Si el mapa ELC resultado va a ser usado por otras herramientas como Representa, ColNucleo, FIGS_R o Tzones, se debe copiar el archivo mapa_elc_pais.grd (el cual no debe ser abierto o editado) y pegar en la carpeta correspondiente (modo local) o se debe subir (modo on server) a través del botón que habilita la selección y posterior subida de archivos.

Los mapas que representan las categorías resultantes de los componentes bioclimático, geofísico y edáfico (mapa_bioclimatico_pais.tif, mapa_geofisico_pais.tif y mapa_edafico_pais.tif) pueden ser abiertos en diferentes software SIG, con la precaución que se encuentran en el sistema de coordenadas WGS84 (latlong). Todos los mapas cuentan con una versión alternativa en .grd, el cual es el formato nativo para el software DIVA-GIS. Inicialmente DIVA-GIS abre los mapas como lo muestra la Fig. 29.

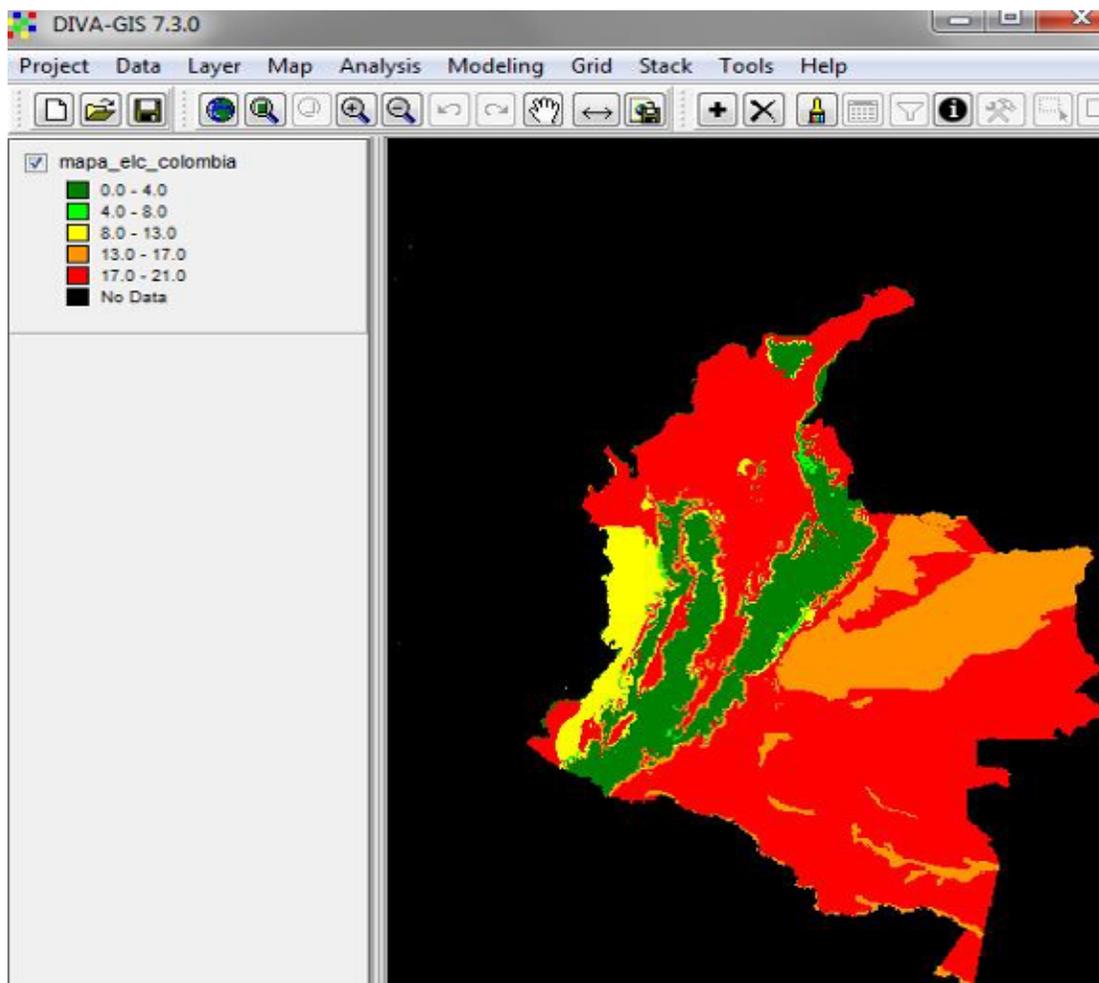


Figura 29. Ejemplo de un mapa de Colombia producido por ELC mapas tal y como lo mostraría en DIVA-GIS sin ningún tipo de edición. Las 20 categorías se muestran agrupadas en 5 rangos.

Sin embargo, su visualización se puede alterar al hacer doble clic sobre el cajón gris de la parte izquierda que representa esa capa en la interfaz de DIVA-GIS. Agregando tantas filas hasta alcanzar el número de categorías que contiene el mapa, y luego aplicando una rampa de colores aleatorios, se puede obtener un mapa como el que muestra la Fig. 30. Es conveniente que los colores sean lo más diferentes posible para que se identifiquen perfectamente las categorías (escenarios ecogeográficos) presentes en el territorio.

NOTA: Recuerde siempre que la categoría denominada como “0” (cero) no hace parte de las categorías ecogeográficas del mapa resultado, sino que es la denominación de todas las zonas para las cuales existe información de uno o dos componentes pero no de los tres. Por ejemplo, por obvias razones es normal que la información de suelos para zonas urbanas o cuerpos de agua sea nula, pero si puede existir información bioclimática e incluso geofísica para estas áreas. Esas áreas serán codificadas como “0”.

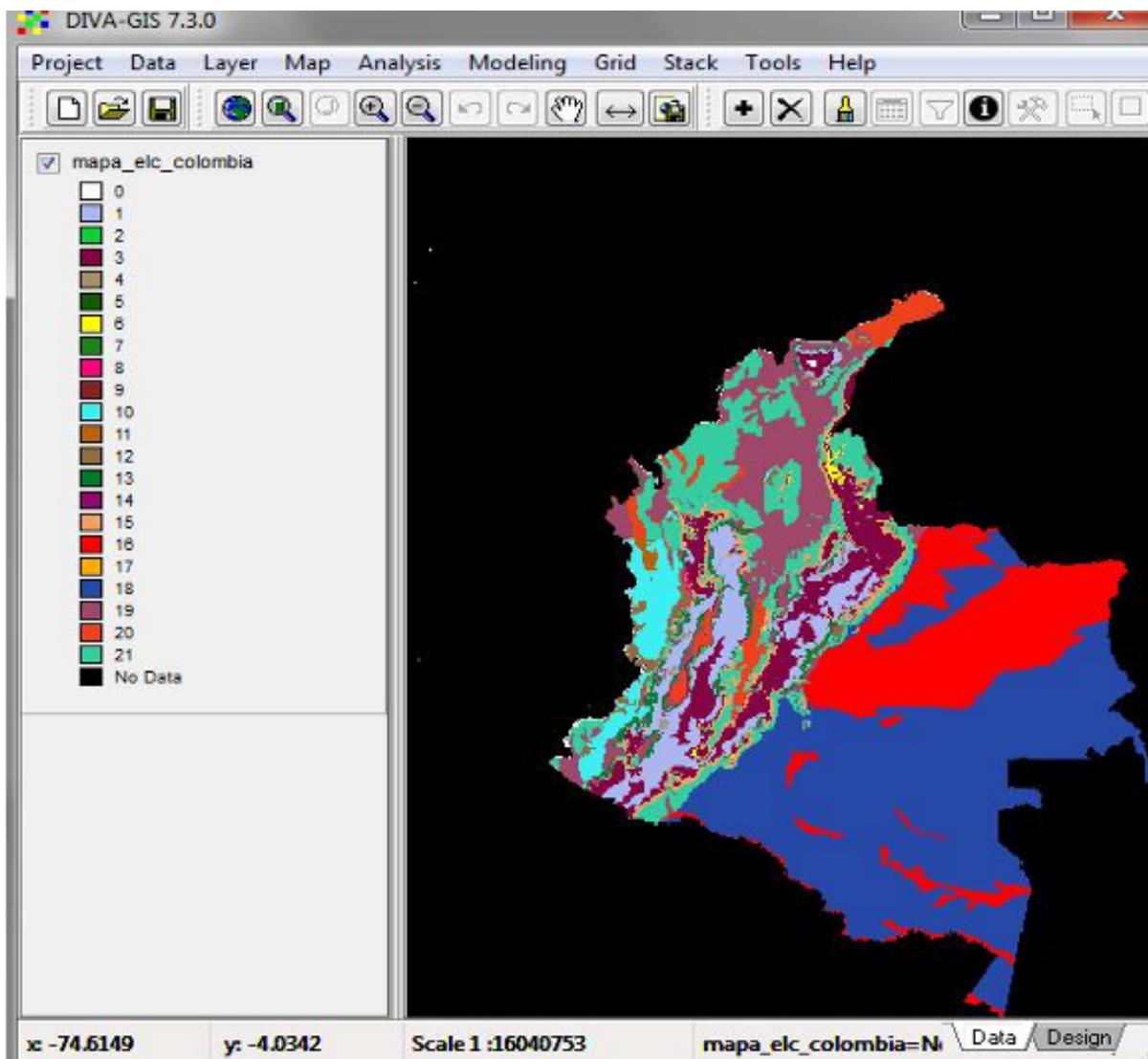


Figura 30. Ejemplo de un mapa de Colombia producido por ELC mapas que muestra un color por cada una de sus categorías. Las propiedades del mapa tal y como lo abre DIVA-GIS fueron alteradas para mostrar cada categoría con un color diferente.

Adicionalmente a los mapas .tif y .grd, se produce un mapa compatible con Google Earth (“mapa_elc_pais.kml”). Siempre que se tenga instalado Google earth, este mapa se podrá abrir como una capa al hacer doble clic sobre el archivo. Este mapa no puede ser manipulado (cambio de colores) y su calidad gráfica no es óptima.

6.6.2 Las tablas

Las tablas producidas por ELC mapas y el resto de las aplicaciones CAPFITOGEN3 se ofrecen en extensiones .txt y .xls. Si bien esta última extensión corresponde a Microsoft Excel®, en realidad es el mismo archivo de texto delimitado por tabulaciones. Entonces, ¿por qué duplicar la misma información en el mismo formato ofreciéndola con dos extensiones diferentes?. La explicación radica en que con la extensión .xls se le indica al sistema operativo que la tabla debe ser abierta en Excel®. Al intentar abrir este archivo de texto, Excel® reconoce que no corresponde a su formato y emite una señal de alerta (ver Fig. 25), la cual puede ser ignorada. El formato .txt se puede abrir con opción “Abrir con” y paso siguiente indicar algún software compatible, entre ellos el mismo Excel®.

6.6.2.1 “Tabla_ELC_celdas_pais.txt”. Esta tabla muestra los valores de las variables seleccionadas y los valores de las categorías ELC (“ELC_CAT”), bioclimáticas, geofísicas y edáficas para cada centroide (fila) de cada celda que compone el territorio del país objeto de estudio. También incluye los valores de latitud y longitud de cada centroide.

6.6.2.2 “numero_categorias_pais.txt”. Contiene un simple recuento de las categorías ecogeográficas producidas y representadas en el mapa ELC resultante (columna “N_ELC_CAT”) y el número de categorías producidas por componente.

6.6.2.3 “Estadist_ELC_pais.txt”, “Estadist_BIOCLIM_pais.txt”, “Estadist_EDAPH_pais.txt”, “Estadist_GEOPHYS_pais.txt” y “Combi_ELC_ecuador.xls”. En estas tablas se encuentran los estadísticos descriptivos (media, valor mínimo, valor máximo y desviación estándar) para cada una de las variables originales involucradas en la creación del mapa ELC y para los mapas de cada uno de los componentes (bioclimático, geofísico y edáfico) representados en el mapa ELC. Estas tablas se asemejan a la tabla suplementaria S2 presentada a manera de descripción de las categorías del mapa ELC en el artículo de Parra Quijano y colaboradores (2012 A). Esta tabla es requerida junto al archivo de mapa mapa_elc_pais.grd y su archivo acompañante mapa_elc_pais.gri cuando se trabaja con las herramientas Representa y ColNucleo (opcionalmente FIGS_R) y debe copiarse junto con los archivos de mapa en la carpeta indicada por cada herramienta (modo local) o subirla en el parámetro statelc (modo on server).

6.6.2.4 “Bioclim_BIC_results.txt”, “edaph_BIC_results.txt” y “edaph_BIC_results.txt”. Estas tablas resultan de la selección del método “bic” en el parámetro 6.5.1.10. Contienen los más importantes parámetros de los análisis de agrupamiento para cada componente, como el modelo y el número de grupos óptimos y utilizados para generar el mapa ELC. También se genera una gráfica en formato jpg que muestra los valores BIC vs el número de agrupamientos para cada componente (bioclimático, edáfico o geofísico).

6.7. Referencias

- Calinski, T., Harabasz, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics*. 3(1): 1-27.
- Contreras-Toledo, A. R., Cortés-Cruz, M., Costich, D. E., Rico-Arce, M., Magos Brehm, J., Maxted, N. 2019. Diversity and conservation priorities of crop wild relatives in Mexico. *Plant Genetic Resources Characterisation and Utilisation*, 17, 140-150.
- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hansen, M. 2017. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534-545.
- Dolnicar, S., Grabler, K., Mazanec, J. A. 1999. A tale of three cities: perceptual charting for analyzing destination images. Pp. 39-62 in: Woodside, A. et al. (eds) *Consumer psychology of tourism, hospitality and leisure*. CAB International, New York.
- Fitzgerald, H., Palmé, A., Asdal, Å., Endresen, D., Kiviharju, E., Lund, B., Rasmussen, M., Thorbjornsson, H., Weibull, J. 2019. A regional approach to Nordic crop wild relative in situ conservation planning. *Plant genetic resources*, 17(2), 196-207.
- Fraley C., Raftery, A.E. 2007. Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* 24:155-181.
- Kaufman, L., Rousseeuw, P.J. 1987, Clustering by means of Medoids, in *Statistical Data Analysis Based on the L1-Norm and Related Methods*. Y. Dodge (eds), North-Holland, 405-416.
- Ketchen, D. J., Shook, C. L. 1996. The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal* 17(6): 441-458.
- Parra-Quijano, M. Iriondo, J.M., De la Cruz, M., Torres, M.E. 2011 A. Strategies for the development of core collections based on ecogeographical data. *Crop Science* 51:656-666
- Parra-Quijano, M. Iriondo, J.M., Torres, M.E., De la Rosa, L. 2011 B. Evaluation and validation of ecogeographical core collections using phenotypic data. *Crop Science* 51:694-703
- Parra-Quijano, M.; Draper, D.; Torres, E., Iriondo, J.M. 2008. Ecogeographical representativeness in crop wild relative ex situ collections. p. 249-273. In Maxted, N.; Ford-Lloyd, B.V.; Kell, S.P.; Iriondo, J.M.; Dulloo, M.E., Turok, J. (ed.) *Crop wild relative conservation and use*. CAB International, Wallingford.
- Parra-Quijano, M. Iriondo, J.M., Torres, M.E. 2012 A. Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genetic Resources and Crop Evolution* 59(2):205-217 DOI 10.1007/s10722-011-9676-7

Parra-Quijano, M. Iriondo, J.M., Torres, M.E. 2012 B. Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodiversity and Conservation* 21:79-96 DOI 10.1007/s10531-011-0167-0

Parra-Quijano, M. Iriondo, J.M., Frese, L., Torres, M.E. 2012 C. Spatial and ecogeographic approaches for selecting genetic reserves in Europe. *En: N. Maxted, M.E. Dulloo, B.V. Ford-Lloyd, L. Frese, J. Iriondo, M.A.A. Pinheiro de Carvalho (ed.) Agrobiodiversity Conservation: securing the diversity of crop wild relatives and landraces.* CABI, Wallingford, UK.

Phillips, J., Asdal, Å., Magos Brehm, J., Rasmussen, M., Maxted, N. 2016. *In situ* and *ex situ* diversity analysis of priority crop wild relatives in Norway. *Diversity and Distributions*, 22(11), 1112-1126.

Rousseeuw, P.J. 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* 20: 53-65. doi:10.1016/0377-0427(87)90125-7.

Tapia, C., Paredes, N., Lima, L. (2019). Representatividad de la diversidad del género musa en el ecuador. *Revista Científica Ecuatoriana*, 6(1).

Taylor, N. G., Kell, S. P., Holubec, V., Parra-Quijano, M., Chobot, K., Maxted, N. (2017). A systematic conservation strategy for crop wild relatives in the Czech Republic. *Diversity and Distributions*, 23(4), 448-462.

Thormann, I. 2012. Applying FIGS to crop wild relatives and landraces in Europe. *Crop Wild Relative* 8 14:16.

Tohme, J., Jones, P., Beebe, S., Iwanaga, M. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. p. 95-107. In Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L., Morales, E.A.V. (eds.) *Core collections of plant genetic resources.* IPGRI, Rome.

Zair, W., Maxted, N., Brehm, J. M., Amri, A. 2020. *Ex situ* and *in situ* conservation gap analysis of crop wild relative diversity in the Fertile Crescent of the Middle East. *Genetic Resources and Crop Evolution*, 1-17.



Taller de Birmingham, UK, Febrero 2014.



7 | Herramienta ECO GEO

7.1. Caracterización ecogeográfica del germoplasma

Se entiende caracterización ecogeográfica como el análisis de toda información ambiental del sitio donde crece un individuo o una población vegetal, directamente relacionada con el proceso de adaptación al entorno biótico o abiótico. En particular, las herramientas CAPFITOGEN permiten analizar sólo el componente abiótico, clasificado por tres aspectos principales, los cuales suelen considerarse en estudios de adaptación de cultivos (Ceballos-Silva y López-Blanco, 2003) o zonificación agrícola (Williams *et al.*, 2008):

- a) **Bioclimático:** Hace referencia a factores relacionados con temperatura y precipitación. También incluye las relaciones entre temperatura y precipitación que se manejan a través de índices.
- b) **Geofísico:** Agrupa los factores topográficos y de relieve de mayor importancia, especialmente relacionados con la radiación solar.
- c) **Edáfico:** Factores relacionados con las condiciones físicas y/o químicas de la fracción del suelo de la que las plantas dependen.

Por tanto, caracterizar ecogeográficamente un conjunto de entradas implica asignar a cada una de ellas, información de tipo bioclimático, geofísico y edáfico del sitio de su recolección.

La información ecogeográfica del sitio de recolección por si sola muestra muchos rasgos adaptativos del germoplasma, y en conjunto con otro tipo de caracterizaciones, como la fenotípica o genotípica, puede resultar muy útil para explicar los patrones genéticos observados. En algunos casos, como los que se dan en ausencia de recursos económicos suficientes para otro tipo de estudios, la caracterización ecogeográfica puede considerarse una alternativa válida, sencilla y barata para facilitar la utilización del germoplasma por parte de mejoradores que buscan parentales con ciertos rasgos adaptativos en las colecciones.

El insumo más importante para realizar una caracterización ecogeográfica son las coordenadas o la descripción del sitio de recolección (de la que se pueda extraer una coordenada), que usualmente se registran en los descriptores de pasaporte en el momento de la recolección. Usando dichas coordenadas se asignará a cada entrada los datos que describen los rasgos ambientales más importantes del sitio donde fue recolectada. Esto implica que la calidad de esas coordenadas es un aspecto crucial para una asignación correcta de información ecogeográfica, razón por la que se sugiere el uso de la herramienta GEOQUAL antes de realizar una caracterización de este tipo. Aparte de las coordenadas como materia prima, una caracterización ecogeográfica requiere de información ambiental que cubra toda el área de trabajo y de un programa especializado para la extracción de la información correspondiente a cada sitio de recolección. Los SIG son el tipo de programa que permiten realizar este proceso.

El producto de una caracterización ecogeográfica es similar al de cualquier otro tipo de caracterización, una matriz de datos, donde las filas suelen corresponder a las entradas y las columnas a los descriptores. A partir de esta matriz inicial es posible realizar análisis multivariados, como los que frecuentemente se realizan con otros tipos de caracterización, que permiten en este caso determinar relaciones de similitud ambiental entre los diferentes sitios de recolección. Un análisis factorial, como por ejemplo el de Componentes Principales (ACP), permitiría también conocer las relaciones entre las diferentes variables originalmente introducidas y crear variables sintéticas no correlacionadas que pudieran describir las afinidades ecogeográficas entre las entradas con un número reducido de componentes.

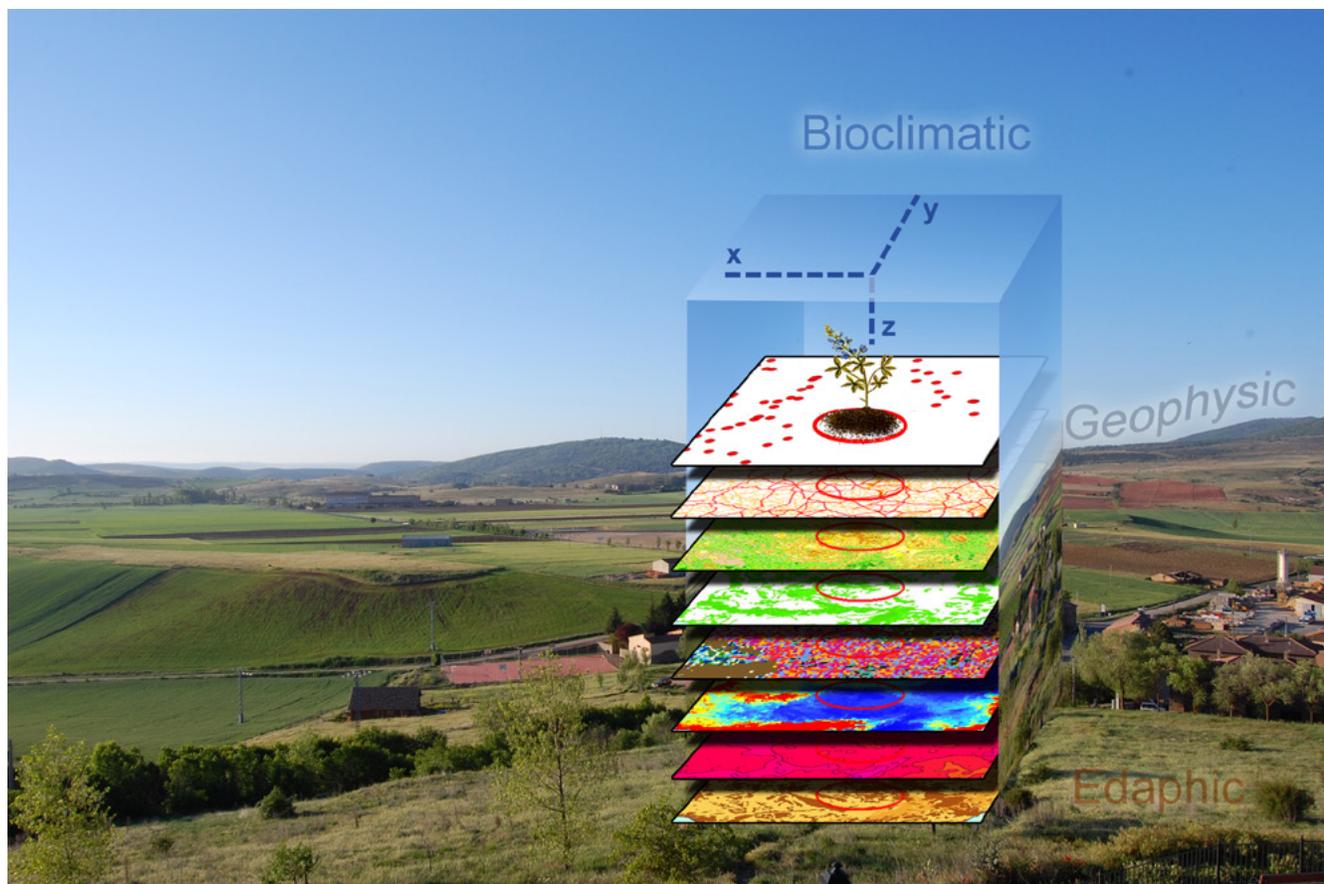


Figura 31. Ilustración del proceso de extracción de información ecogeográfica que se realiza para un sitio de recolección a través de SIG.

Es importante resaltar que la caracterización ecogeográfica arroja información sobre los sitios de recolección, no directamente de la naturaleza del germoplasma. Por esto, los análisis de multivariados que funcionan sobre matrices de distancia o disimilaridad, en este caso lo que reflejan es la afinidad ambiental e, indirectamente, adaptativa que hay entre los diferentes sitios de recolección. Entradas de una misma especie con patrones genotípicos o fenotípicos diferentes pueden ocurrir en escenarios ambientales muy parecidos o incluso indiferenciables.

7.2. Características de ECOGEO

La herramienta ECOGEO realiza para el usuario una extracción de información ecogeográfica de alrededor de 160 variables, para una lista de accesiones que el usuario introduzca en el análisis mediante el formato de datos de pasaporte FAO/Bioversity 2012, con modificaciones menores. Sin embargo, a partir de CAPFITOGEN3, ECOGEO es capaz de realizar, además del proceso completo de caracterización ecogeográfica, análisis multivariados con datos de caracterización fenotípica o genotípica, en el mismo sentido que lo realiza la herramienta DIVmaps. Con esto se logra que ECOGEO sea una herramienta mucho más completa para los técnicos que conservan y caracterizan recursos

fitogenéticos para la alimentación y la agricultura. Si tiene dudas sobre las características de las tablas de caracterización fenotípica o genotípica, podrá encontrar más detalles sobre ellas en el capítulo sobre la herramienta DIVmapas.

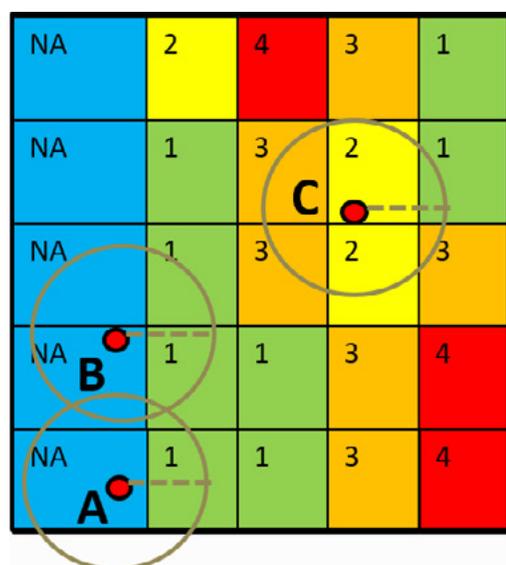
CAPFITOGEN ofrece al usuario toda la información necesaria para una caracterización ecogeográfica con ECOGEO, evitándole a éste tener que descargar, adaptar y uniformizar capas de información desde diferentes sitios de internet. Así, las capas de variables ecogeográficas de CAPFITOGEN están listas para su uso y dispuestas para funcionar con las programaciones R de la herramienta ECOGEO, entre otras (ver Capítulo 2).

El área de trabajo es un segundo aspecto que define el usuario. Frecuentemente el área de trabajo corresponde a los límites territoriales de un estado, tal y como viene definido en la base de datos global sobre áreas administrativas (<http://www.gadm.org>). Las variables o capas ecogeográficas están recortadas según sus límites, de tal manera que si se elige un país determinado y los datos de pasaporte incluyen coordenadas que indican sitios por fuera de este país, a las entradas correspondientes a dichas coordenadas no se les asigna ninguna información. Podrían existir opciones para hacer análisis más allá de las fronteras de un país determinado, que es lo que se considera una región o incluso un continente. Si hubiera disponibilidad de regiones o continentes dentro de la herramienta, (aparecerían en los listados del parámetro “país”), el usuario puede usar estas áreas de trabajo de mayor cubrimiento, sabiendo que seguramente el nivel de resolución de esta información será de menor alcance (tamaños de celda sobre los 10x10 km).

Hay un aspecto importante que introduce la herramienta ECOGEO. Se trata de la forma en la que se extrae la información ecogeográfica para un sitio de recolección. Usualmente las extracciones se realizan para el punto que señala la coordenada. Sin embargo, ocurren dos situaciones en las que la extracción “puntual” no refleje las condiciones abióticas más reales del sitio de recolección:

- a) Cuando se tiene poca información sobre las coordenadas o en general éstas no son de buena calidad, según lo puede señalar GEOQUAL u otras metodologías. Incluso en casos de especies de distribución costera, donde a pesar de una georreferenciación relativamente apropiada, una extracción puntual podría arrojar muchos valores “NA” (no información disponible) dado que los mapas/capas ráster de información ecogeográfica no se ajustan perfectamente al contorno de las líneas costeras.
- b) Cuando por diversas razones el sitio de recolección del germoplasma no corresponde exactamente al sitio donde la planta crece, pero éste se encuentra dentro de un perímetro relativamente conocido (por ejemplo, cuando se recolecta germoplasma en mercados locales).

En estos casos el usuario puede utilizar la extracción de tipo “radial”. En ella, el usuario indica a la herramienta el radio del área circular alrededor del punto que indica la coordenada, para la cual desea extraer información. De esta manera, ECOGEO extrae datos ecogeográficos de todas las celdas que se encuentran al interior del área circular, calcula su valor medio y asigna este valor a la entrada, repitiendo este proceso para todas las variables ecogeográficas por las que se pretende caracterizar el germoplasma (ver Fig. 32). Adicionalmente ECOGEO descarta valores “NA” para el cálculo de medias y la subsecuente asignación de valores. Para indicarle a ECOGEO que se desea realizar una extracción “radial”, el usuario debe activar el parámetro “buffy” y a continuación introducir el valor en metros del radio del área circular de extracción en el parámetro “tamp”. Este procedimiento de extracción de tipo radial está disponible también para otras herramientas.



ACCENUMB	Captured Values*	Circular extraction (mean)	Punctual extraction
A	NA,1,1	1	NA
B	NA,1,1	1	NA
C	3,2,1,3,2,3	2.333	2

*from centroids falling within circular areas

Figura 32. Diferencias entre los valores asignados a partir de una extracción de tipo puntual y una radial. Las celdas en color azul y valores NA representan cuerpos de agua y los puntos de color rojo tres sitios de recolección (identificados con códigos ACCENUMB) ubicados a partir de sus coordenadas.

Una vez el usuario ha preparado su tabla de pasaporte y de caracterización fenotípica y/o genotípica (en su caso) de acuerdo a los formatos preestablecidos, le ha indicado a la herramienta donde se encuentran (modo local) o se las ha proporcionado (modo on server), ha indicado el área de trabajo, el nivel de resolución y la forma de extracción que desea, sólo le resta definir las variables/capas de interés de cada aspecto (bioclimático, geofísico y edáfico) con las que quiere caracterizar ecogeográficamente los sitios de recolección del germoplasma.

Con la definición de estos parámetros, la herramienta ECOGEO buscará las variables/capas de información ecogeográfica de interés, las agrupará y extraerá la información para cada coordenada desde el grupo de capas en un solo paso. Con la información extraída generará una tabla que se salvará en la dirección que el usuario defina en el parámetro "resultados".

Finalmente, si el usuario estuviera interesado en realizar un análisis de agrupamiento o un Análisis de Componentes Principales (ACP), puede indicar a la herramienta que ejecute dichos análisis y además puede señalar que tipo de agrupamiento desea o el número de componentes principales que desea retener, respectivamente. La herramienta ECOGEO producirá gráficos (dendrogramas o biplots) y tablas (valores y vectores principales y puntuaciones para los componentes retenidos) que se guardarán en la carpeta indicada en el parámetro "resultados". Si se han solicitado análisis fenotípicos o genotípicos adicionales a los ecogeográficos, ECOGEO realizará los análisis multivariados y los resultados también serán guardados en "resultados".

7.3. Utilización de la herramienta ECOGEO

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta ECOGEO, se deben especificar una serie de parámetros por parte del usuario.

7.3.1 Parámetros iniciales definidos por el usuario

7.3.1.1 Parámetro: ruta (sólo aplica para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

7.3.1.2 Parámetro: país

Explicación: Seleccione el país/región en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente. También puede utilizar la herramienta rLayer para producir sus propios marcos de trabajo y seleccionarlos aquí.

7.3.1.3 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta 'Pasaporte' pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama 'tabla.txt' y se encuentra en una carpeta llamada 'Stuberosum' que se encuentra dentro de 'Pasaporte', entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

7.3.1.4 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

7.3.1.5 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

7.3.1.6 Parámetro: buffy

Explicación: Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea que la extracción de información ecogeográfica se haga sobre un área circular alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección. Esta opción es muy útil, por ejemplo, cuando la mayoría de los sitios de recolección corresponden a mercados y no directamente las fincas de los agricultores, o cuando existe bastante incertidumbre sobre la real ubicación de los sitios de recolección a pesar de que GEOQUAL pueda otorgar alta calidad de georreferenciación.

7.3.1.7 Parámetro: tamp

Explicación: Solo aplica si buffy ha sido señalado como TRUE (✓ en modo on server). Especifique el radio (en kilómetros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor. La extracción se hará sobre aquellas celdas cuyo centroide esté dentro del área circular.

7.3.1.8 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución o precisión respecto a unas coordenadas (puntos X e Y), pero exige más capacidad y tiempo de cómputo en relación a 5x5 km. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes.

7.3.1.9 Parámetro: ecogeo

Explicación: Indique aquí (con la opción TRUE en modo local o un ✓ en modo on server) si está interesado en realizar una caracterización ecogeográfica de los sitios de recolección/observación del germoplasma.

7.3.1.10 Parámetro: bioclimsn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea caracterizar por variables de tipo bioclimático (temperaturas medias, máximas o mínimas anuales o mensuales, precipitaciones mensuales o anuales, presión de vapor, etc.)

7.3.1.11 Parámetro: bioclimv

Explicación: Solo aplica si bioclimsn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables bioclimáticas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables bioclimáticas), cópielos y péguelos separados por un signo de punto y coma (;). También encontrará una línea bloqueada la selección de las 19 variables bioclim, listas para usar, simplemente eliminando el símbolo # inicial. En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro bioclimv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

7.3.1.11 Parámetro: edaphsn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea caracterizar por variables de tipo edáfico (texturas, profundidad, pH, etc.).

7.3.1.13 Parámetro: edaphv

Explicación: Solo aplica si edaphsn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables edáficas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables edáficas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro edaphv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

7.3.1.14 Parámetro: geophyssn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea caracterizar por variables de tipo geofísico (relativas al relieve y radiación solar).

7.3.1.15 Parámetro: geophysv

Explicación: Solo aplica si geophysn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables geofísicas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables geofísicas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro geophysv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

7.3.1.16 Parámetro: latitud

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la latitud (Y) como una variable geofísica más a ser analizada.

7.3.1.17 Parámetro: longitud

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la longitud (X) como una variable geofísica más a ser analizada.

7.3.1.18 Parámetro: phenotip

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de caracterización fenotípica. Para ello deberá disponer de datos de caracterización o evaluación fenotípica (morfología, fenología, productividad, resistencia, etc.) siguiendo el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama ‘fenotipos’, en este espacio debe escribir ‘fenotipos.txt’. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte, de la estructura de datos de las herramientas CAPFITOGEN3 modo local. Para el modo on server, la tabla se subirá en el momento de configuración de la herramienta o previamente, quedando disponible en la zona de archivos de usuario y resultados.

7.3.1.19 Parámetro: phenot

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización fenotípica en el formato indicado en modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama ‘fenotipos’, en este espacio debe escribir ‘fenotipos.txt’. Para el modo on server, debe hacer click en el botón que le permite buscar, seleccionar y subir el archivo correspondiente.

7.3.1.20 Parámetro: phenotv

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene la tabla que describe la naturaleza de cada variable fenotípica en el formato indicado para el modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'variablesfenotipo', en este espacio debe escribir 'variablesfenotipo.txt'. Esta tabla debe describir todas las variables incluidas en la tabla con los datos de la caracterización (ver parámetro anterior). Para el modo on server, debe hacer click en el botón que le permite buscar, seleccionar y subir el archivo correspondiente.

7.3.1.21 Parámetro: genotip

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de la caracterización genotípica. Para ello deberá disponer de datos de caracterización o evaluación genotípica (como presencia o ausencia de marcador como ceros y uno) siguiendo el formato indicado. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte (modo local) de la estructura de datos de las herramientas CAPFITOGEN. Para el modo on server, la tabla se subirá en el momento de configuración de la herramienta o previamente, quedando disponible en la zona de archivos de usuario y resultados.

7.3.1.22 Parámetro: genot

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización genotípica en el formato indicado para el modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'genotipos', en este espacio debe escribir 'genotipos.txt'. Para el modo on server, debe hacer click en el botón que le permitirá buscar, seleccionar y subir el archivo correspondiente.

7.3.1.23 Parámetro: neigd

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción si desea obtener un mapa del índice promedio de diversidad genética de Nei (1987), un mapa de la proporción promedio de marcadores polimórficos y un mapa de número de entradas analizadas por celda.

7.3.1.24 Parámetro: csimilar

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Indique la opción preferida del coeficiente de similaridad que desea utilizar para generar el mapa de distancia genotípica promedio. Tiene a disposición: 1 = Índice de Jaccard (1901), 2 = SMC de Sokal y Michaner 1958, 3 = Sokal y Sneath (1963) (S5 de Gower y Legendre), 4 = Rogers y Tanimoto (1960), 5 = Dice (1945), 6 = coeficiente de Hamann, 7 = Ochiai (1957), 8 = Sokal y Sneath (1963) (S13 de Gower y Legendre), 9 = Phi de Pearson, 10 = S2 de Gower y Legendre. La distancia (d) se obtiene como $d = \sqrt{1-s}$ siendo s el coeficiente de similaridad.

7.3.1.25 Parámetro: ecogeoclus

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de agrupamientos para las accesiones caracterizadas ecogeográficamente.

7.3.1.26 Parámetro: ecogeoclustype

Explicación: Solo aplica si ecogeoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos ecogeográficos: “single” = vecino más cercano, “complete” = vecindario más compacto, “ward” = método de mínima varianza de Ward, “mcquitty” = método de McQuitty, “average” = similaridad promedio (UPGMA), “median” = similaridad de la mediana, “centroid” = centroide geométrico, “flexible” = Beta flexible.

7.3.1.27 Parámetro: ecogeopca

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (con la opción TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de componentes principales para las accesiones caracterizadas ecogeográficamente.

7.3.1.28 Parámetro: ecogeopcaxe

Explicación: Solo aplica si ecogeopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de componentes a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCA. Esto significa que sólo los componentes principales “retenidos” serán mostrados en las tablas de resultados.

7.3.1.29 Parámetro: phenoclus

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de agrupamientos de todas las accesiones para las cuales se aporta información fenotípica.

7.3.1.30 Parámetro: phenoclustype

Explicación: Solo aplica si phenoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos fenotípicos: ‘single’ = vecino más cercano, ‘complete’ = vecindario más compacto, ‘ward’ = método de mínima varianza de Ward, ‘mcquitty’ = método de McQuitty, ‘average’ = similaridad promedio (UPGMA), ‘median’ = similaridad de la mediana, ‘centroid’ = centroide geométrico, ‘flexible’ = Beta flexible.

7.3.1.31 Parámetro: phenopca

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de componentes/coordenadas principales de todas las accesiones para las cuales se ha aportado información fenotípica.

7.3.1.32 Parámetro: phenopcaxe

Explicación: Solo aplica si phenopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de componentes/coordenadas a retener (siempre menor al número de variables fenotípicas) dentro del análisis PCA/PCoA. Esto significa que sólo los componentes principales “retenidos” serán mostrados en las tablas de resultados.

7.3.1.33 Parámetro: phenovarq

Explicación: Solo aplica si phenopca ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si la totalidad de las variables/descriptores fenotípicos corresponden a variables de tipo cuantitativo.

7.3.1.34 Parámetro: genoclus

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de agrupamientos de todas las accesiones para las cuales se aporta información genotípica.

7.3.1.35 Parámetro: genoclustype

Explicación: Solo aplica si genoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos genotípicos: ‘single’ = vecino más cercano, ‘complete’ = vecindario más compacto, ‘ward’ = método de mínima varianza de Ward, ‘mcquitty’ = método de McQuitty, ‘average’ = similaridad promedio (UPGMA), ‘median’ = similaridad de la mediana, ‘centroid’ = centroide geométrico, ‘flexible’ = Beta flexible.

7.3.1.36 Parámetro: genopco

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (con la opción TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de coordenadas principales de todas las accesiones para las cuales se ha aportado información genotípica.

7.3.1.37 Parámetro: genopcoaxe

Explicación: Solo aplica si genopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de

coordenadas a retener (siempre menor al número de variables genotípicas o marcadores) dentro del análisis PCoA. Esto significa que sólo las coordenadas principales “retenidas” serán mostrados en las tablas de resultados.

7.3.1.38 Parámetro: mantelt

Explicación: Indique (con la opción TRUE en modo local o ✓ en modo on server) si desea analizar las correlaciones matriciales (Mantel, 1967) entre las posibles combinaciones de factores (ecogeográfico vs. fenotípico vs. genotípico). Se realizarán todas las posibles comparaciones de acuerdo a si se introdujeron datos fenotípicos o genotípicos o si se creó una matriz de caracterización ecogeográfica a partir de sitios de recolección. Se generará e incluirá una matriz de distancias geográficas en las comparaciones matriciales pareadas.

7.3.1.39 Parámetro: mantelmeth

Explicación: Solo aplica si mantelt ha sido señalado como TRUE (✓ en modo on server). Seleccione el tipo de correlación a usar en la prueba de Mantel.

7.3.1.40 Parámetro: mantelper

Explicación: Solo aplica si mantelt ha sido señalado como TRUE (✓ en modo on server). Ingrese el número de permutaciones que desea para realizar la prueba de Mantel.

7.3.1.41 Parámetro: resultados (sólo aplica para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

7.4. Resultados de ECOGEO

Después de definir todos los parámetros y rutas (para modo local) que requiere ECOGEO, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, ECOGEO guardará los resultados en la ruta y carpeta que se haya determinado en “resultados” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server.

Independientemente de esto, ECOGEO puede generar hasta tres carpetas diferentes con resultados y un archivo con los parámetros utilizados para generar estos resultados (Parametros.Parameters.ECOGEO.txt). A continuación se detalla los contenidos de las carpetas.

7.4.1 Carpeta ClassicMultivariateResults_pais

7.4.1.1 Gráficas

Se trata de los archivos dendrograma_ecogeo/fenotípico/genotípico.wmf y pca_ecogeo.wmf o pco_fenotípico/genotípico.wmf, que corresponden a figuras vectoriales en formato meta-archivo de Windows. Estas figuras (un dendrograma y un biplot respectivamente) sólo se generan si se ha solicitado a la herramienta realizar un análisis de agrupamientos o un análisis de componentes principales (parámetros ecogeoclus, ecogeopca, phenocclus, phenopca, genocclus y genopco). Pueden abrirse e incluso modificarse en Microsoft Powerpoint o programas de edición de imágenes.

7.4.1.2 Las tablas

Tablas resultado del análisis de componentes principales, las cuales corresponden a los archivos ecogeographic/genotypic/phenotypic_eigenvalues.txt/xls (tabla con los valores propios), ecogeographic/eigenvectors.txt/xls (tabla con los vectores propios), ecogeographic_pcascores.txt/xls (tabla con las puntuaciones de cada entrada para los componentes principales retenidos) y/o phenotypic/genotypic_pcoscores.txt/xls (tabla con las puntuaciones de cada entrada para las coordenadas principales retenidas). Sólo se generarán si se ha solicitado a la herramienta realizar análisis tipo componentes o coordenadas principales.

7.4.2 Carpeta EcogeographicCharacterization_pais

7.4.2.1 Tabla de caracterización ecogeográfica de las entradas: Se trata del archivo TablaVarEcogeograficapais.txt/xls. Corresponde a la matriz inicial de caracterización, con tantas filas como entradas analizadas y tantas columnas como descriptores ecogeográficos.

7.4.3 Carpeta MantelCorrelationResults_pais

En esta carpeta se guardarán todas las tablas con las matrices de distancias calculadas para todas las entradas a la vez (“Matriz_distancia_”) y las tablas con los resultados de las correlaciones matriciales de Mantel (1967). El nombre de cada tabla indica la comparación de información realizada. Para las correlaciones donde intervienen datos genotípicos se usa la matriz de distancia de Dice. Por ejemplo, el archivo “Mantel_genotypic_Vs_phenotypic.txt” contiene los resultados de la correlación matricial entre las distancias genotípicas (Dice) y las distancias fenotípicas (Gower). Es importante anotar que ECOGEO también calcula la matriz de distancias geográficas (calculadas en grados decimales) para poder hacer comparaciones matriciales respecto al componente de distancia geográfica.

7.5. Referencias

Ceballos-Silva, A., López-Blanco, J. 2003. Evaluating biophysical variables to identify suitable areas for oat in Central Mexico: a multi-criteria and GIS approach. *Agriculture, Ecosystems and Environment* 95 (2003) 371–377.

Williams, C.L., Hargrove, W.W., Liebman, M., James, D.E. 2008. Agro-ecoregionalization of Iowa using multivariate geographical clustering. *Agriculture, Ecosystems and Environment* 123 (2008) 161–174



Taller Nacional, Quito, Ecuador, Agosto 2017.



8 | Herramienta Representa

8.1. Concepto de representatividad en colecciones de germoplasma

Los aspectos sensibles que pueden poner en riesgo una conservación *ex situ* de recursos fitogenéticos exitosa se pueden dar en dos momentos particulares, o en el momento de la colecta o durante la conservación propiamente dicha. A través de la aplicación de técnicas apropiadas para el manejo del germoplasma, se puede reducir el riesgo de pérdidas de entradas durante el periodo de conservación. Sin embargo, el germoplasma que se lleva a conservación debe ser el reflejo de la diversidad genética más fiel posible de las poblaciones vegetales que ocurren en el campo. Durante la conservación en el mejor de los casos y sin que medien nuevas recolecciones, ese reflejo se mantendrá intacto. Esta situación pone en evidencia la importancia de realizar recolecciones de germoplasma que aseguren capturar la mayor diversidad genética posible. Qué tanto se parece la muestra de la diversidad conservada *ex situ* respecto a la diversidad genética total que ocurre en la naturaleza, es lo que se denomina **representatividad** de una colección de germoplasma.

La representatividad de una especie dentro de una colección de germoplasma puede determinarse a nivel intra e interpoblacional. Para el caso de una especie cultivada, el equivalente sería niveles intra e intervarietales. Son dos conceptos que al considerarse de forma global la representatividad de una colección son insolubles. A pesar de esto, y por cuestiones prácticas relacionadas con la forma de conservación del germoplasma, ambos conceptos se han trabajado por separado.

La representatividad intrapoblacional ha sido exhaustivamente estudiada como en los múltiples trabajos de Crossa y colaboradores (1994, 1997, 2011), lo cual ha desembocado en estrategias de recolección específicas de acuerdo a la biología reproductiva de la especie, a la distribución espacial de los individuos y al tamaño de la población. Básicamente lo que se busca es calcular, según el caso, el número mínimo de individuos a recolectar para asegurar la captura de la mayor parte de los alelos presentes en la población. En contraste, la forma como debería representarse interpoblacionalmente una especie en una colección ha sido menos abordada. Sin embargo, desde el desarrollo del concepto de colecciones núcleo o nucleares, la representatividad interpoblacional de una especie en una colección ha ganado importancia, dado que éstas subcolecciones sólo operan a este nivel (Brown, 1989; Yonezawa *et al.*, 1995).

Una vez el concepto de representatividad de una colección de germoplasma se implanta en la comunidad de científicos y curadores en el ámbito de los recursos fitogenéticos, el siguiente paso fue determinar la forma más apropiada de calcularla. Si el objetivo de la conservación *ex situ* es capturar y mantener la mayor diversidad genética posible de una especie, la determinación ideal de la representatividad sería en términos genéticos. En consecuencia, la fórmula para determinar la representatividad genética (RG) en porcentaje sería:

$$RG = (NAC * 100) / NAT$$

Donde NAT corresponde al número total de alelos de la suma de todos los loci estudiados que presenta la especie objetivo dentro del ámbito espacial (continente, país, región, etc.) de la colección a evaluar y NAC es el número de alelos de tales loci de esta especie capturados por dicha colección. Esta determinación ideal de la representatividad genética

conlleva un impedimento de orden práctico. Conocer el número total de alelos que puede contener una especie en un territorio suficientemente grande como un país (espacio de trabajo usual de una colección de germoplasma de un Programa Nacional) o incluso a niveles muy inferiores a éste, resulta una tarea en la práctica inabordable para cualquier especie, excepto para aquellas en las que se sabe con certeza que están compuestas de muy pocas poblaciones. Si se considera el contexto de los recursos fitogenéticos para alimentación y agricultura, esta excepción es casi inexistente. Adicionalmente, el intentar calcular la RG lleva a que indirectamente se haya representado el 100% de los alelos, si el muestreo de todas las poblaciones implicase recolección de germoplasma. En otras palabras, si calcular la RG de una colección de germoplasma implica recolectar muestras y germoplasma de todas las poblaciones que componen la especie dentro de un territorio de trabajo, entonces, independientemente de lo difícil que pueda resultar esta tarea, la máxima representatividad ya sería alcanzada siempre que en la recolección se hayan seguido criterios adecuados de representatividad intrapoblacional.

Estas dificultades prácticas y logísticas han llevado a pensar en otras alternativas para determinar la representatividad de una colección. La representatividad ecogeográfica (RE) fue planteada por Parra-Quijano y colaboradores (2008) en el caso especial de colecciones *ex situ* de plantas silvestres emparentadas con cultivadas (CWR en sus siglas en inglés). En ella se plantea la posibilidad de usar mapas de caracterización ecogeográfica del territorio (como los producidos por ELC mapas) para conocer que tantas condiciones ambientales presentes en un marco espacial estarían representados en una colección de germoplasma.

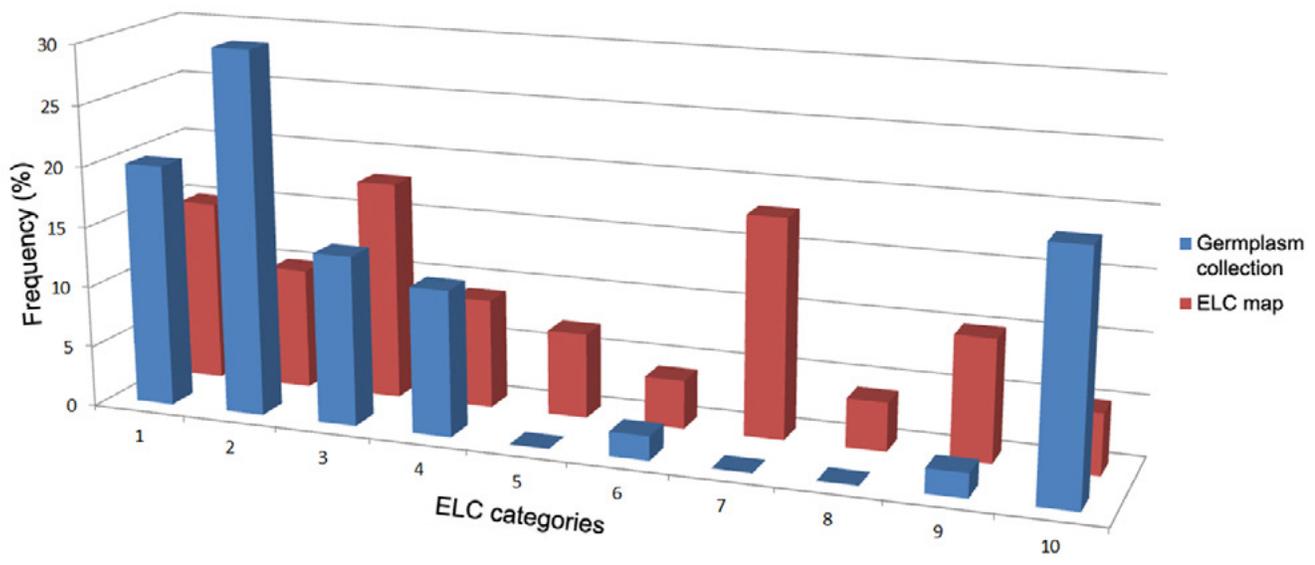


Figura 33. Comparación de la representación de cada categoría ELC en la colección de germoplasma y la disponibilidad total de dichas categorías en el mapa ELC, medida en valores de frecuencias (en porcentaje).

Como ejemplo de dicha aplicación, la Fig. 33 muestra una distribución de las frecuencias para cada categoría ELC de una colección de germoplasma en contraste con la disponibilidad de dichas categorías en el total del marco espacial. Este ejemplo ficticio sirve para mostrar cómo la representatividad de una colección puede o no (como es el caso) estar

sesgada respecto a la abundancia de unidades ambientales presentes en el espacio de trabajo. El contraste de valores para las categorías 2 y 7 dejan entrever que las dos distribuciones no guardan parecido y posiblemente una prueba de Chi cuadrado determinaría una asociación no significativa de las dos distribuciones. Sin embargo, la determinación más exacta de la RE se consigue a través de un análisis de faltantes o vacíos (más conocido por su denominación en inglés como “gap analysis”). Para realizarlo es necesario previamente compilar información de otras fuentes externas a la colección, como otras colecciones de germoplasma o cualquier dato que indique la presencia de poblaciones de la especie objetivo (pliegos de herbario, bases de datos botánicas, referencias bibliográficas, etc.). Luego se compara la distribución de la frecuencia de las recolecciones de la colección a evaluar respecto a la de las fuentes externas y entonces se podrá visualizar claramente qué ambientes están subrepresentados en la colección.

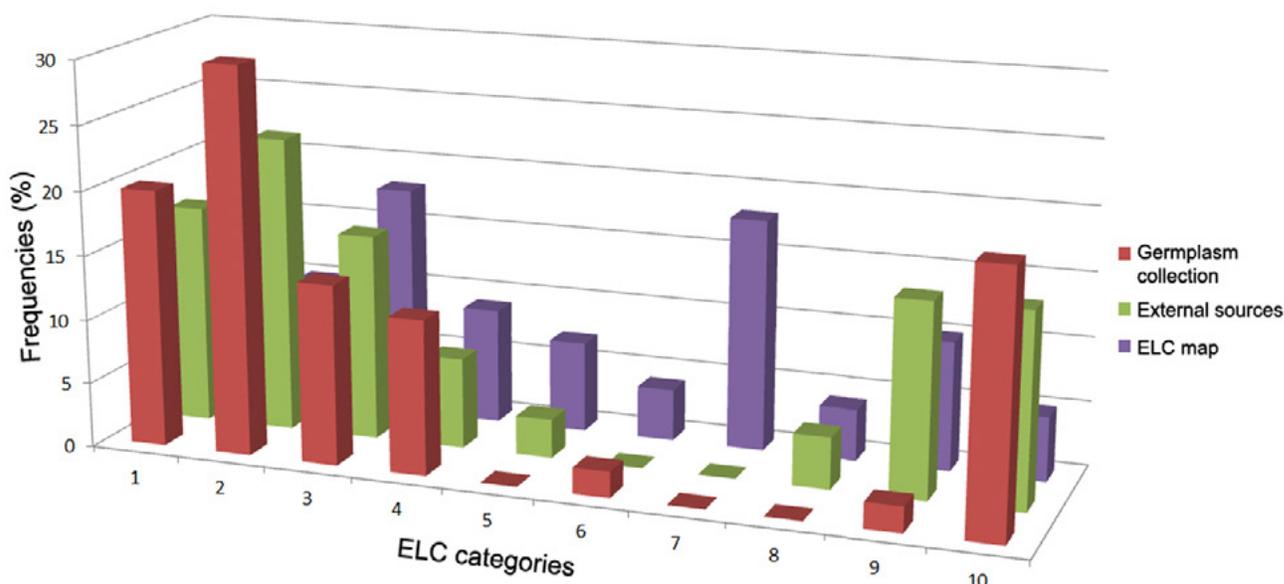


Figura 34. Comparación de distribución de las frecuencias de sitios de recolección de la colección objetivo y de presencias de fuentes externas sobre diez categorías ELC. Se incluye también la distribución de la frecuencia de cada categoría en el total del mapa ELC.

La Fig. 34 ilustra el proceso de comparación previamente mencionado. Usando los mismos datos ficticios del ejemplo de la Fig. 33, en esta gráfica de barras se incluye (en color verde) la distribución de las frecuencias de las categorías ELC para datos de presencia provenientes de fuentes externas. En este caso es evidente el parecido entre las distribuciones colección objetivo y fuentes externas, aunque algunas diferencias son especialmente interesantes. Para las categorías 5 y 8, las fuentes externas indican presencia de la especie en esa unidad ambiental la cual no estaría representado en la colección, evidenciando la presencia de faltantes o vacíos ecogeográficos. Estos vacíos pueden ser útiles para la planificación de nuevas recolecciones de germoplasma, pues se puede priorizar la visita de estos ambientes pues se conoce donde ocurren estas poblaciones provenientes de fuentes externas. Es importante hacer una aclaración respecto al tratamiento que se puede dar a los datos provenientes de fuentes externas. Tomando las presencias provenientes de otro banco de germoplasma como fuente externa, se puede conocer la representatividad

de la colección objetivo globalmente, pero usar dichos datos para determinar sitios prioritarios para nuevas recolecciones puede llevar a recolectar duplicados inter-colecciones.

8.3. Utilización de la herramienta Representa

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Representa, se deben especificar una serie de parámetros por parte del usuario. Note que para que esta herramienta funcione correctamente, antes el usuario debe haber obtenido un mapa ELC con la herramienta ELCmapas y los siguientes archivos producto de esta herramienta deben ser copiados y pegados en la ruta CAPFITOGEN3/ELCmapas/ en el modo local:

- mapa_elc_pais.grd
- mapa_elc_pais.gri
- Estadist_ELC_pais.txt

Siendo “pais” el nombre del país o la región para la cual se realizó el mapa. Para modo on server, estos elementos deben ser subidos a través del botón frente a cada parámetro, que permite seleccionar el archivo localmente, para luego subirlo al servidor.

8.3.1 Parámetros iniciales definidos por el usuario

8.3.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

8.3.1.2 Parámetro: internet

Explicación: Seleccione esta opción si desea utilizar Global Biodiversity Information Facility (GBIF) como fuente externa, conectando directamente a su servidor, para transferir a CAPFITOGEN3 la información de la especie seleccionada.

8.3.1.3 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stube-

rosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

8.3.1.4 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

8.3.1.5 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

8.3.1.6 Parámetro: fext

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si dispone de ocurrencias provenientes de fuentes externas (cualquier fuente de información diferente a la colección objeto del análisis de representatividad) en el formato requerido.

8.3.1.7 Parámetro: fuentex

Explicación: Solo aplica si fext ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo que contiene las ocurrencias de fuentes externas en el formato requerido. Si el archivo se llama "FuentesExternas", en este campo debe aparecer "FuentesExternas.txt" (porque esta tabla debe estar en formato de texto delimitado por tabulaciones). No olvide que este archivo debe encontrarse en la carpeta llamada Pasaporte (CAPFITOGEN3/Pasaporte) para modo local o se deben seleccionar y subir al servidor (a través del botón frente al parámetro).

8.3.1.8 Parámetro: geoqualfe

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si su tabla de ocurrencias de fuentes externas fue evaluada por GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, disponiendo en las columnas indicadas de información sobre evaluación de la calidad de la georreferenciación.

8.3.1.9 Parámetro: totalqualfe

Explicación: Solo aplica si geoqualfe ha sido señalado como TRUE (✓ en modo on server). Determine el valor de TOTALQUAL100 que incluye la tabla de fuentes externas a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

8.3.1.10 Parámetro: duplibg

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si considera como faltante en el banco de germoplasma objetivo, las ocurrencias de fuentes externas que provienen de otros bancos o colecciones de germoplasma (Columna TYPESOURCE con valor 40 en la tabla de fuentes externas). Tenga en cuenta que, al marcar esta opción Ud. podría llegar a realizar recolecciones de poblaciones que otras colecciones ya tienen representadas, generando duplicación inter-colecciones. Precaución: Configurar duplibg como TRUE o ✓ cuando todos los registros de la tabla de fuentes externas vienen de bancos de pasaporte (TYPESOURCE=40) generará un error, pues todos ellos serán descartados en el análisis de representatividad.

8.3.1.11 Parámetro: gbifFE

Explicación: Solo aplica si el parámetro “internet” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea descargar datos de ocurrencia de fuentes externas desde el portal de Global Biodiversity Information Facility (GBIF). Esta opción requiere conexión a internet en conexión. Esta opción es incompatible con la inclusión de ocurrencias de fuentes externas aportadas por el usuario. En caso de marcar esta opción y aportar a la vez una tabla con ocurrencia de fuentes externas, solo se tendrá en cuenta esta última.

8.3.1.12 Parámetro: genero

Explicación: Solo aplica si el parámetro “gbifFE” ha sido señalado como TRUE (✓ en modo on server). Escriba el nombre del género de la especie objetivo del presente análisis. Este será el género para el cual se descargará información desde el portal de GBIF. No olvide primera letra del género debe ir en mayúscula.

8.3.1.13 Parámetro: especie

Explicación: Solo aplica si el parámetro “gbifFE” ha sido señalado como TRUE (✓ en modo on server). Escriba el nombre de la especie (solo el epíteto) objetivo del presente análisis. Este nombre se unirá al del género para realizar la solicitud y descarga de información desde GBIF. Si desea descargar información para todo el género, solamente escriba aquí un *. Todo el epíteto debe estar escrito en minúsculas.

8.3.1.14 Parámetro: mapaelc

Explicación: Indique el nombre del archivo que contiene el mapa ELC (producto de la aplicación de la herramienta

ELCmapas), el cual debe encontrarse en la carpeta ELCmapas, que corresponde a una de las carpetas que componen el directorio CAPFITOGEN3 (modo local). Este mapa debe estar en formato DIVA-GIS, compuesto por los dos archivos con extensiones “.grd” y “.gri”, tal y como lo produce ELCmapas. En esta caja de texto se debe escribir el nombre del archivo con la extensión “.grd”. Así, si el nombre del mapa es “mapa_elc_spain”, ud. debe escribir “mapa_elc_spain.grd”. Para el modo on server, seleccione y suba el archivo del mapa ELC en el servidor haciendo click con el botón en frente del parámetro en el momento de utilizar la herramienta, o previamente, guardando los archivos .grd y .gri en la zona de archivos de usuario y resultados.

8.3.1.15 Parámetro: **statelc**

Explicación: Indique el nombre del archivo que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (esta herramienta suele nombrar a este archivo como “Estadist_ELC_” más el nombre del país) para el modo local. Al igual que el mapa ELC, este archivo debe estar situado en la carpeta ELCmapas (CAPFITOGEN3/ELCmapas). De igual forma el nombre debe escribirse junto con la extensión, en este caso al ser una tabla, la extensión es txt. Así si el archivo se llama “Estadist_ELC_spain” se debe escribir “Estadist_ELC_spain.txt”. Para el modo on server, seleccione y suba el archivo .txt en el servidor haciendo click con el botón en frente del parámetro en el momento de utilizar la herramienta, o previamente, guardando el archivo .txt en la zona de archivos de usuario y resultados.

8.3.1.16 Parámetro: **distdup**

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas. La determinación de la distancia depende de condiciones biológicas (flujo de genes) y espaciales (tamaños promedios de las poblaciones). Se trata de un parámetro específico para la especie objetivo y muchas veces será necesario consultar el concepto de un experto.

8.3.1.17 Parámetro: **resultados (sólo para modo local)**

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

8.4. Resultados de Representa

En la ruta y carpeta que haya determinado en “resultados” (modo local) o en la zona de archivos de usuario y resultados (modo on server) encontrará hasta cinco mapas y hasta cinco tablas.

8.4.1 Los mapas

Se trata de dos mapas de puntos de tipo vectorial (shapefiles) y tres mapas ráster (formatos .grd y.tif) que pueden

visualizarse directamente en DIVA-GIS o en otros software SIG. Si no se introducen datos de fuentes externas, la lista de mapas que se producirán se reduciría a tres.

8.4.1.1 “mapa_Class_ELC”. Corresponde a un mapa que agrupa en cuatro las categorías originales del mapa ELC. Estos grupos corresponden a la partición de las categorías según su frecuencia en la totalidad del territorio. La frecuencia se divide a partir de cuartiles. El grupo señalado como 1 corresponde a la frecuencia baja (por debajo del cuartil 0.25), el grupo 2 corresponde a la frecuencia media-baja (entre cuartiles 0.25 y 0.5 o mediana), el grupo 3 corresponde a la frecuencia media-alta (entre cuartiles 0.5 o mediana y 0.75) y grupo 4 que corresponde a frecuencia alta (por encima del cuartil 0.75). Cuando este mapa en formato .grd se abre en DIVA-GIS, por defecto se muestran cinco colores, por lo que es importante cambiar la visualización y que sólo se muestren cuatro colores, cada uno abarcando un rango que cubra cada valor de grupo (1 a 4). Este cambio se ilustra en la Fig. 35.

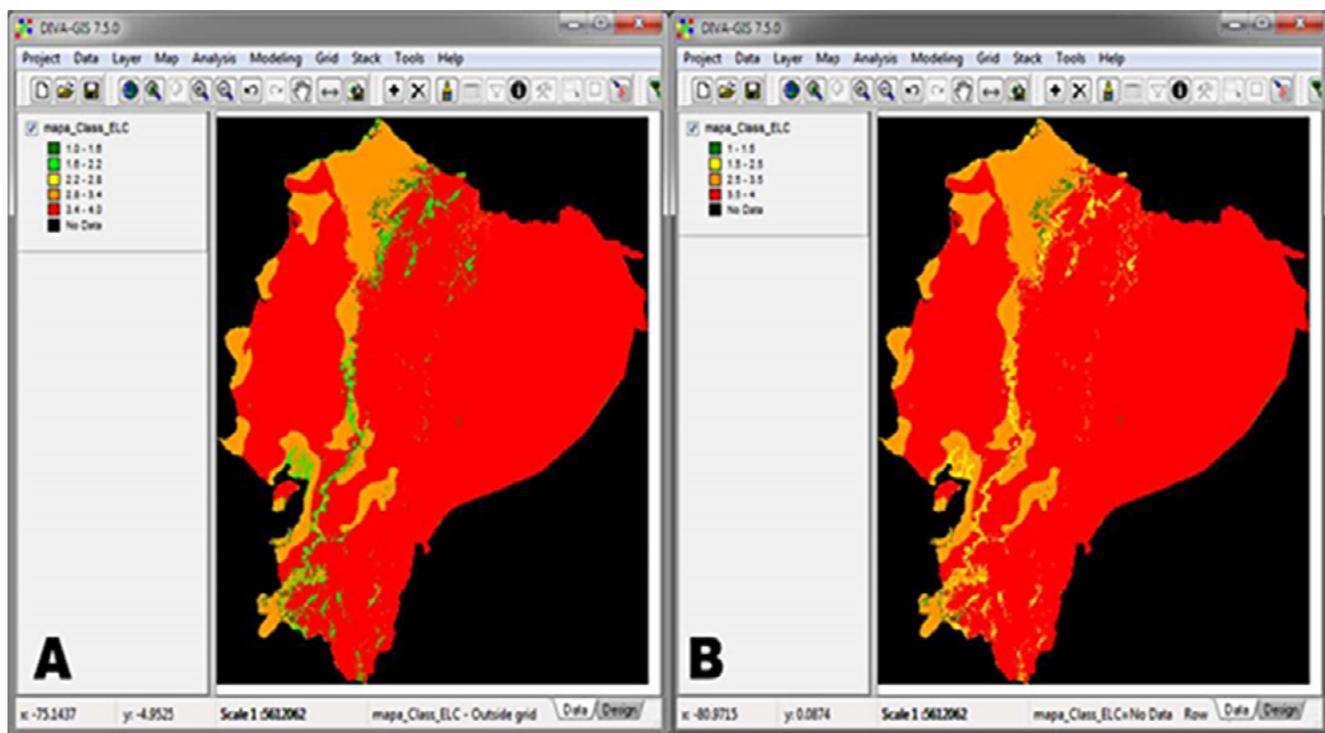


Figura 35. Configuración visual apropiada para mapas ráster de Representa. A) Visualización mapa_Class_ELC.grd tal y como se abre en DIVA-GIS, B) visualización ajustada a cuatro colores por cada grupo de frecuencia. Los escenarios adaptativos menos frecuentes (grupos baja y media baja frecuencia) aparecen en colores verde y amarillo.

8.4.1.2 “mapa_Class_Sp”. Corresponde a un mapa que agrupa en cuatro las categorías originales del mapa ELC. Estos grupos corresponden a la partición de las categorías ELC según la frecuencia en la ocurren las entradas. La frecuencia se divide a partir de cuartiles. El grupo señalado como 1 corresponde a la frecuencia baja (por debajo del cuartil 0.25), el grupo 2 corresponde a la frecuencia media-baja (entre cuartiles 0.25 y 0.5 o mediana), el grupo 3 corresponde a la frecuencia media-alta (entre cuartiles 0.5 o mediana y 0.75) y grupo 4 que corresponde a frecuencia alta (por encima del cuartil 0.75).

8.4.1.3 “mapa_Tipo_faltante”. Este mapa es otra reclasificación de las categorías originales del mapa ELC. Este mapa sólo se produce cuando el usuario introduce datos de fuentes externas. Esta reclasificación corresponde a unos criterios expuestos en la tabla siguiente:

Tabla 1. Clasificación de categorías de mapa ELC de acuerdo a criterios de prioridad para futuras exploraciones.

Clase	Diferencia Fuentes externas vs Banco de germoplasma (DIF) ¹	Clasificación frecuencia por ocurrencia de la especie ²	Clasificación por frecuencia de categoría en mapa ELC ³
0	No aplica	No aplica	No aplica
1	1	Baja o media baja	Baja o media baja
2	1	Baja o media baja	Media alta o alta
3	1	Media alta o alta	Baja o media baja
4	1	Media alta o alta	Media alta o alta
5	0.99-0.5	Baja o media baja	Baja o media baja
6	0.99-0.5	Baja o media baja	Media alta o alta
7	0.99-0.5	Media alta o alta	Baja o media baja
8	0.99-0.5	Media alta o alta	Media alta o alta
9	0.01-0.499	Baja o media baja	Baja o media baja
10	0.01-0.499	Baja o media baja	Media alta o alta
11	0.01-0.499	Media alta o alta	Baja o media baja
12	0.01-0.499	Media alta o alta	Media alta o alta
13	0 y NA	No aplica	No aplica

Estas clases están relacionadas con la prioridad de visita o exploración que tendría cada categoría ecogeográfica en una futura recolección. La clase 1 agrupa las categorías que tendrían la máxima prioridad, mientras que la clase 2 tendría menos prioridad que la clase 1, y así consecutivamente hasta la clase 13.

1 Este valor sale de la comparación entre ocurrencias de fuentes externas y bancos/colecciones de germoplasma en cada categoría usando la siguiente fórmula $DIF = FE / (FE + BG)$ siendo FE el número de ocurrencias de fuentes externas y BG las correspondientes a banco de germoplasma.

2 Esta clasificación es la misma que se muestra en el mapa 8.4.1.2

3 Esta clasificación es la misma que se muestra en el mapa 8.4.1.1

El mapa, al abrirse en DIVA-GIS no muestra las 13 clases con un color individual para cada clase, sino que agrupa los 13 valores en cinco colores. Su visualización correcta se logra en DIVA-GIS cuando se añaden 8 colores más y se ajustan los rangos de valores de cada color (como en los mapas anteriores) al valor de una clase.

8.4.1.4 “Shapefile_Puntos_BG.shp”. Mapa vectorial (shapefile) que representa los sitios de recolección de germoplasma del banco o la colección bajo evaluación de representatividad. La tabla que acompaña este mapa de puntos contiene todos los campos del formato de pasaporte FAO/Bioversity 2012.

8.4.1.5 “Shapefile_FE_class.shp”. Mapa vectorial (shapefile) que representa las ocurrencias de fuentes externas. La tabla que acompaña este mapa de puntos presenta los siguientes campos adicionales al formato de datos de fuentes externas:

FE_cat: Categoría del mapa ELC en la que ocurre la presencia

FE_BG_dif: Valor DIF (ver Tabla 1) para la categoría ELC en la que ocurre la presencia.

Class_Sp: Indica el cuartil al que pertenece la categoría en la que ocurre la presencia de fuente externa, de acuerdo a la frecuencia de la especie.

Class_ELC: Indica el cuartil al que pertenece la categoría en la que ocurre la presencia de fuente externa, de acuerdo a la frecuencia de la misma categoría en el mapa ELC.

Tipo_falt: Indica a que clase pertenece la categoría en la que ocurre la presencia de fuente externa de acuerdo a la clasificación de la Tabla 1.

8.4.2 Las tablas

De igual manera que con los mapas, la lista de tablas puede reducirse de cinco a cuatro, dependiendo si el usuario introduce o no datos de fuentes externas.

8.4.2.1 “Tabla_Fuentes_Externas_clasificadas_ExternalSourcesClassified”. Esta, como el resto de las tablas en CAPFITOGEN3, se ofrece en formato de texto separado por tabulaciones con extensiones .txt y .xls. Corresponde a la misma tabla acompañante del shapefile del apartado 8.4.1.5, conteniendo las mismas variables.

8.4.2.2 “Tabla_Resultados_Representatividad_RepresentativenessResults”. En esta tabla se presentan los resultados finales de evaluación de la representatividad, se hayan o no incluido datos de fuentes externas. Con esta tabla es posible crear en Excel las gráficas de barras como las de las Fig. 33 y 34. Finalmente en esta tabla se presenta toda la información requerida para el cálculo de los parámetros de la Tabla 1, incluyendo el valor de clase con el que se definen las prioridades.

8.4.2.3 “Tabla_Resultados_X2_Results”. Esta tabla muestra los resultados de la prueba de Chi cuadrado para determinar el grado de asociación entre dos distribuciones. Si se han introducido datos de fuentes externas, esta tabla contendrá los resultados de dos pruebas de Chi cuadrado: distribución banco/colección (o BG) vs. fuentes externas (FE) y banco/colección vs. distribución de frecuencias totales de las categorías en el mapa ELC.

8.4.2.4 “TablaClasificacionCuartilesEspecie_QuartileClassificationSpecies” y “TablaClasificacionCuartilesMapaELC_QuartileClassificationELCmap”. Estas dos tablas muestran los valores de los cuartiles 0.25, 0.5 (mediana) y 0.75 para la distribución de frecuencias de la especie y de las categorías del mapa ELC.

8.4.2.5 “Genbank_ELC”. Esta tabla corresponde a la misma información introducida en el formato de pasaporte más la columna ‘BGcat’ donde se puede ver la categoría del mapa ELC que le corresponde a cada entrada de acuerdo a su localización.

8.5. Referencias

Brown, A.H.D. 1989. The case for core collections. In: Brown, A.H.D., Frankel, O.H., Marshall, D.R., Williams, J.T. (ed.) The use of plant genetic resources. Cambridge University Press, Cambridge, UK.

Crossa, J., Vencovsky, R. 1994. Implications of the variance effective population size on the genetic conservation of monoecious species. *Theoretical and Applied Genetics* 89:936–942

Crossa, J., Vencovsky, R. 1997. Variance effective population size for two-stage sampling of monoecious species. *Crop Science* 37:14–26

Crossa, J., Vencovsky, R. 2011 Chapter 5: Basic sampling strategies: theory and practice. In: Guarino, L., Ramanatha Rao, V., Goldberg, E. (ed.) *Collecting Plant Genetic Diversity: Technical Guidelines – 2011 Update*. Bioversity International. Available online (accessed 6 November 2013) http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=671

Parra-Quijano, M.; Draper, D.; Torres, E., Iriondo, J.M. 2008. Ecogeographical representativeness in crop wild relative *ex situ* collections. p. 249-273. In Maxted, N.; Ford-Lloyd, B.V.; Kell, S.P.; Iriondo, J.M.; Dulloo, M.E., Turok, J. (ed.) *Crop wild relative conservation and use*. CAB International, Wallingford.

Yonezawa, K.; Nomura, T., Morishima, H. 1995. Sampling strategies for use in stratified germplasm collections. P. 35-53. In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L., Morales, E.A.V. (ed.) *Core collections of plant genetic resources*. John Willey & sons, Chichester, UK.



Taller Regional, Bogotá, Colombia, Marzo 2013.



9 | Herramienta DIVmapas

9.1. Representación espacial de la diversidad local

En 2012 fue publicado un estudio sobre la presentación de patrones espaciales de diversidad genética a partir de marcadores neutrales tipo microsatélite para el caso de *Annona cherimola* (van Zonneveld *et al.*, 2012). En este estudio se propone una forma diferente de mostrar cómo está distribuida la diversidad genotípica, basado en la estimación de parámetros propios de genética de poblaciones, pero que antes de ser aplicados a todas las muestras a la vez, en este caso la diversidad se estima a nivel local con la determinación de vecindades o áreas de influencia. El resultado de juntar los resultados de cada vecindad fue un mapa que muestra claramente donde se localizan los “puntos calientes” de diversidad. Las aplicaciones de esta metodología en la conservación *ex situ* e *in situ* de los recursos fitogenéticos (variedades locales y especies silvestres emparentadas con cultivadas) son evidentes.

Esta no ha sido la primera aproximación SIG o de geoestadística en el análisis de la diversidad genética, pues previamente se han realizado interpolaciones de datos genéticos (Hoffman *et al.*, 2003). Sin embargo, la metodología de van Zonneveld y colaboradores representa una aplicación muy práctica y más sencilla en su análisis e interpretación.

Posteriormente Thomas y colaboradores (2012) aplicaron esa misma metodología para 993 individuos caracterizados por microsatélites de cacao (*Theobroma cacao*), junto con otros análisis, para identificar procesos evolutivos en esta planta cultivada.

A partir de la publicación de estos desarrollos fue posible entender paso a paso cómo opera el proceso de obtención de un mapa de este tipo. La metodología era evidentemente repetible y sólo varía el parámetro genético que se calcula entre las muestras que componen una vecindad. Así, si el parámetro expresa que tan diferente genéticamente son las muestras de una vecindad, el mapa podría denominarse como “mapa de diversidad”. A partir de esta base, fue posible desarrollar la herramienta “DIV mapas” y generalizar su aplicación más allá de datos de caracterización genotípica.

Es muy importante destacar que estos mapas muestran diversidad genotípica a nivel intraespecífico, un aspecto que los diferencia notablemente de los mapas de riqueza de especies o mapas de diversidad filogenética, los cuales trabajan a nivel interespecífico.

Mostrar la diversidad en forma de mapas tiene múltiples ventajas respecto a las formas en las que usualmente se presentan estos resultados. Los mapas de diversidad, basados en el desarrollo original de van Zonneveld y colaboradores (2012) permiten identificar de manera sencilla y rápida, las zonas o regiones donde se concentra una alta variabilidad. Un mapa de estas características se convierte en una poderosa herramienta para la toma de decisiones en materia de conservación *ex situ* e *in situ*.

9.1.1 ¿Por qué un mapa de la diversidad ecogeográfica?

La diversidad ecogeográfica de un conjunto de accesiones es una medida de las diferencias que ocurren entre los escenarios adaptativos de donde provienen dichas accesiones, o en otras palabras, los sitios de recolección. Se recurre

al término “escenario adaptativo” antes que “ambiente”, porque para el cálculo de la diversidad ecogeográfica sólo se considerarían las características ambientales de tipo abiótico de mayor influencia en la distribución y ocurrencia de la especie objetivo, no todas las características ambientales disponibles.

La diversidad ecogeográfica, como otro tipo de diversidad, se determina a partir de datos de caracterización del germoplasma. La caracterización ecogeográfica se realiza mediante la extracción de información para cada coordenada mediante la utilización de un SIG, el cual ha sido previamente cargado con capas de información ambiental.

La visualización de la diversidad ecogeográfica en forma de mapa similar a los que ha desarrollado van Zonneveld y colaboradores (2012), facilita la comparación entre zonas o regiones en base a la diferencia de los escenarios adaptativos donde ocurren las entradas. Las zonas o regiones donde ocurren mayores diferencias se pueden traducir directamente en zonas donde cabe esperar germoplasma con adaptaciones más divergentes e indirectamente una posible ocurrencia de una mayor diversidad genotípica o fenotípica. Evidentemente la determinación de zonas con mayor diversidad genotípica o fenotípica es óptima cuando se realiza a través de datos de caracterización genotípica y fenotípica respectivamente. Sin embargo, a falta de éstos, un mapa de diversidad ecogeográfica puede servir de solución intermedia mientras se consiguen caracterizar genotípica y/o fenotípicamente las entradas. En cualquier caso, el escenario ideal para el análisis de la diversidad bajo esta nueva metodología, es cuando se pueden obtener mapas para los tres tipos de caracterización, dado que su contraste ofrece una visión biológica muy completa de la situación de los recursos fitogenéticos que ocurren dentro de un marco de trabajo.

9.2. Procedimiento de obtención de mapas de diversidad de DIVmapas

DIVmapas es una aplicación inspirada en la aplicación de van Zonneveld y colaboradores (2012) para chirimoya (*Annona cherimola*), pero incluye algunas diferencias respecto a la metodología original las cuales son más evidentes al contrastar los dos procesos. En este apartado se mostrará paso a paso como DIVmapas logra obtener mapas de diversidad.

DIVmapas determina medidas de diversidad local, es decir, entre las accesiones recolectadas en una zona con forma de cuadrícula con determinado tamaño y sus vecindades (zona de influencia), usando como insumo información ecogeográfica, fenotípica o genotípica. Nótese que a partir de este punto nos referiremos a entradas y no a muestras, dado que la herramienta está pensada para ser utilizada en el entorno de recursos fitogenéticos, lo cual no implica que no pueda usarse en otros ámbitos biológicos. Como resultado, DIVmapas ofrece una ilustración gráfica que refleja los valores de las medidas de diversidad distribuidas en un espacio a manera de mapa, el cual facilita la visualización de zonas calientes de diversidad genética.

Para empezar, es importante anotar que DIVmapas así como otras herramientas incluidas en el presente manual y muchas otras aplicaciones de SIG y ecogeografía en recursos fitogenéticos, requiere que cada entrada este georreferenciada y que además lo esté de manera adecuada. El apartado 3 de este manual hace referencia a la herramien-

ta GEOQUAL que permite conocer la calidad de la georreferenciación del sitio de recolección del germoplasma. Es aconsejable usar esta herramienta antes de usar DIVmapas, de manera que sólo entradas con suficiente calidad en su georreferenciación serán tenidas en cuenta en la obtención de mapas de diversidad. En cualquier caso, entradas sin coordenadas (campos DECLATITUDE o LATITUDE y DECLONGITUDE o LONGITUDE) no podrán ser incluidas en el análisis que realiza DIVmapas.

El segundo punto importante es que, si se necesita obtener mapas de diversidad fenotípicos o genotípicos, los datos de caracterización de cada tipo deben ser arreglados según el formato que se suele proveer en la carpeta "Formatos" (archivos .xls de Excel). Si el tipo de mapa de diversidad requerido es de tipo ecogeográfico, DIVmapas incluye el mismo proceso de caracterización ecogeográfica de germoplasma que la herramienta ECOGEO (capítulo 5), por lo tanto para este aspecto no es necesario preparar tablas o matrices de datos de caracterización sino simplemente señalar las variables ecogeográficas por las cuales se quiere caracterizar las entradas.

DIVmapas aprovechará toda la información de caracterización disponible y válida, y en base a ella realizará los mapas de diversidad para cada aspecto individualmente. En ese sentido, el conjunto de entradas caracterizadas genotípica, fenotípica o ecogeográficamente pueden coincidir (lo cual favorece la interpretación de resultados) o no. Lo que es imprescindible es que los códigos de identificación de las entradas en las tablas de caracterización genotípica o fenotípica, deben estar incluidos en la tabla de pasaporte FAO/Bioversity 2012 que contiene la información de georreferenciación de los sitios de recolección.

Una vez claras estas condiciones, los siguientes puntos muestran cómo DIVmapas obtiene los mapas de diversidad, independientemente de los datos de caracterización que use para ello.

9.2.1. Distribución de sitios de recolección y generación de retícula o malla

A partir de las coordenadas de cada sitio de recolección se genera un espacio de trabajo (x-min, y-min, x-max e y-max siendo x latitud e y longitud), sobre el cual se sobrepone una retícula o conjunto de celdas de forma cuadrada de un tamaño definido por el usuario (ver Fig. 36). Adicionalmente se carga una capa con los centroides de cada celda de la retícula. Cada centroide tiene un código de identificación.

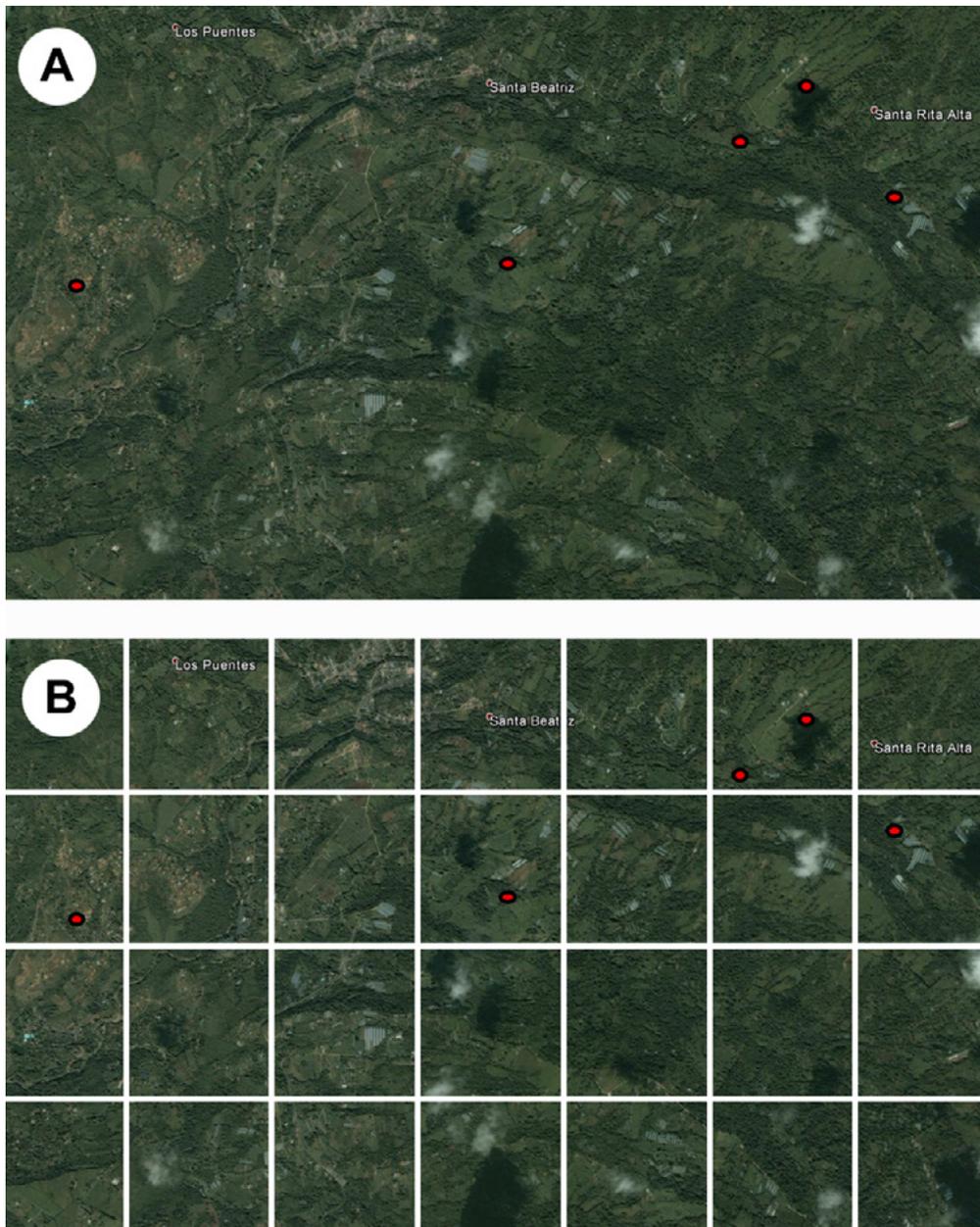


Figura 36. Primer paso. A) Distribución espacial de los sitios de recolección, B) superposición de retícula de dimensiones de celda (resolución) seleccionadas por el usuario.

9.2.2. Selección de celdas con entradas y celdas vecindario

Del total de celdas de la retícula se seleccionan aquellas donde ocurren entradas. Adicionalmente el usuario determina un área de influencia, que señala como el radio de un área circular, la cual está relacionada con la biología reproductiva de la especie, el flujo genético y el manejo y la dispersión que puede generar el hombre, especialmente si se trata de una forma cultivada. Esta área de influencia se utilizará para determinar las celdas vecindario, que son

las celdas donde no ocurren entradas pero que se encuentran suficientemente cerca a las inicialmente seleccionadas (celdas con entradas). Para ser celda vecindario, el centroide de ésta debe caer dentro de la proyección del área circular de influencia que se traza a partir del centroide de cada celda con entradas. El proceso de selección de celdas con entradas y de celdas vecindario se muestra en la Fig 37.

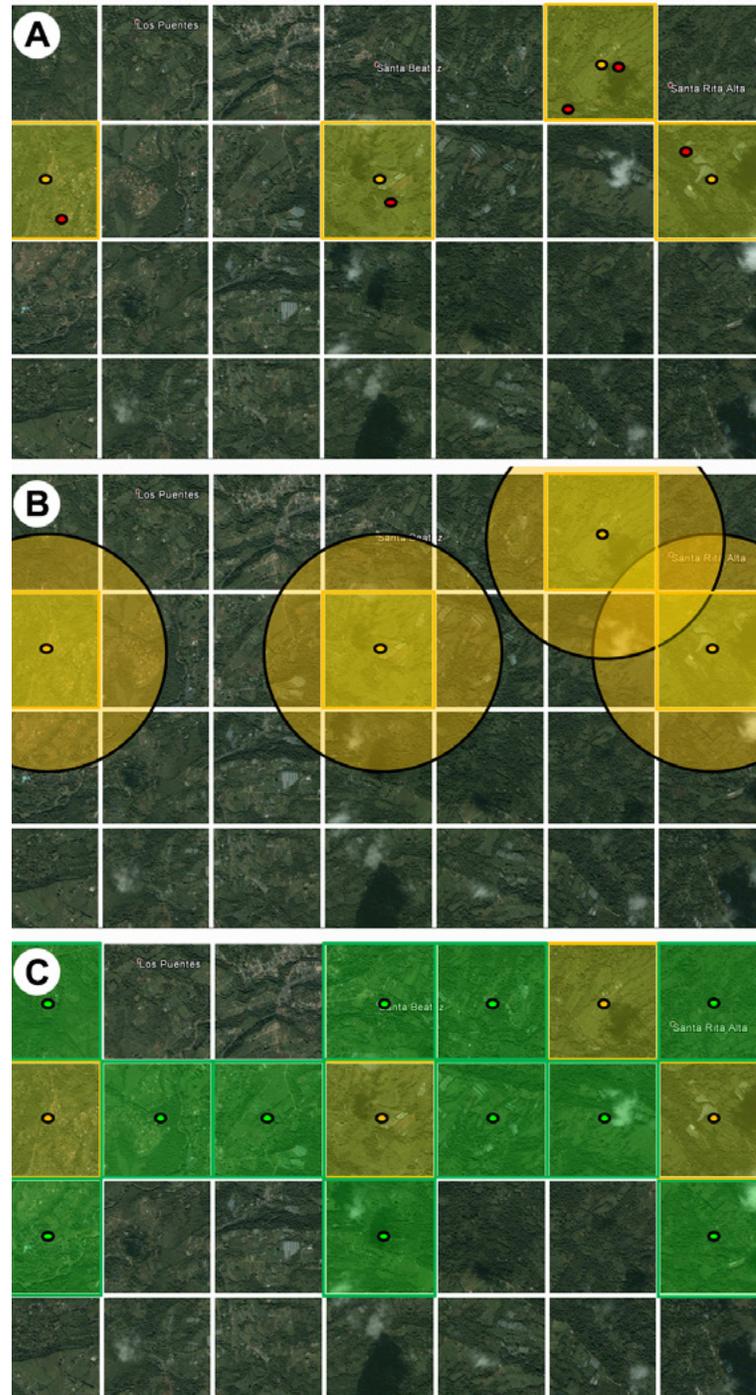


Figura 37. Segundo paso. A) Determinación de celdas con entradas y sus centroides, B) proyección de las áreas de influencia desde los centroides de celdas con entradas, C) determinación de celdas vecindario.

9.2.3. Determinación de entradas asociadas a celda con entradas y celdas vecindario

A partir de los centroides de las celdas con entradas y celdas vecindario se proyectan de nuevo las áreas circulares de influencia y a la lista resultante de entradas que caen en cada área se le asigna el código de identificación de su respectivo centroide (ver Fig. 38).

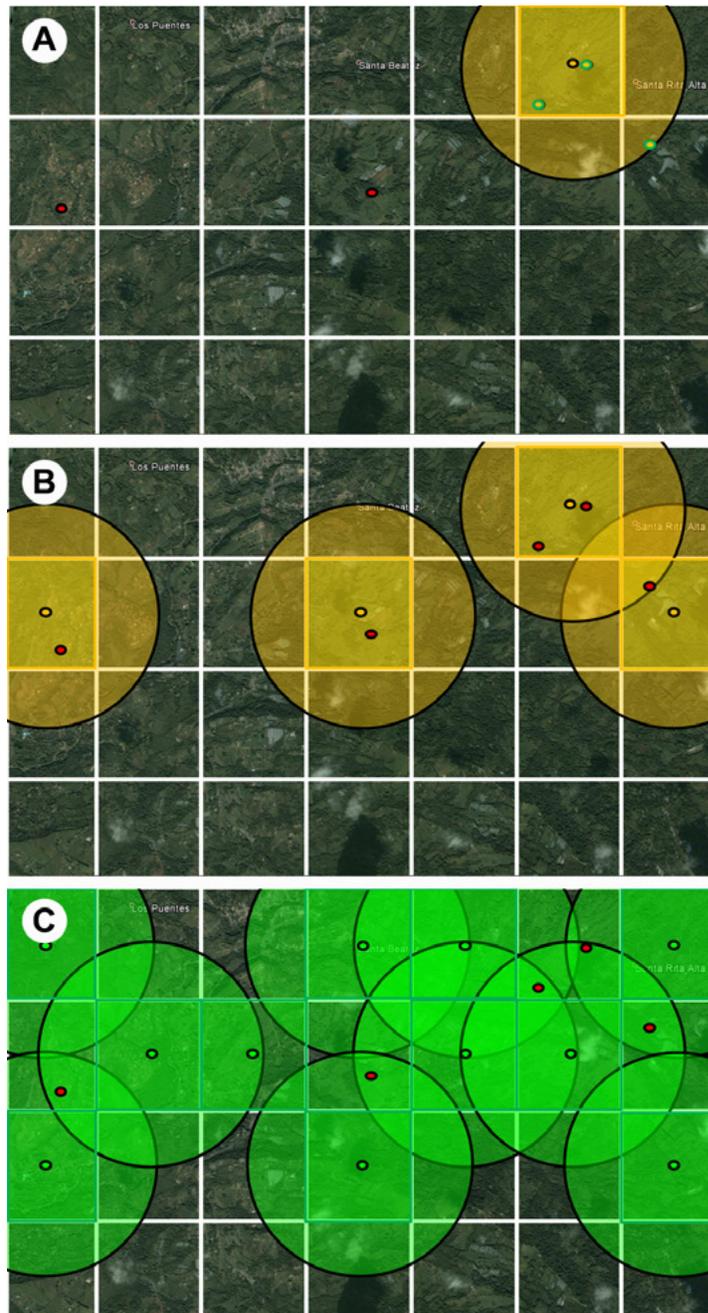


Figura 38. Tercer paso. A) Determinación de las entradas que ocurren dentro del área de influencia de una celda en singular, B) determinación de entradas que ocurren dentro de áreas de influencia de celdas con entradas, C) determinación de entradas que ocurren dentro de áreas de influencia de celdas vecindario.

9.2.4. Obtención de los mapas finales de diversidad

La lista de entradas por celda servirá para obtener matrices iniciales, con los datos de caracterización fenotípica, genotípica o ecogeográfica (según los datos que haya ingresado el usuario) como columnas y las entradas de cada centroide identificadas por su valor de ACCENUMB como filas. De esta forma se obtendrán tantas matrices iniciales como celdas con entradas y celdas vecindario han sido determinadas. A cada matriz inicial se aplica estandarizaciones de datos si se tratase de variables cuantitativas y posteriormente un coeficiente de distancia o similaridad/disimilitud, el cual producirá a la vez, una matriz diagonal de distancia. A partir de esta matriz se calcula la distancia promedio de las entradas involucradas y ese valor promedio de distancia se adjudicará a cada código de centroide y a su respectiva celda. Esto permite a R producir mapas de celdas ráster) que reflejan los valores asignados (ver Fig. 39).

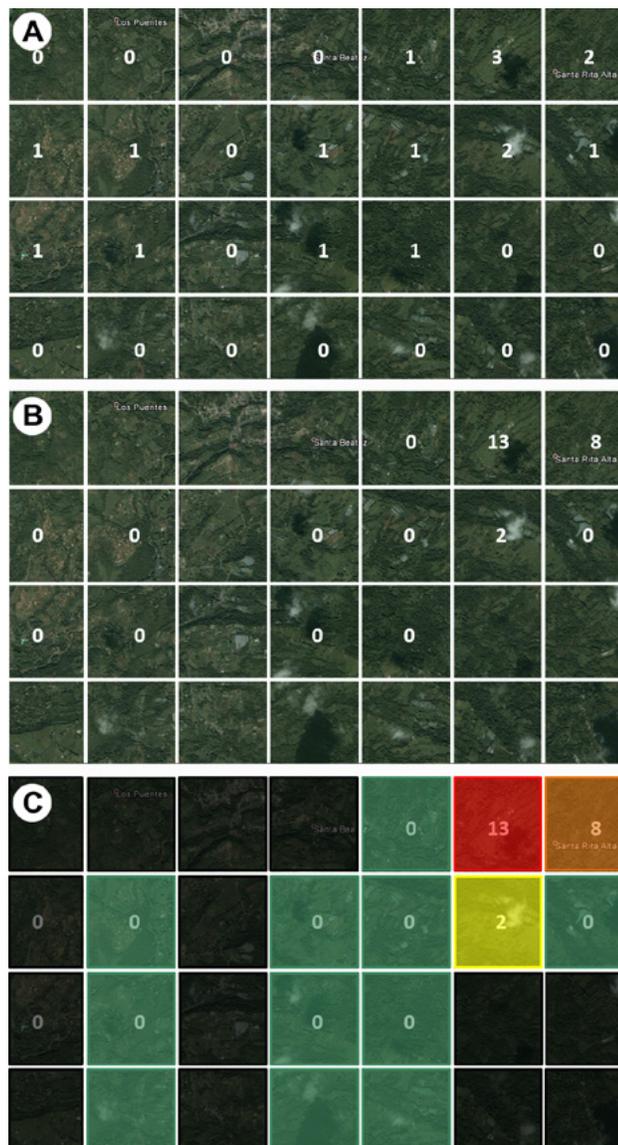


Figura 39. Cuarto paso. A) Número de entradas analizadas por celda, B) valores asignados a celdas de distancia genotípica, fenotípica o ecogeográfica promedio, C) Asignación de colores de una gradación acorde con los valores promedio de distancia.

Para el caso de caracterizaciones genotípicas, adicionalmente a la distancia o disimilaridad promedio, se puede calcular otros parámetros genéticos como el de la diversidad genética de Nei (1987) o la proporción de marcadores polimórficos para cada grupo de entradas dentro de cada área de influencia. Estos parámetros los calcula R desde las matrices iniciales de caracterización.

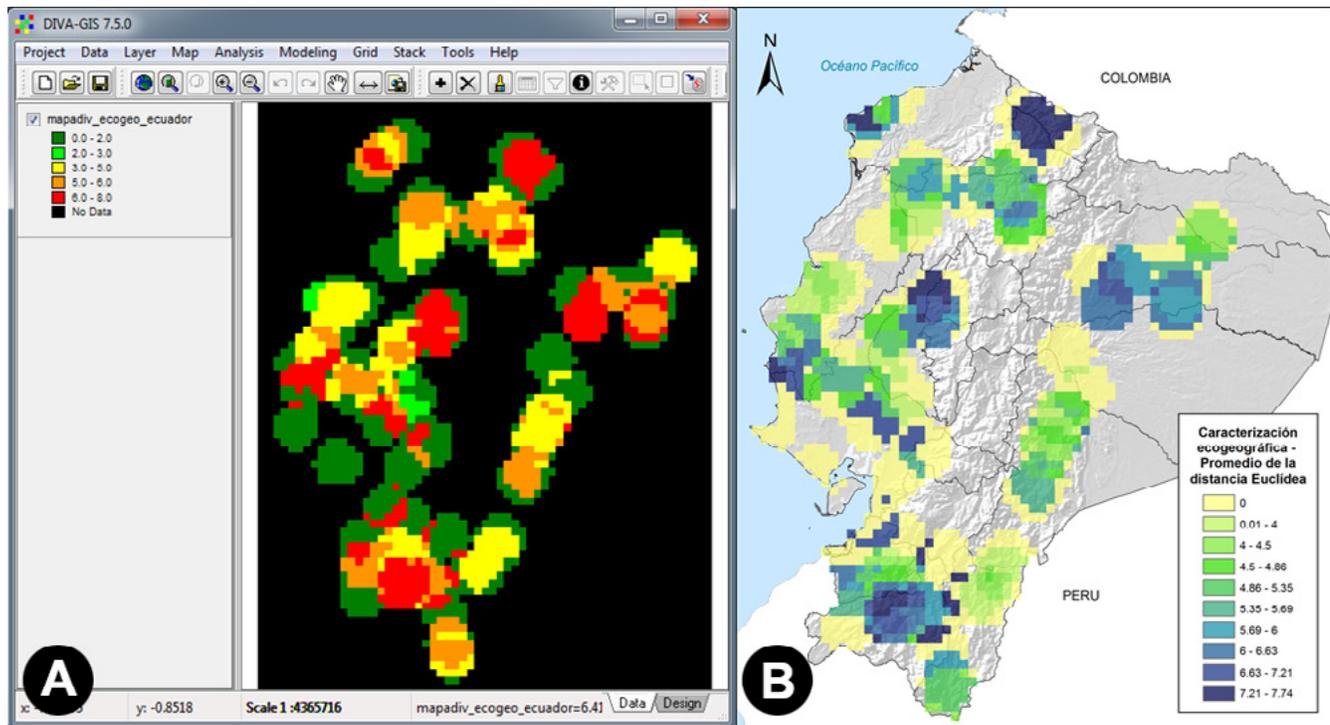


Figura 40. Visualización de mapas de diversidad. A) Mapa de diversidad ecogeográfica para la colección ecuatoriana de maní (*Arachis hypogaea*) producto de DIVmapas abierto en DIVA-GIS, B) mismo mapa en su versión final.

Finalmente, al visualizarse en DIVA-GIS el archivo ráster compuesto por las celdas a las que se le han asignado valores de los parámetros de diversidad, lo que hace el programa informático es asignar un color a cada celda de una gradación de colores, permitiendo observar rápidamente cuales son los lugares donde se presentan mayores niveles de diversidad, medida por valores promedio de distancia/disimilaridad u otros parámetros genéticos (ver Fig. 40). Si se han introducido datos de caracterización de diferente tipo, se producirán varios mapas, ocurriendo siempre la siguiente relación: uno para el aspecto ecogeográfico, uno para el aspecto fenotípico y uno o varios para el aspecto genotípico, dependiendo de si el usuario ha solicitado el cálculo de uno o varios parámetros. También se produce un mapa del número de entradas analizadas por celda como se muestra en la Fig. 41, que corresponde a lo que representa la Fig. 39 parte A. Este último mapa puede servir de apoyo para determinar la existencia de posibles sesgos en la recolección o en la interpretación de los patrones que se puedan encontrar en los mapas de diversidad.

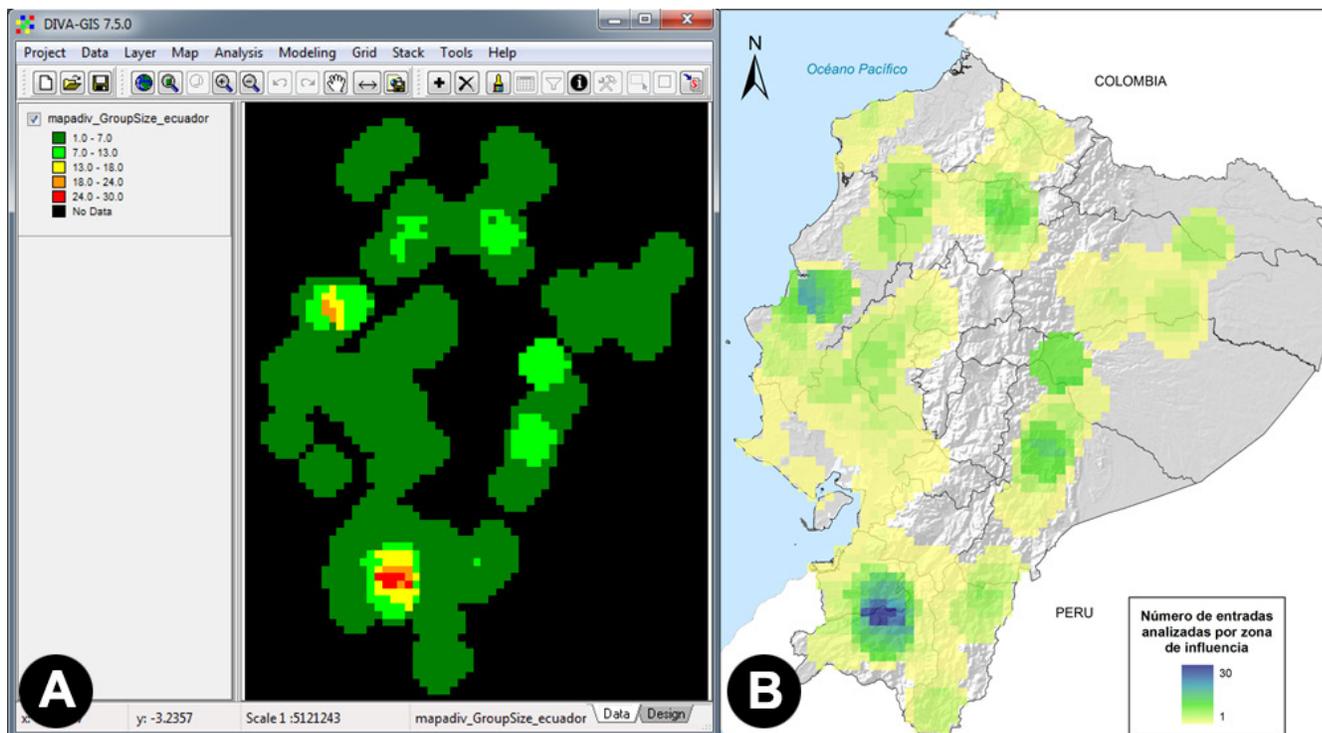


Figura 41. Visualización de mapa de número de entradas analizadas por celda. A) Mapa abierto en DIVA-GIS, B) mismo mapa en su versión final.

9.2.5. Otros análisis

DIVmapas también permite realizar otros tipos de análisis, particularmente cuando se han introducido datos de caracterización de diferentes tipos. De esta manera, DIVmapas pregunta al usuario si desea realizar análisis de agrupamientos, análisis de ordenación de la misma manera que se realiza en la herramienta ECOGEO. Adicionalmente el usuario puede solicitar realizar comparaciones matriciales de Mantel (1967) entre matrices de distancia para el total de las entradas. De manera automática, DIVmapas genera una matriz de distancias geográficas entre todos los sitios de recolección e introduce esta matriz en las correlaciones matriciales pareadas.

9.3. Formatos de datos de entrada en DIVmapas

Para el correcto funcionamiento de la herramienta DIVmapas, es necesario introducir información de diverso tipo en los formatos indicados. Usualmente estos formatos se encuentran en la carpeta "Formatos" dentro de la estructura de carpetas y archivos de las herramientas CAPFITOGEN. Dentro de esta carpeta encontraremos otra con el nombre "Formatos DIVmapas" y dentro de ella cinco archivos Excel.

9.3.1. Modelo de datos de pasaporte

Como en otras herramientas CAPFITOGEN, los datos de pasaporte deben ingresar en formato FAO/Bioversity 2012 con modificaciones menores (archivo “ModeloDatosPasaporte_FAO_BIOVERSITY_2012.xls” disponible para los usuarios desde <https://drive.google.com/drive/folders/1xCnllZgzW0uDeCldvxcBAdV9H583xzpn?usp=sharing>), formato descrito en el apartado sobre la herramienta TesTable. Como DIV mapas permite trabajar con datos evaluados por su calidad de georreferenciación mediante la herramienta GEOQUAL, aparte del modelo de datos de pasaporte normal, se dispone de un modelo que incluye campos adicionales para los valores de evaluación GEOQUAL. Sin embargo, la manera más fácil de usar datos de pasaporte evaluados por GEOQUAL es introducir directamente la tabla de pasaporte con datos de evaluación que produce GEOQUAL como resultado y que usualmente lleva el nombre de “PasaporteOriginalEvaluadoGEOQUAL.txt”. Recuerde que esta tabla debe estar en formato de texto delimitado por tabulaciones, por lo que debe ser exportada desde Excel en este formato, y debe ser ubicada en la carpeta “Pasaporte” de la estructura de carpetas y archivos de las herramientas CAPFITOGEN.

9.3.2. Modelo de datos fenotípicos

Al visualizar el contenido del formato de datos fenotípicos (archivo “ModeloDatosFenotipicos.xlsx”) aparecerá una columna en verde (de relleno obligatorio) denominada “ACCENUMB” que corresponde al mismo código de identificación ACCENUMB de la tabla de datos de pasaporte. El orden en que se encuentren los códigos es indiferente. Dado que no siempre se cuenta con datos de caracterización fenotípica para todas las entradas presentes en los datos de pasaporte, el número de entradas presente en la tabla de datos fenotípicos puede ser menor al de datos de pasaporte. La situación que no debe ocurrir es que en la tabla de datos fenotípicos aparezcan entradas o códigos ACCENUMB que no ocurran en la tabla de datos de pasaporte. Esto último generaría un error de procesamiento.

El resto de las columnas en el formato aparecen con los nombres “D1”, “D2” y “D3”. Estos nombres representan los nombres de los descriptores fenotípicos 1, 2 y 3. El formato sólo incluye tres columnas para los descriptores, pero teóricamente pueden ser tantos descriptores como el usuario disponga, alargando la secuencia desde “D4” hasta donde sea necesario. Los nombres pueden ser cambiados (por ejemplo “D1” por “NGRANOS”), para facilidad del usuario. Si desea cambiar los nombres, tenga en cuenta dos recomendaciones. La primera es que no existan espacios dentro del nombre, que en lo posible el nombre este compuesto por menos de once caracteres y que especialmente no se repita ningún nombre. Esta última condición puede generar error.

La codificación de las variables fenotípicas tiene ciertas condiciones. Las variables, ya sean cuantitativas o categóricas deben estar expresadas numéricamente. Para las variables categóricas, los estados que durante la caracterización hayan sido descritos mediante códigos que incluyen caracteres alfabéticos o de otros tipos, deben ser transformados a códigos exclusivamente numéricos, sin líneas, puntos, comas o espacios. Si existen datos faltantes, éstos deben codificarse con los caracteres “NA”.

Finalmente tenga en cuenta que la herramienta DIVmapas sólo reconoce la información en tablas cuando esta se encuentra en formato de texto separado por tabulaciones. En consecuencia, una vez la tabla de datos fenotípicos

en Excel esté completa y acorde a los requerimientos previamente señalados, se debe exportar en formato de texto separado por tabulaciones y este archivo de texto debe ubicarse en la carpeta “Pasaporte” junto con el resto de las tablas de datos de caracterización y la tabla de datos de pasaporte.

9.3.3. Modelo de tabla de naturaleza de las variables fenotípicas

En el caso de que disponga de datos de caracterización fenotípica y que desee utilizarla para generar un mapa de diversidad con DIVmapas, además de aportar la tabla de datos fenotípicos del apartado 9.3.2, es imprescindible rellenar la tabla que aparece con el nombre de “ModeloTablaNaturalezaVariables.xlsx”, la cual indicará la naturaleza de cada variable o descriptor fenotípico incluido en la tabla de datos fenotípicos. Este archivo Excel contiene dos hojas de cálculo. En la primera (“Natvariables”) se encuentra la tabla de naturaleza de variables fenotípicas, que sólo contiene tres columnas. La primera es “ID” y en ella simplemente se asigna un número consecutivo a cada variable (1, 2, 3,...), de esta manera cada fila en esta tabla corresponde a una variable o descriptor fenotípico de la tabla de datos fenotípicos. La segunda es “NOMVAR” y corresponde exactamente a los nombres asignados a las variables o descriptores en la tabla de datos fenotípicos. La tercera y última columna es “NATVAR”, en la que se debe indicar la naturaleza de la variable o descriptor correspondiente. Al situar el cursor sobre esta celda, se desplegará la lista de posibles valores para esta columna, a saber: binario simétrico, binario asimétrico, nominal, ordinal, cuantitativo.

Finalmente, en la hoja de cálculo “Observaciones” se encuentran algunas indicaciones y ayudas para rellenar la tabla de la hoja de cálculo “Natvariables”.

Al finalizar, la tabla de naturaleza de variables debe ser exportada en formato de texto delimitado por tabulaciones y debe situarse en la carpeta “Pasaporte”, de igual manera que el resto de las tablas de ingreso de información.

9.3.4. Modelo de datos genotípicos

Como se ha señalado previamente, DIVmapas ofrece la posibilidad de hacer mapas de diversidad a partir de caracterizaciones genotípicas de germoplasma, analizando la información proveniente de marcadores moleculares, como si estos fueran de tipo dominante. Esto implica que la tabla de datos genotípicos (archivo Excel “ModeloDatosGenotípicos0_1.xlsx”) contendrá variables de ausencia/presencia codificados como 0 y 1 respectivamente. Así, la estructura de esta tabla es muy similar a la de datos fenotípicos y se debe rellenar de manera similar, excepto porque todas las variables o descriptores en la tabla de datos genotípicos corresponden a variables binarias asimétricas y se codificarán con valores 0 y 1.

Al igual que con la información fenotípica, tenga en cuenta que la herramienta DIVmapas sólo reconoce la información en tablas cuando ésta se encuentra en formato de texto separado por tabulaciones. En consecuencia, una vez la tabla de datos genotípicos en Excel esté completa y acorde a los requerimientos señalados, se debe exportar en formato de texto separado por tabulaciones y este archivo de texto debe ubicarse en la carpeta “Pasaporte” junto con el resto de las tablas de datos de caracterización y la tabla de datos de pasaporte.

9.4. Utilización de la herramienta DIVmapas

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta DIVmapas, se deben especificar una serie de parámetros por parte del usuario.

9.4.1 Parámetros iniciales definidos por el usuario

9.4.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

9.4.1.2 Parámetro: país

Explicación: Seleccione el país/región en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente. También puede utilizar la herramienta rLayer para producir sus propios marcos de trabajo y seleccionarlos aquí.

9.4.1.3 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta 'Pasaporte' pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama 'tabla.txt' y se encuentra en una carpeta llamada 'Stuberosum' que se encuentra dentro de 'Pasaporte', entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

9.4.1.4 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

9.4.1.5 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

9.4.1.6 Parámetro: buffy

Explicación: Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea que la extracción de información ecogeográfica se haga sobre un área circular alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección. Esta opción es muy útil, por ejemplo, cuando la mayoría de los sitios de recolección corresponden a mercados y no directamente las fincas de los agricultores, o cuando existe bastante incertidumbre sobre la real ubicación de los sitios de recolección a pesar de que GEOQUAL pueda otorgar alta calidad de georreferenciación. La explicación de esta característica se encuentra explicada en el capítulo dedicado a la herramienta ECOGEO.

9.4.1.7 Parámetro: tamp

Explicación: Solo aplica si buffy ha sido señalado como TRUE (✓ en modo on server). Especifique el radio (en kilómetros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor. La extracción se hará sobre aquellas celdas cuyo centroide esté dentro del área circular.

9.4.1.8 Parámetro: ecogeo

Explicación: Indique aquí (con la opción TRUE en modo local o un ✓ en modo on server) si está interesado en obtener un mapa de diversidad ecogeográfica de los sitios de recolección/observación del germoplasma.

9.4.1.9 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución o precisión respecto a unas coordenadas (puntos X e Y), pero exige más capacidad y tiempo de cómputo en relación a 5x5 km. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes.

9.4.1.10 Parámetro: bioclimsn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea introducir variables de tipo bioclimático (temperaturas medias, máximas o mínimas anuales o mensuales, precipitaciones mensuales o anuales, presión de vapor, etc.) en el

cálculo de la diversidad ecogeográfica.

9.4.1.11 Parámetro: bioclimv

Explicación: Solo aplica si bioclimsn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables bioclimáticas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables bioclimáticas), cópielos y péguelos separados por un signo de punto y coma (;). También encontrará una línea bloqueada la selección de las 19 variables bioclim, listas para usar, simplemente eliminando el símbolo # inicial. En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro bioclimv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

9.4.1.12 Parámetro: edaphsn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea introducir variables de tipo edáfico (texturas, profundidad, pH, etc.) en el cálculo de la diversidad ecogeográfica.

9.4.1.13 Parámetro: edaphv

Explicación: Solo aplica si edaphsn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables edáficas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables edáficas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro edaphv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

9.4.1.14 Parámetro: geophyssn

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea introducir variables de tipo geofísico (relativas al relieve y radiación solar) en el cálculo de la diversidad ecogeográfica.

9.4.1.15 Parámetro: geophysv

Explicación: Solo aplica si geophysn ha sido señalado como TRUE (✓ en modo on server). Liste (modo local) o seleccione (modo on server) las variables geofísicas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables geofísicas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro geophysv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

9.4.1.16 Parámetro: latitud

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la latitud (Y) como una variable geofísica más a ser analizada.

9.4.1.17 Parámetro: longitud

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la longitud (X) como una variable geofísica más a ser analizada.

9.4.1.18 Parámetro: phenotip

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un mapa de diversidad fenotípica. Para ello deberá disponer de datos de caracterización o evaluación fenotípica (morfología, fenología, productividad, resistencia, etc.) siguiendo el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama ‘fenotipos’, en este espacio debe escribir ‘fenotipos.txt’. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte, de la estructura de datos de las herramientas CAPFITOGEN3 modo local. Para el modo on server, la tabla se subirá en el momento de configuración de la herramienta o previamente, quedando disponible en la zona de archivos de usuario y resultados.

9.4.1.19 Parámetro: phenot

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización fenotípica en el formato indicado en modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama ‘fenotipos’, en este espacio debe escribir ‘fenotipos.txt’. Para el modo on server, debe hacer click en el botón que le permite buscar, seleccionar y subir el archivo correspondiente.

9.4.1.20 Parámetro: phenotv

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene la tabla que describe la naturaleza de cada variable fenotípica en el formato indicado para el modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'variablesfenotipo', en este espacio debe escribir 'variablesfenotipo.txt'. Esta tabla debe describir todas las variables incluidas en la tabla con los datos de la caracterización (ver parámetro anterior). Para el modo on server, debe hacer click en el botón que le permite buscar, seleccionar y subir el archivo correspondiente.

9.4.1.21 Parámetro: genotip

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un mapa de diversidad genotípica. Para ello deberá disponer de datos de caracterización o evaluación genotípica (como presencia o ausencia de marcador como ceros y uno) siguiendo el formato indicado. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte (modo local) de la estructura de datos de las herramientas CAPFITOGEN. Para el modo on server, la tabla se subirá en el momento de configuración de la herramienta o previamente, quedando disponible en la zona de archivos de usuario y resultados.

9.4.1.22 Parámetro: genot

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización genotípica en el formato indicado para el modo local. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'genotipos', en este espacio debe escribir 'genotipos.txt'. Para el modo on server, debe hacer click en el botón que le permitirá buscar, seleccionar y subir el archivo correspondiente.

9.4.1.23 Parámetro: neigd

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción si desea obtener un mapa del índice promedio de diversidad genética de Nei (1987), un mapa de la proporción promedio de marcadores polimórficos y un mapa de número de entradas analizadas por celda.

9.4.1.24 Parámetro: csimilar

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Indique la opción preferida del coeficiente de similaridad que desea utilizar para generar el mapa de distancia genotípica promedio. Tiene a disposición: 1 = Índice de Jaccard (1901), 2 = SMC de Sokal y Michaner 1958, 3 = Sokal y Sneath (1963) (S5 de Gower y Legendre), 4 = Rogers y Tanimoto (1960), 5 = Dice (1945), 6 = coeficiente de Hamann, 7 = Ochiai (1957), 8 = Sokal y Sneath (1963) (S13 de Gower y Legendre), 9 = Phi de Pearson, 10 = S2 de Gower y Legendre. La distancia (d) se obtiene como $d = \sqrt{1-s}$ siendo s el coeficiente de similaridad.

9.4.1.25 Parámetro: rgrid

Explicación: Elija el tamaño de celda (en km) que tendrá el/los mapas de diversidad que va a generar. Este parámetro está restringido a los siguientes valores: 1, 2.5, 5, 7.5, 10, 15, 20, 25, 30, 40, 50 y 100 km (si escoge otro valor se producirá un error).

9.4.1.26 Parámetro: buffer

Explicación: Elija el radio del área circular de influencia o vecindad (en km). Esta área se creará a partir de cada centroide de celda del mapa con sitios de recolección y genera grupos con las entradas cuyo sitio de recolección queda incluido. El valor de índices y distancias promedio de cada grupo será asignado a la celda desde cuyo centroide se trazó el área de influencia.

9.4.1.27 Parámetro: ecogeoclus

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de agrupamientos para las accesiones caracterizadas ecogeográficamente.

9.4.1.28 Parámetro: ecogeoclustype

Explicación: Solo aplica si ecogeoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos ecogeográficos: “single” = vecino más cercano, “complete” = vecindario más compacto, “ward” = método de mínima varianza de Ward, “mcquitty” = método de McQuitty, “average” = similaridad promedio (UPGMA), “median” = similaridad de la mediana, “centroid” = centroide geométrico, “flexible” = Beta flexible.

9.4.1.29 Parámetro: ecogeopca

Explicación: Solo aplica si el parámetro “ecogeo” ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (con la opción TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de componentes principales para las accesiones caracterizadas ecogeográficamente.

9.4.1.30 Parámetro: ecogeopcaxe

Explicación: Solo aplica si ecogeopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de componentes a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCA. Esto significa que sólo los componentes principales “retenidos” serán mostrados en las tablas de resultados.

9.4.1.31 Parámetro: phenoclus

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea obtener un análisis de agrupamientos de todas las accesiones para las cuales se aporta información fenotípica.

9.4.1.32 Parámetro: phenoclustype

Explicación: Solo aplica si phenoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos fenotípicos: 'single' = vecino más cercano, 'complete' = vecindario más compacto, 'ward' = método de mínima varianza de Ward, 'mcquitty' = método de McQuitty, 'average' = similaridad promedio (UPGMA), 'median' = similaridad de la mediana, 'centroid' = centroide geométrico, 'flexible' = Beta flexible.

9.4.1.33 Parámetro: phenopca

Explicación: Solo aplica si phenotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de componentes/coordenadas principales de todas las accesiones para las cuales se ha aportado información fenotípica.

9.4.1.34 Parámetro: phenopcaxe

Explicación: Solo aplica si phenopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de componentes/coordenadas a retener (siempre menor al número de variables fenotípicas) dentro del análisis PCA/PCoA. Esto significa que sólo los componentes principales "retenidos" serán mostrados en las tablas de resultados.

9.4.1.35 Parámetro: phenovarq

Explicación: Solo aplica si phenopca ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si la totalidad de las variables/descriptores fenotípicos corresponden a variables de tipo cuantitativo.

9.4.1.36 Parámetro: genoclus

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de agrupamientos de todas las accesiones para las cuales se aporta información genotípica.

9.4.1.37 Parámetro: genoclustype

Explicación: Solo aplica si genoclus ha sido señalado como TRUE (✓ en modo on server). Elija el tipo de agrupamiento

jerárquico que desea utilizar para los agrupamientos genotípicos: 'single' = vecino más cercano, 'complete' = vecindario más compacto, 'ward' = método de mínima varianza de Ward, 'mcquitty' = método de McQuitty, 'average' = similaridad promedio (UPGMA), 'median' = similaridad de la mediana, 'centroid' = centroide geométrico, 'flexible' = Beta flexible.

9.4.1.38 Parámetro: **genopco**

Explicación: Solo aplica si genotip ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (con la opción TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de coordenadas principales de todas las accesiones para las cuales se ha aportado información genotípica.

9.4.1.39 Parámetro: **genopcoaxe**

Explicación: Solo aplica si genopca ha sido señalado como TRUE (✓ en modo on server). Indique aquí el número de coordenadas a retener (siempre menor al número de variables genotípicas o marcadores) dentro del análisis PCoA. Esto significa que sólo las coordenadas principales "retenidas" serán mostrados en las tablas de resultados.

9.4.1.40 Parámetro: **mantelt**

Explicación: Indique (con la opción TRUE en modo local o ✓ en modo on server) si desea analizar las correlaciones matriciales (Mantel, 1967) entre las posibles combinaciones de factores (ecogeográfico vs. fenotípico vs. genotípico). Se realizarán todas las posibles comparaciones de acuerdo a si se introdujeron datos fenotípicos o genotípicos o si se creó una matriz de caracterización ecogeográfica a partir de sitios de recolección. Se generará e incluirá una matriz de distancias geográficas en las comparaciones matriciales pareadas.

9.4.1.41 Parámetro: **mantelmeth**

Explicación: Solo aplica si mantelt ha sido señalado como TRUE (✓ en modo on server). Seleccione el tipo de correlación a usar en la prueba de Mantel.

9.4.1.42 Parámetro: **mantelper**

Explicación: Solo aplica si mantelt ha sido señalado como TRUE (✓ en modo on server). Ingrese el número de permutaciones que desea para realizar la prueba de Mantel.

9.4.1.43 Parámetro: **resultados (sólo aplica para modo local)**

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

9.5. Resultados de DIVmapas

Después de definir todos los parámetros y rutas (para modo local) que requiere DIVmapas, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, DIVmapas guardará los resultados en la ruta y carpeta que se haya determinado en “resultados” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server. Dado que la ejecución de DIVmapas puede llegar a producir muchos resultados y que éstos pueden ser organizados de acuerdo al tipo de datos y/o análisis de donde provienen, DIVmapas crea varias carpetas dentro de la ruta indicada en el parámetro “resultados” (modo local) o en la zona de archivos de usuario y resultados (modo on server). De esta manera, los resultados se guardan en las correspondientes carpetas temáticas que se explicarán en los siguientes apartados, y sólo se ubicará fuera de ellas el mapa de puntos correspondiente a los sitios de recolección en dos versiones (“ShapefilePuntosPasaporte.shp” y “mapa_puntospas_google.kml”) y el archivo Parametros.Parameters.DIVmapas.txt con un registro de los parámetros utilizados.

9.5.1 Carpeta “ClassicMultivariateResults pais”

Esta carpeta contiene gráficas (formato .wmf) y tablas (.txt y .xls) producto de análisis multivariados (análisis de agrupamientos y análisis de componentes/coordenadas principales). Dependiendo de los datos introducidos, aparecerá en el nombre de los archivos los términos “ecogeographic” (proviene de la caracterización ecogeográfica), “genotypic” (proviene de la caracterización genotípica) y “phenotypic” (proviene de la caracterización fenotípica).

9.5.2 Carpeta “EcogeographicResults pais”

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad ecogeográfica (parámetro ecogeo). Este mapa de diversidad se ofrece en cuatro formatos (imagen “mapadiv_ecogeo_pais.png”, Google earth “mapadiv_ecogeo_pais.kml”, DIVA-GIS “mapadiv_ecogeo_pais.grd” y geotif “mapadiv_ecogeo_pais.tif”). En estos mapas aparecen representados con diferentes colores el promedio de las distancias ecogeográficas de las entradas del área de influencia de cada celda. La distancia corresponde a la euclídea y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas hayan sido recolectadas en ambientes idénticos) hasta infinito.

Respecto a las tablas, pueden encontrarse las siguientes:

9.5.2.1 “tabla_estadisticas_mapadiv_ecogeo.txt”. Esta tabla muestra las estadísticas del mapa de diversidad ecogeográfica en términos de la distancia, es decir, la media, desviación estándar y valores máximos y mínimos de distancia, determinados en el conjunto de celdas que componen el mapa.

9.5.2.2 “TablaVarEcogeograficaspais.txt”. Esta tabla contiene los datos de caracterización ecogeográfica de las entradas analizadas.

9.5.3 Carpeta “PhenotypicResults_pais”

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad fenotípica (parámetro phenotip) y se haya aportado la tabla con los datos correspondientes. Contiene el mapa de diversidad en tres formatos (imagen “mapadiv_phenot_pais.png”, Google earth “mapadiv_phenot_pais.kml”, y DIVA-GIS “mapadiv_phenot_pais.grd”). En estos mapas aparecen representados con diferentes colores el promedio de las distancias fenotípicas de las entradas del área de influencia de cada celda. La distancia corresponde a 1-coeficiente general de similitud de Gower (1971) y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas presenten fenotipo idéntico) hasta 1 (máxima diferencia).

Respecto a las tablas, se encontrará sólo “tabla_estadisticas_mapadiv_phenot” con extensiones .txt y .xls. Esta tabla muestra las estadísticas del mapa de diversidad fenotípica en términos de la distancia, es decir, la media, desviación estándar y valores máximos y mínimos de distancia, determinados para todo el conjunto de celdas que componen el mapa.

9.5.4 Carpeta “GenotypicResults_pais”

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad genotípica (parámetro genotip) y se haya aportado la tabla con los datos correspondientes.

Al interior de la carpeta se encontrarán los siguientes mapas:

9.5.4.1 “mapadiv_GenotDistance_pais”. Corresponde al mapa de diversidad genotípica medida en distancias promedio en cuatro formatos (imagen “.png”, Google earth “.kml”, DIVA-GIS “.grd” y geotif “.tif”). En estos mapas aparecen representados con diferentes colores el promedio de las distancias genotípicas de las entradas caracterizadas del área de influencia de cada celda. La distancia corresponde a 1-coeficiente de similitud de Dice (1945) y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas presenten fenotipo idéntico) hasta 1 (máxima diferencia).

9.5.4.2 “mapadiv_GroupSize_pais”. Corresponde al mapa de número de entradas analizadas por celda (imagen “.png”, Google earth “.kml”, DIVA-GIS “.grd” y geotif “.tif”). En estos mapas aparecen representados con diferentes colores el número de las entradas de cada área de influencia de cada celda.

9.5.4.3 “mapadiv_NeisGeneDiversity_pais”. Corresponde al mapa de diversidad genotípica medida por el índice de diversidad de Nei (1987) en cuatro formatos (imagen “.png”, Google earth “.kml”, DIVA-GIS “.grd” y geotif “.tif”). En estos mapas aparecen representados con diferentes colores el índice de diversidad antes mencionado, obtenido de las entradas caracterizadas del área de influencia de cada celda.

9.5.4.4 “mapadiv_ProportionVariableMarkers_pais”. Corresponde al mapa que muestra la proporción de marcadores polimórficos en tres formatos (imagen “.png”, Google earth “.kml”, DIVA-GIS “.grd” y geotif “.tif”). En estos mapas aparecen representados con diferentes colores la proporción de marcadores moleculares polimórficos obteni-

da de las entradas caracterizadas del área de influencia de cada celda.

Respecto a las tablas, pueden encontrarse las siguientes:

9.5.4.5 “tabla_estadisticas_mapa_GenotDistance.txt”. Esta tabla muestra las estadísticas del mapa de diversidad genotípica en términos de la distancia de Dice (1945), es decir, la media, desviación estándar y valores máximos y mínimos de dicha distancia, determinados para todo el conjunto de celdas que componen el mapa.

9.5.4.6 “tabla_estadisticas_mapa_NeiGeneDiversity.txt”. Esta tabla muestra las estadísticas del mapa de diversidad genotípica en términos del índice de diversidad genética de Nei (1987), es decir, la media, desviación estándar y valores máximos y mínimos para este índice, determinados para todo el conjunto de celdas que componen el mapa.

9.5.5 Carpeta “MantelCorrelationResults_pais”

En esta carpeta se guardarán todas las tablas con las matrices de distancias calculadas para todas las entradas a la vez (“Matriz_distancia_”) y las tablas con los resultados de las correlaciones matriciales de Mantel (1967). El nombre de cada tabla indica la comparación de información realizada. Para las correlaciones donde intervienen datos genotípicos se usa la matriz de distancia de Dice. Por ejemplo, el archivo “Mantel_genotypic_Vs_phenotypic.txt” contiene los resultados de la correlación matricial entre las distancias genotípicas (Dice) y las distancias fenotípicas (Gower). Es importante anotar que DIV mapas también calcula la matriz de distancias geográficas (calculadas en grados decimales) para poder hacer comparaciones matriciales respecto al componente de distancia geográfica.

9.6. Referencias

Damme, P., Garcia, W., Tapia, C., Romero, J., Manuel Siqueñas, M., Hormaza, J.I. 2012. Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. PLoS ONE 7(1): e29845. doi:10.1371/journal.pone.0029845

Dice, L.R. 1945. Measures of the Amount of Ecologic Association Between Species. Ecology 26:297–302.

FAO, BIOVERSITY. 2015. FAO/Bioversity multi-crop Passport descriptors V.2. URL: <https://bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v21-mcpd-v21/>

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27: 857★74.

Hoffmann, M.H., Glaß, A.S., Tomiuk, J., Schmutz, H., Fritsch, R.M., Bachmann, K. 2003. Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). Molecular Ecology, 12: 1007–1019

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209-220.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., van Etten, J. 2012. Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS ONE* 7(10): e47676. doi:10.1371/journal.pone.0047676

van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P, et al. (2012) Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. *PLoS ONE* 7(1): e29845. doi:10.1371/journal.pone.0029845



Taller Nacional, Esperanza, Santa Fe, Argentina, Abril 2016.



10 | Herramienta ColNucleo

10.1. Concepto de colección núcleo

Una colección núcleo o nuclear (en adelante núcleo) es una subcolección, o una fracción de una colección original, que se puede realizar por diversos motivos. El tamaño de la colección original es la mayor determinante a la hora de tomar la decisión de crear una colección núcleo. Las colecciones núcleo suelen considerarse una solución cuando el tamaño de las colecciones originales se convierte en un problema. Una colección de mayor tamaño puede representar un problema a la hora de multiplicar, caracterizar o evaluar germoplasma, en particular cuando los recursos económicos son limitados, seleccionar materiales para programas de mejoramiento, crear colecciones activas o de trabajo, etc. Dependiendo de las condiciones de cada sitio, una colección de mayor tamaño puede ser una conformada por 500, 1000, 2000 o más entradas.

Una colección núcleo suele tener un tamaño del 10% del total de entradas de la colección original, si bien hay estudios que sitúan el porcentaje óptimo por encima o por debajo de este valor (Parra Quijano *et al.*, 2011a). A este porcentaje se le denomina “intensidad de muestreo”.

Es importante anotar que de ninguna manera la determinación de una colección núcleo debe poner en riesgo la conservación de las entradas no seleccionadas, las que se conocen como “colección de reserva”. Una colección núcleo puede facilitar la toma de decisiones en ciertas actividades de conservación y determinación de prioridades bajo condiciones de recursos insuficientes, lo cual no exime la responsabilidad de conservar la colección por completo. Por ejemplo, cuando es necesario realizar una multiplicación de germoplasma, se cuenta con una colección núcleo y los recursos son limitados, se podría plantear multiplicar en primer término las entradas de la colección núcleo y con recursos adicionales en otro ciclo de multiplicación, el resto de la colección.

Independiente de los motivos que lleven a su creación, la principal característica de una colección núcleo respecto a otro tipo de subcolecciones es que ésta debe ser representativa de la diversidad genética contenida en la colección original. Esto implica que una colección núcleo debe contener entradas lo más disímiles entre sí, de tal manera que desde un punto de vista genético no se incluyan entradas duplicadas o altamente emparentadas (Brown, 1995). Por consiguiente, para obtener un subconjunto de entradas disímiles genéticamente, es imprescindible contar con información sobre la composición genética de la colección, en otras palabras, disponer de datos de caracterización.

En este punto aparece una de las primeras limitantes para obtener colecciones núcleo, en un escenario de recursos limitados, caracterizar una colección de más de 1000 o 2000 entradas puede resultar inabordable. Las caracterizaciones genotípicas y fenotípicas suelen demandar ingentes esfuerzos económicos y humanos que muchas instituciones no pueden asumir o lo hacen de manera parcial. Por lo tanto, para superar este tipo de impedimentos se puede utilizar otro tipo de datos de caracterización.

En ciertos casos, cuando se ha requerido la creación de colecciones núcleo y no se contaban con datos de caracterización para su obtención, la solución propuesta fue el uso de datos de pasaporte, particularmente aquellos datos administrativos que describen la localidad del sitio de la recolección (país, estado, provincia). Con ello, se pretendía asimilar diferentes unidades administrativas de recolección a diferentes ambientes y así lograr una colección núcleo representativa de todas las unidades administrativas y ambientes. De esta forma se crearon diversas colecciones

núcleo administrativas en especies como cacahuete (Upadhyaya *et al.*, 2003), guandul (Reddy *et al.*, 2005), sésamo (Xiourong *et al.*, 2000) o sorgo (Grenier *et al.*, 2001). Este tipo de colecciones no asegurarían que la colección núcleo esté conformada por las entradas más disímiles en términos del ambiente donde fueron colectadas, dado que las diferentes unidades administrativas son divisiones antrópicas que no corresponden necesariamente a ambientes diferentes.

10.1.1 Estrategia de agrupación

El primer paso para la obtención de una colección núcleo es la formación de grupos en la colección original por afinidad. Como se ha comentado previamente, para obtener una colección núcleo es necesario contar con datos de caracterización ecogeográfica, genotípica, fenotípica o en el caso de colecciones núcleo administrativas, datos de pasaporte. Esta información se usa para crear grupos de entradas afines o similares. La creación de los grupos suele realizarse a través de métodos de clasificación multivariada a partir de los datos de caracterización del germoplasma.

En una variante para colecciones núcleo de tipo ecogeográfico, se propone no usar la caracterización del germoplasma sino del territorio donde este ocurre, a través del uso de mapas ecogeográficos de caracterización del territorio (mapas ELC). Así las entradas son agrupadas de acuerdo a la categoría del mapa en el que ocurren. Esto facilitaría la inclusión de nuevas entradas a la colección nuclear, dado que no sería necesario repetir el análisis de agrupamientos, simplemente conocer el grupo (categoría ecogeográfica del mapa) de pertenencia de la nueva entrada (Parra Quijano *et al.*, 2011b).

10.1.2 Determinación de cuotas por estrategias de afijación

Posteriormente, se determina el número de entradas que serán seleccionadas de cada grupo de afinidad. Ese número o cuota la define la estrategia de afijación (*allocation strategy* en inglés) que el curador considere más adecuada. A medida que las colecciones núcleo se han popularizado, se han propuesto numerosas estrategias de afijación. Gradualmente la complejidad y sofisticación de dichas estrategias se ha ido incrementando. Sin embargo algunos estudios comparativos muestran que las estrategias más complejas no necesariamente son las que producen colecciones nucleares más representativas (Parra Quijano *et al.*, 2011b). Las estrategias más recurrentes, sencillas y testadas son las siguientes (Yonezawa *et al.*, 1995):

1. Aleatoria (R): Las entradas son seleccionadas de manera aleatoria del total de la colección. Los grupos creados por la estratificación son ignorados.
2. Constante (C): El mismo número de entradas se selecciona de cada grupo, sin tener en cuenta el número de entradas que conforman el grupo.
3. Proporcional (P): El número de entradas a seleccionar de cada grupo es proporcional al tamaño (número de entradas) de cada grupo.
4. Logarítmico (L): El número de entradas a seleccionar de cada grupo es proporcional al logaritmo del tamaño (número de entradas) de cada grupo.

5. Dependiente de la diversidad (G): El número de entradas a seleccionar de cada grupo es proporcional a la diversidad contenida en cada grupo. Esta estrategia requiere acceso a datos de caracterización, adicionalmente a los agrupamientos provenientes de la estratificación.

10.1.3 Información sobre disponibilidad de entradas

En buena parte de los estudios científicos sobre la creación de colecciones núcleo, se realizan simulaciones para determinar cuál es la estrategia de agrupamiento y de afijación que produce la colección núcleo más representativa para cada caso, utilizando para ellos el total de la colección. Sin embargo, estas aproximaciones teóricas y simulaciones pueden producir colecciones núcleo que en la práctica podrían no constituirse debido a que las entradas seleccionadas no están disponibles. La disponibilidad de una entrada para ser parte de una colección núcleo, puede estar influenciada por varios factores, entre ellos el número de semillas disponible, si la entrada está sólo representada en la colección base o si la entrada tiene algún tipo de restricción en su uso y distribución. Por esta razón es importante que en el caso de la obtención de una colección núcleo de carácter práctico, se tenga en cuenta la información de disponibilidad que pueda aportar el curador de la colección.

10.2. Colecciones núcleo ecogeográficas

La caracterización ecogeográfica representa una alternativa para la obtención de colecciones núcleo. Teniendo en cuenta la relación entre fenotipo, genotipo y ambiente, una colección núcleo creada a partir de datos de caracterización ecogeográfica puede ser representativa no sólo de las condiciones ambientales de las poblaciones de procedencia de las entradas contenidas en la colección original sino también de sus fenotipos y genotipos, siempre que la representatividad sea evaluada mediante características fenotípicas o genotípicas relacionadas con la adaptación (Parra Quijano *et al.*, 2011a).

El uso de datos de caracterización ecogeográfica del germoplasma en la obtención de colecciones núcleo se ha documentado desde 1995, en el caso del establecimiento de una colección núcleo de *Phaseolus vulgaris* en el Centro Internacional de Agricultura Tropical (CIAT) (Tohme *et al.*, 1995). Dado que la amplia disponibilidad de SIG aplicados a recursos fitogenéticos y de capas de información ecogeográfica no fue posible hasta la década 2000-2010, las colecciones núcleo ecogeográficas no reaparecen en el contexto científico internacional hasta 2008, con el caso de *Trifolium spumosum* (Ghamkhar *et al.*, 2008). Posteriormente, en un par de estudios sobre diferentes tipos de colecciones núcleo ecogeográficas, se determinó que la combinación mapa de caracterización ecogeográfica del territorio (mapa ELC) como estrategia de agrupamiento y el tipo proporcional como estrategia de afijación, produjo colecciones núcleo de alta representatividad ecogeográfica y fenotípica para colecciones de *Lupinus spp.* y *Phaseolus vulgaris*, respectivamente (Parra Quijano *et al.*, 2011a, 2011b). En dichos estudios, hasta 16 combinaciones diferentes de estrategias de agrupamiento y afijación, dieron resultados similares o inferiores en términos de representatividad ecogeográfica y fenotípica que la combinación mapa ELC junto con afijación proporcional.

10.3. Obtención de colecciones núcleo ecogeográficas en ColNucleo

Siguiendo las recomendaciones de algunos estudios científicos sobre colecciones núcleo y representatividad, la herramienta ColNucleo permite obtener colecciones núcleo ecogeográficas mediante la combinación de agrupamiento por mapa ELC y tres métodos de afijación (C, P y L). El mapa ELC debe ser generado con la herramienta “ELC mapas” (ver Capítulo sobre esta herramienta).

La Fig. 42 muestra que, en un primer paso, se extrae la categoría del mapa ELC correspondiente al sitio de recolección de cada entrada con coordenadas. Posteriormente las entradas se agrupan de acuerdo a la categoría del mapa ELC asignada. De acuerdo a la estrategia de afijación y la intensidad de muestreo que el usuario haya seleccionado, ColNucleo fija las cuotas o número de entradas de cada grupo que conformarán en conjunto la colección núcleo. Después, ColNucleo determina si la cuota se puede cumplir con entradas que no correspondan a duplicados geográficos (no necesariamente genético) y que el curador ha designado como “disponibles” en el caso de que el usuario haya señalado la opción de utilizar datos de disponibilidad. Las entradas no duplicado tendrán prioridad sobre entradas duplicado. Si la cuota es de menor tamaño que el número de entradas no duplicadas disponibles, se seleccionará aleatoriamente entre éstas. Si la cuota es de mayor tamaño se seleccionarán todas las entradas disponibles no duplicadas y el faltante se cubrirá con entradas duplicadas seleccionadas aleatoriamente. Finalmente, las entradas seleccionadas serán señaladas con el número 1 (uno) en una nueva columna que se añadirá a la tabla de pasaporte de entrada. En el caso de haberse utilizado datos de disponibilidad, las colecciones núcleo obtenidas pueden ser incompletas al no disponerse de suficientes entradas para representar una o varias categorías ELC. Por esta razón ColNucleo genera una tabla adicional en la que señala qué accesiones serían necesarias hacer disponibles para que la colección núcleo represente todas las categorías ELC de acuerdo a las cuotas fijadas.

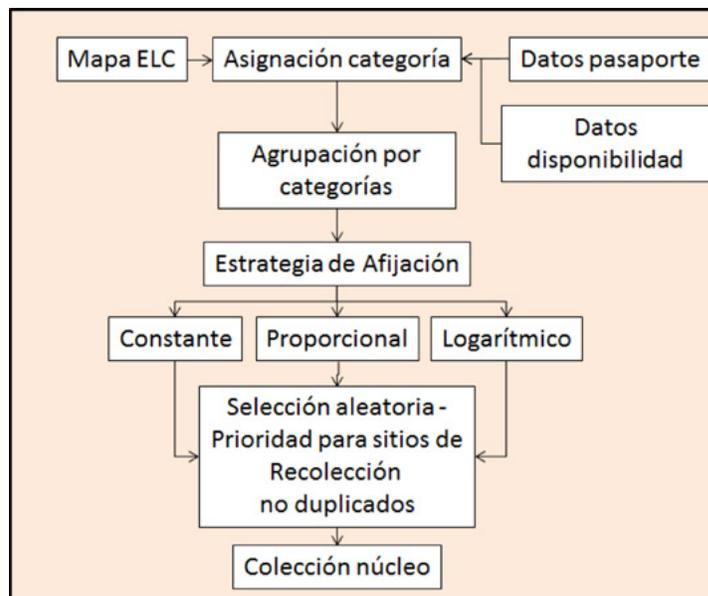


Figura 42. Ilustración del proceso que sigue la herramienta ColNucleo para la obtención de colecciones núcleo ecogeográficas.

10.4. Formato de tabla de pasaporte para ColNucleo

ColNucleo utiliza la tabla de pasaporte FAO/Bioiversity 2012 con modificaciones que utiliza las herramientas GEOQUAL, Representa y ECOGEO, con la adición de un campo en la parte derecha, denominado “AVAILAB” que determina la disponibilidad de cada entrada. De esta manera, en la columna AVAILAB se codificará con el número 1 (uno) las entradas disponibles, con un 0 (cero) las no disponibles y con las letras NA las entradas para las cuales no se tiene información.

10.5. Utilización de la herramienta ColNucleo

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta ColNucleo, se deben especificar una serie de parámetros por parte del usuario. Note que para que esta herramienta funcione correctamente, antes el usuario debe haber obtenido un mapa ELC con la herramienta ELCmapas y los siguientes archivos producto de esta herramienta deben ser copiados y pegados en la ruta CAPFITOGEN3/ELCmapas (modo local) o subidos al servidor cuando los respectivos parámetros lo requieran:

- mapa_elc_pais.grd
- mapa_elc_pais.gri
- Estadist_ELC_pais.txt

Siendo “pais” el nombre del país o la región para la cual se realizó el mapa.

10.5.1 Parámetros iniciales definidos por el usuario

10.5.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

10.5.1.2 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archi-

vos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Ésta tabla contiene una columna adicional denominada “AVAILAB” respecto a la tabla de pasaporte usada como modelo para otras herramientas CAPFITOGEN. Esta columna adicional indica la disponibilidad de cada entrada a ser seleccionada para una colección núcleo/nuclear.

10.5.1.3 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro ‘pasaporte’) la tabla resultado de GEOQUAL denominada “PasaporteOriginalEvaluadoGEOQUAL.txt” con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

10.5.1.4 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

10.5.1.5 Parámetro: mapaelc

Explicación: Indique el nombre del archivo (modo local) o suba el archivo al servidor (modo on server) que contiene el mapa ELC producto de la aplicación de la herramienta ELCmapas, el cual debe encontrarse en la carpeta CAPFITOGEN3/ELCmapas (modo local). Este mapa, en formato DIVA-GIS (extensión .grd, como lo produce ELCmapas) debe escribirse con la extensión. Así, si el nombre del mapa es “mapa_elc_spain”, ud. debe escribir (modo local) o buscar el archivo (modo on server) “mapa_elc_spain.grd”.

10.5.1.6 Parámetro: statelc

Explicación: Indique el nombre del archivo (modo local) o seleccione y suba al servidor (modo on server) que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (esta herramienta suele nombrar a éste archivo como “Estadist_ELC_” más el nombre del país o región). Al igual que el mapa ELC, este archivo debe estar situado en la carpeta CAPFITOGEN3/ELCmapas. De igual forma el nombre debe escribirse junto con la extensión (modo local), en este caso al ser una tabla, la extensión es txt. Así, si el archivo se llama “Estadist_ELC_spain” se debe escribir “Estadist_ELC_spain.txt”.

10.5.1.7 Parámetro: distdup

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia

o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas. La determinación de la distancia depende de condiciones biológicas (flujo de genes) y espaciales (tamaños promedios de las poblaciones). Se trata de un parámetro específico para la especie objetivo y muchas veces será necesario consultar el concepto de un experto.

10.5.1.8 Parámetro: porcol

Explicación: Corresponde a la intensidad de muestreo. Indique el tamaño que desea para la colección núcleo/nuclear, expresado en porcentaje del tamaño de la colección original (valores de 0 a 100). Por ejemplo, si la colección original contiene 2000 entradas y se desea una colección núcleo/nuclear de 200 entradas, el valor a indicar aquí será 10, o para una colección núcleo/nuclear de 300 entradas, el valor será 15.

10.5.1.9 Parámetro: estratcol

Explicación: Seleccione la estrategia de afijación o fijación de cuotas de representación de cada categoría ecogeográfica del mapa ELC. Puede escoger entre las estrategias “C” constante (la misma cuota para todas las categorías), “P” proporcional (cuotas proporcionales a la cantidad de entradas de cada categoría) o “L” logarítmica (cuotas proporcionales al logaritmo de la cantidad de entradas de cada categoría).

10.5.1.10 Parámetro: availab

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea utilizar la columna sobre disponibilidad de entradas para ser seleccionadas con destino a una colección núcleo. Recuerde que, para esta herramienta, la tabla de pasaporte incluye una columna denominada “AVAILAB”, donde se indica la disponibilidad de las entradas de la colección original para integrar colecciones núcleo, a través de los códigos 0 (entrada no disponible), 1 (entrada disponible) y NA (sin información/no disponible). Si no desea utilizar información sobre disponibilidad, la selección de entradas se realizará sobre el total de las entradas. La disponibilidad es un criterio propio del curador y podría estar determinado por el número de semillas conservadas, su germinación u otros diversos factores.

10.5.1.11 Parámetro: resultados (sólo para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

10.6. Resultados de ColNucleo

Una vez finalizado el análisis, ColNucleo producirá un total de cuatro tablas en la ruta determinada en el parámetro “resultados” (modo local) o la zona de archivos de usuario y resultados (modo on server), si el usuario ha indicado utilizar los datos de disponibilidad de germoplasma (columna AVAILAB en la tabla pasaporte) o tres en el caso que haya decidido no utilizar dicha información.

10.6.1.1 “CoreCollection.txt/xls”.

Esta tabla contiene la tabla de pasaporte de las entradas que ColNucleo ha seleccionado como colección núcleo ecogeográfica, con una columna adicional en el extremo derecho de la tabla denominada “BGcat”, que indica el grupo o categoría del mapa ELC al que pertenece la entrada, de acuerdo con la ubicación del sitio de recolección.

10.6.1.2 “CoreCollect_Properties.txt/xls”.

Esta tabla muestra varios de los parámetros introducidos por el usuario y que ColNucleo ha utilizado para obtener la colección núcleo. Los campos que incluye son “Allocation_strategy” o estrategia de afijación, “Sample_size” o porcentaje de intensidad de muestreo, “Use_availability_data” o si se usa o no datos de disponibilidad, “No_access_sampled”, tamaño de la colección núcleo obtenida y “No_access_to_be_multiplied” (solo para el caso de usar datos de disponibilidad) o número de entradas que no están disponibles o no se tiene información sobre su disponibilidad, y que deberían estarlo para lograr una colección núcleo completa.

10.6.1.3 “CoreCollect_stats.txt/xls”.

Esta tabla contiene estadísticas por cada categoría del mapa ELC (identificadas en la columna “ELC_CAT”). Contiene las siguientes columnas a la derecha de “ELC_CAT”: “FREC_W_DUPL” que indica el número de entradas, incluyendo duplicados geográficos, cuyo sitio de recolección ocurre dentro de cada categoría, “FREC_WO_DUPL” indica lo mismo que la anterior columna pero excluyendo duplicados geográficos, “Porcent_W_DUPL” indica el porcentaje de entradas (contando duplicados) que ocurre en cada categoría, “FreqClass_W_DUPL” indica la clasificación en cuartiles de la frecuencia de ocurrencia en cada categoría, “Duplicates” es el número de entradas duplicados geográficos por cada categoría, “N_Availab” es el número de entradas disponibles en total (duplicados más no duplicados) para cada categoría, “N_AvailabWO” es el número de entradas disponibles no duplicadas para cada categoría, “Q_Even” o “Q_Prop” o “Q_Log” (aparecería el título de columna dependiendo del método de afijación seleccionado) es el número de entradas que deben representar *a priori* cada categoría, es decir, la cuota, y finalmente la columna “CCfinal” que corresponde al número de entradas que componen la colección núcleo ecogeográfica obtenida por ColNucleo en base a los parámetros introducidos y, en su caso, disponibilidad de entradas.

10.6.1.4 “EntriesToBeMultiplied.txt/xls”.

Se trata de una tabla con la misma estructura de columnas que “CoreCollection.txt/xls” solo que en “EntriesToBeMultiplied.txt/xls” aparecen las entradas que ColNucleo ha seleccionado como parte de la colección núcleo pero para las cuales no hay disponibilidad. En principio se asume, por el nombre del archivo, que pueden ser entradas que requerirían ser multiplicadas para ser disponibles y entonces hacer parte de la colección núcleo. Sin embargo, las razones por las cuales no están disponibles pueden ser variadas como se explica en el apartado 10.1.3.

10.7. Referencias

- Brown, A.H.D. 1995. The core collection at the crossroads. p. 3–19. En Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.
- Ghamkhar, K., R. Snowball, B.J. Wintle, Brown, A.H.D. 2008. Strategies for developing a core collection of bladder clover (*Trifolium spumosum* L.) using ecological and agro-morphological data. *Aust. J. Agric. Res.* 59:1103–1112.
- Grenier, C., Hamon, P., Bramel-Cox, P.J.. 2001. Core collection of sorghum: II. Comparison of three random sampling strategies. *Crop Science.* 41:241–246.
- Parra Quijano, M., Iriondo, J.M., Torres, M.E., De la Rosa, L. 2011a. Evaluation and validation of ecogeographical core collections using phenotypic data. *Crop Science* 51:694-703.
- Parra-Quijano, M., Iriondo, J.M., de la Cruz, M., Torres, M.E. 2011b. Strategies for the development of core collections based on ecogeographical data. *Crop Science* 51:656-666.
- Reddy, L.J., H.D. Upadhyaya, C.L.L. Gowda, S. Singh. 2005. Development of core collection in pigeonpea (*Cajanus cajan* (L.) Millspaugh) using geographic and qualitative morphological descriptors. *Genetic Resources and Crop Evolution* 52:1049–1056.
- Tohme, J., Jones, P., Beebe, S., Iwanaga, M. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. p. 95–107. In Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.
- Upadhyaya, H.D., Ortiz, R., Bramel, P.J., S. Singh, S. 2003. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genet. Resour. Crop Evol.* 50:139–148.
- Xiurong, Z., Yingzhong, Z., Yong, C., Xiangyun, F., Qingyuan, G., Mingde, Z., Hodgkin, T. 2000. Establishment of sesame germplasm core collection in China. *Genet. Resour. Crop Evol.* 47:273– 279.
- Yonezawa, K., Nomura, T., Morishima, H. 1995. Sampling strategies for use in stratified germplasm collections. p. 35–53. In Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.



Taller Regional, Pretoria, Sudáfrica, Abril 2015.



11 | Herramienta FIGS_R

11.1. La Estrategia de Selección Focalizada de Germoplasma

La técnica de selección de germoplasma con fines de utilización denominada “Estrategia de Selección Focalizada de Germoplasma” o FIGS por sus siglas en inglés (Focused Identification of Germplasm Strategy), proviene de un concepto originalmente desarrollado por Mackay (1990).

Esta técnica busca identificar entradas de una colección que tienen potencial de uso por parte de mejoradores. El potencial para uso en mejoramiento está basado en el uso de información ecogeográfica de los sitios de recolección y su asociación con caracteres de interés para el mejorador de cultivos (Mackay y Street, 2004).

Dado que FIGS usa variables ecogeográficas de tipo abiótico para la selección de germoplasma, la asociación entre las variables ecogeográficas y el carácter de interés para el mejoramiento es directa si el carácter de interés es de tipo abiótico

o indirecto para un carácter biótico. De esta manera, si un mejorador busca germoplasma con potencial y el carácter de interés es adaptación a condiciones de sequía, buscará directamente germoplasma cuyo sitio de recolección esté caracterizado por bajas precipitaciones. Si la característica de interés fuera de tipo biótica, como la resistencia a un patógeno, primero se debería establecer una relación entre una serie de variables ecogeográficas y la resistencia a dicho patógeno, para consecuentemente seleccionar germoplasma cuyo sitio de recolección reúna las condiciones ecogeográficas asociadas a la resistencia al patógeno.

Se pueden distinguir dos técnicas para la selección de germoplasma mediante FIGS. La primera es la de filtrado de entradas y la segunda es la técnica de calibración.

La técnica de filtrado selecciona entradas de una colección caracterizada ecogeográficamente, escogiendo aquellas que cumplen con ciertos valores o rangos para alguna de las variables por las cuales han sido caracterizadas. En ocasiones lo que se selecciona es una fracción de la distribución de una variable ecogeográfica en la colección caracterizada. Los valores, rangos o la fracción de la distribución y la variable ecogeográfica de selección los determina el investigador, curador o mejorador en base al conocimiento de la especie, de la variable ecogeográfica y del carácter de interés. Un ejemplo de la aplicación de este método fue la selección FIGS indirecta para resistencia en trigo ante la plaga *Eurygaster integriceps* (El Bouhssini *et al.*, 2009). Otro caso es la aplicación de FIGS de forma directa para la identificación de recursos genéticos de *Vicia faba* con adaptación a sequía (Khazaei *et al.*, 2013). La técnica de calibración requiere que la colección haya sido caracterizada ecogeográficamente en su totalidad o casi totalidad (entradas con coordenadas) y que además se haya evaluado al menos parcialmente por el carácter de interés. La técnica de calibración se desarrolla en dos fases. En la primera fase, se utiliza análisis matemáticos y estadísticos para establecer la relación entre la presencia o ausencia del carácter de interés y una o varias variables ecogeográficas. Una vez probada esta relación, se realiza una predicción de presencia o ausencia del carácter de interés sobre la fracción de la colección no evaluada usando para ello la información ecogeográfica disponible para la totalidad de la colección. La predicción señalaría qué entradas serían potencialmente de interés para el mejoramiento de cultivos. La aplicación de la técnica de calibración se aprecia en los estudios de Endresen y colaboradores para cebada y trigo (Endresen, 2010; Endresen *et al.*, 2012).

Por su propia naturaleza, la técnica de calibración se usa para FIGS de tipo indirecto, mientras la técnica de filtrado se ha usado para ambos tipos. La técnica de calibración es metodológicamente más compleja y sus resultados se asumen más precisos en la detección de entradas que poseen la característica de interés respecto a la técnica de filtrado. Sin embargo, la técnica de calibración tiene como desventaja el requisito de tener que contar con datos de evaluación parcial de la colección y que esos datos sean suficientemente fiables como para que la relación que se pueda establecer entre variable ecogeográfica y carácter de interés sea válida. Este hecho hace que su aplicación se vea restringida a un 22% de las colecciones, que es el porcentaje de colecciones nacionales que cuentan con algún tipo de evaluación por factores bióticos en 40 países, según el Segundo Reporte del Estado de los Recursos Fitogenéticos para la Alimentación y la Agricultura (FAO, 2010).

Independientemente de la forma en cómo se obtenga la subcolección FIGS, es altamente recomendable que se valide mediante ensayos de adaptación, tolerancia o resistencia, que efectivamente las entradas seleccionadas poseen el rasgo de interés por la cual fueron seleccionadas a través de las condiciones ecogeográficas de sus sitios de recolección.

11.2. Subcolecciones FIGS y colecciones núcleo

Una subcolección FIGS es el conjunto de entradas con potencial uso en el mejoramiento de una especie cultivada y que proviene de una selección FIGS.

Las subcolecciones FIGS, a diferencia de una colección núcleo no tiene necesariamente que ser representativas de la variabilidad de la colección original. Una subcolección FIGS convencional conlleva un sesgo (el interés del mejorador de cultivos) muy pronunciado en el momento de su selección, por lo que no cabe esperar encontrar alta representatividad.

Otra diferencia entre una colección núcleo y una subcolección FIGS es que para ésta última pueden establecerse tantas por especie como características de interés existan. En contraste, colecciones núcleo se suelen establecer una por especie.

Sin embargo, como en el caso de las colecciones núcleo, establecer una o varias subcolecciones FIGS no debe poner en riesgo la conservación de las entradas no seleccionadas. Si una colección núcleo sirve, por ejemplo, para priorizar algunas labores como la caracterización o evaluación de unas entradas en una colección cuando no se cuenta con medios para realizar la labor para la totalidad de colección, una subcolección FIGS lo que busca es potenciar la utilización del germoplasma contenido en una colección, facilitando al mejorador de cultivos, quien es su principal usuario, la búsqueda de material con potencial para ser integrado en sus programas de mejora.

11.3. Obtención de subcolecciones FIGS con la herramienta FIGS_R

Con la herramienta FIGS_R se puede obtener una subcolección FIGS mediante la técnica de filtrado. FIGS_R permite utilizar hasta tres variables de selección de manera jerárquica, es decir, una variable primaria (de uso obligado) que

ejerce el primer proceso de filtrado, una secundaria (opcional) que filtra sobre el subconjunto producto del primer filtrado y una variable terciaria (opcional y sólo funcional si se ha utilizado una variable secundaria) que filtra sobre el subconjunto producto del segundo filtrado. Como variable primaria, secundaria o terciaria, se puede escoger una de las 177 variables ecogeográficas que se tienen a disposición en las herramientas CAPFITOGEN3. La Fig. 43 muestra el proceso que sigue FIGS_R para la obtención de subcolecciones FIGS.

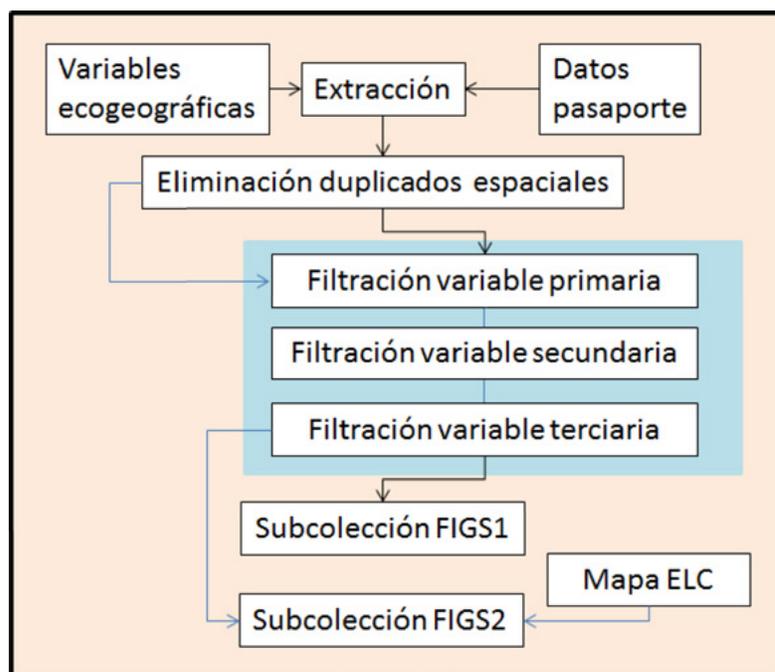


Figura 43. Ilustración del proceso que sigue la herramienta FIGS_R para la obtención de subcolecciones FIGS.

Para cada variable de selección, el usuario de la herramienta FIGS_R puede escoger la forma de realizar la selección. La primera forma es la de la determinación por parte del usuario de un rango de valores que debe satisfacer una entrada para ser incluida en la subcolección FIGS. La segunda forma es la determinación de una fracción de la colección, expresada en porcentaje, cuyas entradas son las que poseen los valores más altos o bajos en relación a la variable de selección.

FIGS_R adopta algunos términos o definiciones del mejoramiento genético de cultivos, como intensidad de selección o diferencial de selección. El primero se usa para definir el porcentaje de la colección inicial que quedará incluido en la subcolección FIGS. El segundo hace referencia a la diferencia entre la media de la colección original y la subcolección FIGS para la o las variables de selección.

Adicionalmente, la herramienta FIGS_R permite crear subcolecciones FIGS balanceadas ecogeográficamente. En otras palabras, si el usuario previamente ha generado un mapa ELC (a través de la herramienta ELC mapas) y usa exclusivamente la segunda forma de selección (fracción de la colección), puede hacer que la herramienta: 1. Asigne categorías a cada entrada en base a la categoría del mapa ELC donde ocurre su sitio de recolección, y 2. seleccione de cada catego-

ría la fracción de entradas con los valores más altos o bajos para la variable de selección. El balance mediante un mapa ELC tiende a producir subcolecciones FIGS con una mayor representatividad ecogeográfica aunque conservando en buena medida el potencial de utilización en programas de mejoramiento por la característica de interés.

Finalmente es importante tener en cuenta que la herramienta FIGS_R puede tener en cuenta también información sobre disponibilidad de las entradas para ser seleccionadas. De esta manera, utiliza el mismo formato de entrada de datos (datos de pasaporte) que la herramienta ColNucleo, es decir, el formato usado para la herramienta GEOQUAL más la adición del campo “AVAILAB”.

11.4 Utilización de la herramienta FIGS_R

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta FIGS_R, se deben especificar una serie de parámetros por parte del usuario. Note que es posible que para el correcto funcionamiento de esta herramienta cuando se ha seleccionado el parámetro “mapaelc” antes el usuario debe haber obtenido un mapa ELC con la herramienta ELCmapas y los siguientes archivos producto de esta herramienta deben ser copiados y pegados en la ruta CAPFITOGEN3/ELCmapas:

- mapa_elc_pais.grd
- mapa_elc_pais.gri
- Estadist_ELC_pais.txt

Siendo “pais” el nombre del país o la región para la cual se realizó el mapa. Si el usuario no selecciona el parámetro “mapaelc”, no requiere ningún mapa ELC y puede realizar sin inconvenientes el filtrado del germoplasma con FIGS_R.

11.4.1 Parámetros iniciales definidos por el usuario

11.4.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

11.4.1.2 Parámetro: país

Explicación: Seleccione el país/región en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente. También puede utilizar la herramienta rLayer para producir sus propios marcos de trabajo y seleccionarlos aquí.

11.4.1.3 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta 'Pasaporte' pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama 'tabla.txt' y se encuentra en una carpeta llamada 'Stuberosum' que se encuentra dentro de 'Pasaporte', entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Esta tabla contiene una columna adicional denominada "AVAILAB" respecto a la tabla de pasaporte usada como modelo para otras herramientas CAPFITOGEN. Esta columna adicional indica la disponibilidad de cada entrada a ser seleccionada para una subcolección FIGS.

11.4.1.4 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

11.4.1.5 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

11.4.1.6 Parámetro: controlelc

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea usar un mapa ELC previamente creado, para determinar la distribución de las entradas que componen la subcolección FIGS sobre las diferentes categorías del mapa. Para usuarios avanzados, esta opción habilita la obtención de una subcolección FIGS adicional, en la que selección de las entradas se hace sobre cada categoría ELC. Se requiere para ello usar métodos de selección sobre fracciones de la distribución para todas las variables consideradas.

11.4.1.7 Parámetro: mapaelc

Explicación: Solo aplica si controlelc ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo (modo local) o suba el archivo al servidor (modo on server) que contiene el mapa ELC producto de la aplicación de la herramienta ELCmapas, el cual debe encontrarse en la carpeta CAPFITOGEN3/ELCmapas (modo local). Este mapa, en formato DIVA-GIS (extensión .grd, como lo produce ELCmapas) debe escribirse con la extensión. Así, si el nombre del mapa es “mapa_elc_spain”, ud. debe escribir (modo local) o buscar el archivo (modo on server) “mapa_elc_spain.grd”.

11.4.1.8 Parámetro: statelc

Explicación: Solo aplica si controlelc ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo (modo local) o seleccione y suba al servidor (modo on server) que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (esta herramienta suele nombrar a éste archivo como “Estadist_ELC_” más el nombre del país o región). Al igual que el mapa ELC, este archivo debe estar situado en la carpeta CAPFITOGEN3/ELCmapas. De igual forma el nombre debe escribirse junto con la extensión (modo local), en este caso al ser una tabla, la extensión es txt. Así, si el archivo se llama “Estadist_ELC_spain” se debe escribir “Estadist_ELC_spain.txt”.

11.4.1.9 Parámetro: distdup

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas. La determinación de la distancia depende de condiciones biológicas (flujo de genes) y espaciales (tamaños promedios de las poblaciones). Se trata de un parámetro específico para la especie objetivo y muchas veces será necesario consultar el concepto de un experto.

11.4.1.10 Parámetro: availab

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea utilizar la columna sobre disponibilidad de entradas para ser seleccionadas como integrantes de la subcolección FIGS. Esto implica dar prioridad a entradas disponibles pero no restringir la posible consideración de no disponibles. Recuerde que para esta herramienta, la tabla de pasaporte incluye una columna denominada “AVAILAB”, donde se indica la disponibilidad de las entradas de la colección original para integrar colecciones núcleo, a través de los códigos 0 (entrada no disponible), 1 (entrada disponible) y NA (sin información/no disponible). La disponibilidad es un criterio propio del curador y podría estar determinado por el número de semillas conservadas, su germinación u otros diversos factores.

11.4.1.11 Parámetro: soloavailab

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea restringir la selección de entradas con destino a la subcolección FIGS exclusivamente a las entradas designadas como disponibles (valor 1 en el campo “AVAILAB”).

11.4.1.12 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución, pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes.

11.4.1.13 Parámetro: buffy

Explicación: Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea que la extracción de información ecogeográfica se haga sobre un área circular alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección. Esta opción es muy útil, por ejemplo, cuando la mayoría de los sitios de recolección corresponden a mercados y no directamente las fincas de los agricultores, o cuando existe bastante incertidumbre sobre la real ubicación de los sitios de recolección a pesar de que GEOQUAL pueda otorgar alta calidad de georreferenciación. La explicación de esta característica se encuentra explicada en el capítulo dedicado a la herramienta ECOGEO.

11.4.1.14 Parámetro: tamp

Explicación: Solo aplica si buffy ha sido señalado como TRUE (✓ en modo on server). Especifique el radio (en kilómetros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor. La extracción se hará sobre aquellas celdas cuyo centroide esté dentro del área circular.

11.4.1.15 Parámetro: variab1v

Explicación: Seleccione una (1) variable ecogeográfica primaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Si elige seleccionar entradas por una o dos variables adicionales (variable secundaria y terciaria), la variable seleccionada aquí será con la que se realice el primer filtrado.

11.4.1.16 Parámetro: variab1rang

Explicación: Solo aplica si variab1cola ha sido señalado como FALSE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable primaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

11.4.1.17 Parámetro: variab1min

Explicación: Solo aplica si variab1rang ha sido señalado como TRUE (✓ en modo on server). Especifique el valor mínimo para la variable primaria con la que determinar el rango que se usará para seleccionar entradas con destino a la

subcolección FIGS.

11.4.1.18 Parámetro: **variab1max**

Explicación: Solo aplica si **variab1rang** ha sido señalado como TRUE (✓ en modo on server). Especifique el valor máximo para la variable primaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

11.4.1.19 Parámetro: **variab1cola**

Explicación: Solo aplica si **variab1rang** ha sido señalado como FALSE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable primaria usando una fracción de la distribución, es decir, un porcentaje de la colección original cuyos valores son los más altos o bajos respecto a la variable primaria.

11.4.1.20 Parámetro: **variab1vpor**

Explicación: Solo aplica si **variab1cola** ha sido señalado como TRUE (✓ en modo on server). Determine la fracción de la distribución (en porcentaje) que desea seleccionar para conformar la subcolección FIGS. Valores permitidos entre 0 y 100.

11.4.1.21 Parámetro: **variab1vhl**

Explicación: Solo aplica si **variab1cola** ha sido señalado como TRUE (✓ en modo on server). Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable primaria.

11.4.1.22 Parámetro: **variab2**

Explicación: Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea utilizar una variable secundaria para seleccionar entradas con destino a una subcolección FIGS. Con valores de esta variable se realizará una selección de entradas sobre el subconjunto previamente seleccionado con la variable primaria.

11.4.1.23 Parámetro: **variab2v**

Explicación: Solo aplica si **variab2** ha sido señalado como TRUE (✓ en modo on server). Seleccione una (1) variable ecogeográfica secundaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Puede ser la misma variable primaria.

11.4.1.24 Parámetro: **variab2rang**

Explicación: Solo aplica si **variab2** ha sido señalado como TRUE (✓ en modo on server) y **variab2cola** ha sido señalado

como FALSE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable secundaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

11.4.1.25 Parámetro: **variab2min**

Explicación: Solo aplica si **variab2rang** ha sido señalado como TRUE (✓ en modo on server). Especifique el valor mínimo para la variable secundaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

11.4.1.26 Parámetro: **variab2max**

Explicación: Solo aplica si **variab2rang** ha sido señalado como TRUE (✓ en modo on server). Especifique el valor máximo para la variable secundaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

11.4.1.27 Parámetro: **variab2cola**

Explicación: Solo aplica si **variab2** ha sido señalado como TRUE (✓ en modo on server) y **variab2rang** ha sido señalado como FALSE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable secundaria usando una fracción de la distribución remanente, es decir, un porcentaje del subconjunto seleccionado por la variable primaria cuyos valores son los más altos o bajos respecto a la variable secundaria.

11.4.1.28 Parámetro: **variab2vpor**

Explicación: Solo aplica si **variab2cola** ha sido señalado como TRUE (✓ en modo on server). Determine la fracción de la distribución remanente (en porcentaje) que desea seleccionar para conformar la subcolección FIGS, usando la variable secundaria. Valores permitidos entre 0 y 100.

11.4.1.29 Parámetro: **variab2vhl**

Explicación: Solo aplica si **variab2cola** ha sido señalado como TRUE (✓ en modo on server). Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable secundaria.

11.4.1.30 Parámetro: **variab3**

Explicación: Solo aplica si **variab2** ha sido señalado como TRUE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea utilizar una variable terciaria para seleccionar entradas con destino a una subcolección FIGS. Con valores de esta variable se realizará una selección de entradas sobre el subconjunto previamente seleccionado con las variables primaria y secundaria. Si no se ha determinado previamente la utilización

de una variable secundaria, seleccionar una terciaria no tendrá ningún efecto en la conformación de la subcolección FIGS.

11.4.1.31 Parámetro: **variab3v**

Explicación: Solo aplica si **variab3** ha sido señalado como TRUE (✓ en modo on server). Seleccione una (1) variable ecogeográfica terciaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Puede ser la misma variable primaria o secundaria.

11.4.1.32 Parámetro: **variab3rang**

Explicación: Solo aplica si **variab3** ha sido señalado como TRUE (✓ en modo on server) y **variab3cola** ha sido señalado como FALSE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable terciaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

11.4.1.33 Parámetro: **variab3min**

Explicación: Solo aplica si **variab3rang** ha sido señalado como TRUE (✓ en modo on server). Especifique el valor mínimo para la variable terciaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

11.4.1.34 Parámetro: **variab3max**

Explicación: Solo aplica si **variab3rang** ha sido señalado como TRUE (✓ en modo on server). Especifique el valor máximo para la variable terciaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

11.4.1.35 Parámetro: **variab3cola**

Explicación: Solo aplica si **variab3** ha sido señalado como TRUE (✓ en modo on server) y **variab3rang** ha sido señalado como FALSE (en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea seleccionar entradas respecto a la variable terciaria usando una fracción de la distribución remanente, es decir, un porcentaje del subconjunto seleccionado por la variable primaria y secundaria cuyos valores son los más altos o bajos respecto a la variable terciaria.

11.4.1.36 Parámetro: **variab3vpor**

Explicación: Solo aplica si **variab3cola** ha sido señalado como TRUE (✓ en modo on server). Determine la fracción de la distribución remanente (en porcentaje) que desea seleccionar para conformar la subcolección FIGS, usando la variable terciaria. Valores permitidos entre 0 y 100.

11.4.1.37 Parámetro: variab3vhl

Explicación: Solo aplica si variab3cola ha sido señalado como TRUE (✓ en modo on server). Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable terciaria.

11.4.1.38 Parámetro: resultados (sólo para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

11.5. Resultados de FIGS_R

Después de definir todos los parámetros y rutas (para modo local) que requiere FIGS_R, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta. FIGS_R producirá entre tres y cinco tablas.

Cuando no se incluye un mapa ELC en el análisis (parámetro controlelc), sólo aparecerán las siguientes tres tablas:

11.5.1 “FIGS regular.txt/xls”

Ésta tabla contiene la identificación de las entradas seleccionadas con destino a la subcolección FIGS (campo “ACCENUMB”), las coordenadas del sitio de recolección (“DECLATITUDE” Y “DECLONGITUDE”), el campo de disponibilidad (“AVAILAB”) y a continuación tantas columnas como variables de selección usadas.

11.5.2 “FIGS stat table.txt/xls”

Corresponde a la tabla que resume las características tanto de la colección original como de la subcolección FIGS a través de estadísticas como la intensidad de selección alcanzada, media, valor máximo y mínimo y diferencial de selección para cada una de las variables de selección utilizadas.

11.5.3 “Passport FIGS R.txt/xls”

Ésta tabla corresponde a la tabla de pasaporte que el usuario introduce al análisis más un campo adicional por cada variable de selección utilizada llamado “SEL_VAR” seguido por el número 1, 2 o 3. Estos campos señalan con el número 1 las entradas que hacen parte de la subcolección FIGS y con “NA” las que no fueron seleccionadas. Así, en el proceso selectivo con la variable primaria (definida en el parámetro 11.4.1.15) se habrían seleccionado las entradas identificadas con un “1” en el campo “SEL_VAR1”, si se ha utilizado una variable de

selección secundaria, las entradas seleccionadas en el segundo proceso de filtrado aparecerán identificadas con un “1” en el campo “SEL_VAR2”, y finalmente si se ha utilizado una variable de selección terciaria, las entradas seleccionadas en el tercer proceso de filtrado aparecerán identificadas con un “1” en el campo “SEL_VAR3”.

Cuando se incluye un mapa ELC para conocer las características ecogeográficas de la subcolección FIGS generada, aparecerá una nueva tabla:

11.5.4 “FIGS_freq_ELCmap.txt/xls”

Esta tabla muestra valores de frecuencia, número de duplicados y número de entradas disponibles para cada categoría ecogeográfica de manera similar a la tabla de la herramienta ColNucleo (“CoreCollect_stats.txt/xls”). Adicionalmente aparecen en la parte izquierda de la tabla hasta tres nuevos campos identificados con el prefijo “FIGS_var” y a continuación los números 1, 2 o 3. De esta manera, en el campo “FIGS_var1” aparecerá el número de entradas seleccionadas por la variable primaria que fueron recolectadas en cada categoría ELC, en “FIGS_var2” el número de entradas seleccionadas por la variable secundaria en el segundo proceso de filtrado de cada categoría ELC y en “FIGS_var3” el número de entradas seleccionadas por la variable terciaria en el tercer proceso de filtrado de cada categoría ELC.

Finalmente, si para las variables de selección primaria, secundaria y terciaria se ha utilizado exclusivamente la segunda forma de selección (fracción de la colección), es decir, se ha marcado la opción “variab1cola”, “variab2cola” y “variab3cola” (parámetros 11.4.1.19, 11.4.1.27 y 11.4.1.35 respectivamente), los resultados incluirán una quinta tabla:

11.5.5 “FIGS_UnderELC.txt/xls”.

Corresponde a una tabla con los mismos campos descritos anteriormente para la tabla “Passport_FIGS_R.txt” (apartado 11.5.3) pero en este caso sólo contiene las entradas que constituyen la subcolección FIGS balanceada mediante el mapa ELC, e incluyen los campos “SEL_VAR1”, “SEL_VAR2” y “SEL_VAR3” que indican con el valor 1 si esas entradas hubieran sido también seleccionadas en un esquema FIGS sin el uso de mapas ELC. Adicionalmente, al costado izquierdo de la tabla aparecen hasta 3 nuevos campos, identificados como “var_eco1”, “var_eco2” y “var_eco3” dependiendo de cuantas variables de selección se hayan utilizado. En cada uno de estos campos aparecerán los valores correspondientes a la extracción de cada variable de selección para cada sitio de recolección (“var_eco1” valores de la variable primaria, “var_eco2” valores de la variable secundaria y “var_eco3” valores de la variable terciaria).

Adicionalmente, la tabla “FIGS_freq_ELCmap.txt/xls” (apartado 11.5.4) incluirá hasta tres nuevos campos en el costado izquierdo, con las denominaciones “No_by_var1”, “No_by_var2” y “No_by_var3”. En estos campos se muestra el número de entradas seleccionadas para la subcolección FIGS balanceada por el mapa ELC en cada proceso de selección, “No_by_var1” para el primer proceso de filtrado por la variable primaria, “No_by_var2” para el segundo proceso de filtrado por la variable secundaria y “No_by_var3” para el tercer proceso de filtrado por la variable terciaria.

11.6. Referencias

El Bouhssini, M. E., Street, K., Joubi, A., Ibrahim, Z., Rihawi, F. 2009. Sources of wheat resistance to Sunn pest, *Eurygaster integriceps* Puton, in Syria. *Genetic Resources and Crop Evolution* 56: 1065– 1069.

Endresen, D.T.F. 2010. Predictive association between trait data and ecogeographic data for Nordic barley landraces. *Crop Science* 50: 2418-2430.

Endresen, D.T.F., Street, K., Mackay, M., Bari, A., Amri, A., De Pauw, E., Nazari, K., Yahyaoui, A. 2012. Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using Focused Identification of Germplasm Strategy. *Crop Science* 52: 764-773.

FAO 2010. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome

Khazaei, H., Street, K., Bari, A., Mackay, M., Stoddard, F.L. 2013. The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS ONE* 8(5): e63107. doi:10.1371/journal.pone.0063107

Mackay, M.C. 1990. Strategic planning for effective evaluation of plant germplasm. p. 21-25 En: Srivastava, J.P., Damania, A.B. (eds). *Wheat genetic resources: Meeting diverse needs*. John Wiley & Sons, Chichester, UK.

Mackay, M. C., Street, K. 2004. Focused identification of germplasm strategy – FIGS. p. 138-141. En: Black, C.K., Panozzo, J.F., Rebetzke, G.J. (eds). *Cereals 2004. Proceedings of the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders' Assembly, 21-24 September 2004, Canberra, Australian Capital Territory (ACT)*. Cereal Chemistry Division, Royal Australian Chemical Institute, Melbourne, Australia.



Taller Nacional, Colonia, Uruguay, Abril 2014.



12 | Herramienta rLayer

12.1. Creación de capas de información “a medida”

Resulta frecuente que la distribución de las recolecciones almacenadas en una colección de germoplasma o datos de presencia de poblaciones de una especie superen las fronteras de los estados. A veces lo que ocurre es que las distribuciones son muy limitadas y ni siquiera alcanzan los bordes de una provincia. Esto puede suponer un problema a la hora de aplicar las herramientas CAPFITOGEN, ya que la información ecogeográfica (capas) viene organizada por países, con algunas excepciones en las que se ofrece cobertura global o para subcontinentes (Mesoamérica o Suramérica), continentes (Europa).

Para estos casos se ha creado la herramienta rLayer. Con ella el usuario podrá crear conjuntos de capas de información (para 177 variables, excluyendo latitud y longitud que no son necesarias en los análisis que hace CAPFITOGEN con las capas SIG) que cubran la distribución de sus recolecciones o de las poblaciones/individuos que reflejan sus datos de pasaporte, sin importar si éstas sobrepasan o no las fronteras de los países.

Estos conjuntos son muy útiles para poder realizar análisis más realistas, ajustados o precisos con herramientas como ELCmapas, ECOGEO, DIVmapas, ColNucleo, FIGS_R o Modela.

12.2. Cómo funciona rLayer

A partir de CAPFITOGEN3, la herramienta rLayer crea nuevos conjuntos de capas de información ecogeográfica recortando las capas de cobertura global a medida de los datos de distribución de recolecciones o presencias de poblaciones del usuario, a través de tres vías que se describen a continuación y se ilustran en la Fig. 44:

- Método “polygon”: el usuario proporciona a la herramienta un archivo de mapa vectorial (shapefile) de tipo polígono, usualmente compuesto de un solo polígono, el cual servirá de molde para recortar las 177 capas de información ecogeográfica. Este mapa molde debe estar en el sistema de coordenadas WGS84 y debe representar áreas terrestres mayores a la resolución seleccionada para el conjunto de capas ecogeográficas recordadas (1x1, 5x5, 10x10 o 20x20 km).
- Método “square”: el molde para recortar las capas de información ecogeográfica resulta de la distribución de los sitios de recolección o presencia de poblaciones de la tabla de pasaporte. En este caso se dibuja un rectángulo donde cada lado coincide con los valores máximos y mínimos de latitud y longitud de la distribución.
- Método “buffer”: el molde para recortar las capas de información ecogeográfica resulta de áreas circulares alrededor de cada sitio de recolección o presencia de poblaciones de la tabla de pasaporte. Esta área se traza a partir de un valor de radio que el usuario define. Las áreas circulares que se solapan crean áreas contiguas.

El usuario debe indicar a la herramienta la resolución que desea para el nuevo conjunto de capas, que debe concordar con las resoluciones disponibles según el archivo descargado (1x1, 5x5, 10x10 o 20x20 km). También se deberá indicar el nombre para el nuevo conjunto de capas, eligiendo entre “user1”, “user2” o “user3” (para modo local) o cualquier nombre diferente al de un país que desee el usuario (modo on server). Una característica muy importante de rLayer en su actualización para CAPFITOGEN3 es que las capas que recorta y que produce están en formato .tif, no en formato .grd. Esto implica un ahorro en espacio en disco.

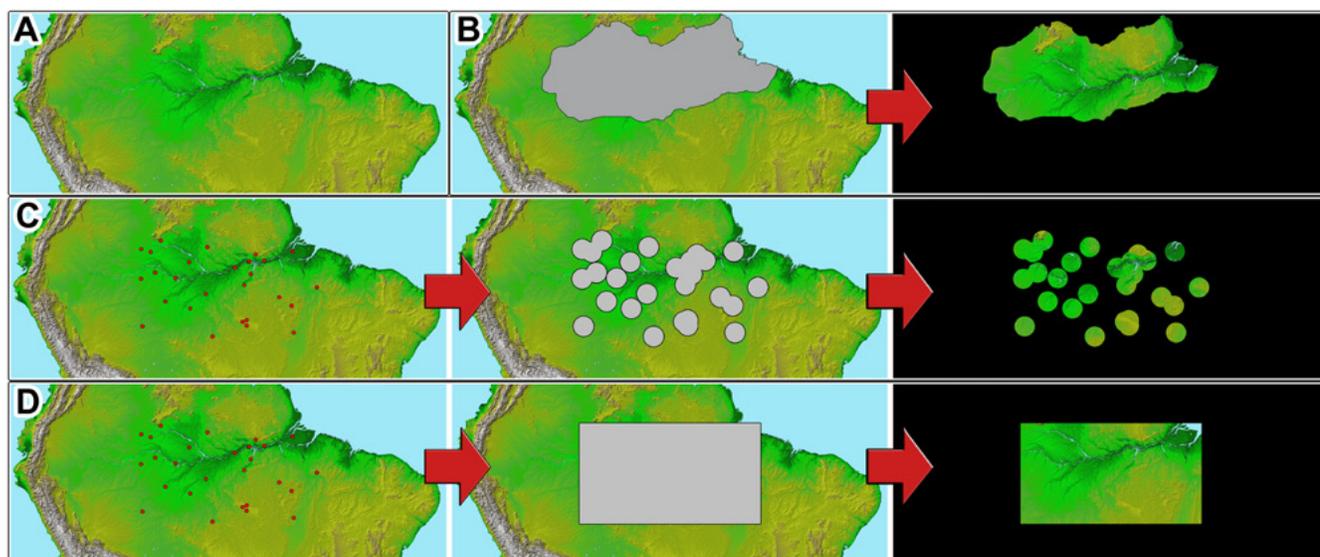


Figura 44. Ilustración del proceso de recorte de una capa realizada por rLayer. En fondo oscuro se observa el resultado del proceso de corte en cada caso. A. Capa global original (1x1 km) para contenido de elevación (altitud) que va a ser recortada. B. Proceso de recorte a través de un shapefile que contiene un polígono suministrado por el usuario (método “polygon”). C. Proceso de recorte basado en áreas circulares trazados alrededor de los sitios de recolección de germoplasma u observación de poblaciones (método “buffer”). D. Proceso de recorte basado en un rectángulo trazando sus lados con los valores máximos y mínimos de latitud y longitud de los puntos de la distribución de sitios de recolección o poblaciones observadas (método “square”).

Una vez finalizado el proceso, el usuario puede seleccionar en diferentes herramientas como ECOGEO, DIVmapas, ColNucleo o FIGS_R, en el parámetro “pais”, el nuevo conjunto de capas. Sin importar que nombre haya elegido el usuario, en el parámetro “pais” aparecerán al final del listado de países y regiones las tres opciones de rLayer (“user1”, “user2” y “user3”) en el modo local. Tanto en modo local o modo on server, el usuario debe recordar el nombre que asignó en rLayer y entonces seleccionarlo, si su deseo es usar el nuevo conjunto de capas.

Así mismo, el usuario debe recordar las resoluciones disponibles de acuerdo a los archivos descargados y descomprimidos, pues el tamaño de celda será requerido en el parámetro “resol1”, el cual aparece en las herramientas que usan capas de información ecogeográfica.

Para el método “square” rLayer identifica los valores máximos y mínimos de latitud y longitud de la distribución de recolecciones o presencias que el usuario introduce en el parámetro “pasaporte” y añade un ‘borde’ (sumando a máximos y restando a mínimos) que según las resoluciones de las capas globales serán: 1x1 km = 0.05 grados decimales (aproximadamente 6 km en el Ecuador), 5x5 km = 0.10 grados decimales (aproximadamente 11 km en el Ecuador), 10x10 km = 0.15 grados decimales (aproximadamente 17 km en el Ecuador) y 20x20 km = 0.20 grados decimales (aproximadamente 22 km en el Ecuador). Con los valores finales de mínimos y máximos (más o menos los bordes), rLayer recorta una a una las 177 capas de información ecogeográfica de cubrimiento global que se encuentran en la

carpeta “world” de la ruta “CAPFITOGEN2/rdatamaps/world” y las guarda en una carpeta en la misma ruta con el nombre seleccionado (“user1”, “user2” o “user3”) para el modo local. Para el modo on server, las variables quedarán disponibles para el usuario en el parámetro “país”, desde la lista desplegable con los nombres de países. Allí el usuario debe buscar el nombre que asignó en la ejecución de rLayer al conjunto de capas ecogeográficas recortadas de manera personalizada.

12.3. Utilización de la herramienta rLayer

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta rLayer, se deben especificar una serie de parámetros por parte del usuario.

12.3.1 Parámetros iniciales definidos por el usuario

12.3.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

12.3.1.2 Parámetro: cropway

Explicación: Seleccione uno de los tres métodos para recortar las capas: “polygon”, “square” o “buffer”.

12.3.1.3 Parámetro: buffer

Explicación: Solo aplica si cropway = “buffer”. Parámetro numérico en kilómetros (km) y con el que se indica el radio que se usará para generar las áreas circulares alrededor de cada coordenada/sitio de recolección o población observada.

12.3.1.4 Parámetro: shapefile

Explicación: Solo aplica si cropway = “polygon”. En este parámetro se debe indicar el nombre del archivo shapefile, que debe estar en sistema de coord lat-long WGS84 y contener un sólo polígono por el cual se quiere cortar. El shapefile, compuesto entre cuatro a siete archivos, entre los que debe estar los imprescindibles .shp, .dbf y .shx) se deben encontrar en la carpeta Pasaporte (CAPFITOGEN3/Pasaporte) en modo local. En modo on server los archivos del shapefile deben ser subidos previamente a la zona de archivos de usuario y resultados, particularmente en la carpeta “uploads”. De esta manera el usuario sólo debe especificar en este parámetro el nombre del archivo.

12.3.1.5 Parámetro: pasaporte

Explicación: Solo aplica si cropway = “square” o cropway = “buffer”. Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados

12.3.1.6 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro ‘pasaporte’) la tabla resultado de GEOQUAL denominada “PasaporteOriginalEvaluadoGEOQUAL.txt” con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

12.3.1.7 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

12.3.1.8 Parámetro: resol1

Explicación: Seleccione el nivel de resolución de las capas de información ecogeográfica que desea recortar. En modo local, asegúrese que cuenta con las capas a nivel global (carpeta “world” dentro de “rdatamaps”) en la resolución que indica en este parámetro (carpetas 1x1, 5x5, 10x10 o 20x20).

12.3.1.9 Parámetro: unname

Explicación: Para modo local, seleccione uno de los siguientes nombres para el nuevo conjunto de capas recortadas: “user1”, “user2” o “user3”. Para modo on server, elija un nombre sencillo sin espacios ni caracteres especiales.

12.4. Resultados de rLayer

Una vez finalizado el análisis y de acuerdo a la configuración de los parámetros anteriormente detallados, rlayer producirá el nuevo conjunto de capas ecogeográficas y las guardará en la carpeta “user1”, “user2” o “user3” dentro de la carpeta “rdatamaps” (modo local) o quedarán disponibles dentro de la lista desplegable de países y conjuntos de capas de parámetros como “pais” (modo on server).



BANCO PORTUQUÊS
DE
CERNOPLASMA VEGETAL

Taller Nacional, Braga, Portugal, Junio 2015.



13 | Herramienta Complementa

13.1. Conservación *in situ* de la agrobiodiversidad

La estrategia de conservación *in situ* de biodiversidad es la única que permite la evolución natural y adaptación de los organismos a la dinámica cambiante de las condiciones ambientales.

La conservación *in situ* para el caso de los recursos fitogenéticos para la alimentación y la agricultura tiene dos vertientes claramente definidas:

- Áreas protegidas donde se conserva todo el ecosistema en su estado natural con la esperanza de que las especies silvestres que allí ocurren mantengan su viabilidad a largo plazo
- Fincas de agricultores o de comunidades agrícolas, donde variedades tradicionales y/o locales (principalmente, aunque también se pueden conservar especies parientes silvestres) se mantienen en las condiciones agroecológicas que les dieron lugar o que les ha permitido subsistir.

Constantemente se insiste en la necesidad de complementar los esfuerzos de conservación *ex situ* con actividades y proyectos de conservación *in situ* para las mismas especies objetivo. Uno de los documentos que inicialmente expuso esta necesidad de forma más nítida fue el primer Plan de Acción Mundial (FAO, 1997). En el año 2010, el segundo Plan de Acción Mundial hace de nuevo un llamado para la integración de esfuerzos *in situ-ex situ*. Este documento indica que las estrategias de conservación “alcanzan la máxima eficacia cuando son complementarias y están bien coordinadas. Es necesario integrar plenamente a todos los niveles la conservación *in situ* y *ex situ* y la utilización sostenible” (FAO, 2012).

Respecto a la conservación *in situ* de la agrobiodiversidad, las actividades más importantes señaladas en los planes de acción son las siguientes:

- Estudiar e inventariar los recursos fitogenéticos para alimentación y agricultura.
- Proveer apoyo para el manejo y mejoramiento en fincas.
- Asistir a los agricultores en situaciones de desastre y restauración de los sistemas productivos que favorecen la conservación.
- Promover la conservación *in situ* de especies parientes silvestres.

El objetivo de la primera actividad es facilitar la elaboración, aplicación y vigilancia de las estrategias de conservación y elaborar y aplicar métodos para la preparación de inventarios de los RFG conservados *in situ* y *ex situ*, incluyendo SIG y marcadores moleculares. Con estos estudios se busca identificar, localizar, catalogar y evaluar las amenazas hacia la agrobiodiversidad, en particular las derivadas del uso de la tierra y el cambio climático (FAO, 2012).

Estas disposiciones hacen necesarios los medios tecnológicos suficientes para determinar, evaluar y monitorear los sitios destinados para actividades de conservación *in situ*, ya sea en áreas protegidas como en fincas.

13.2 Concepto de complementariedad

A partir de los años 70, la selección de sitios para la implantación de áreas protegidas para la conservación de la biodiversidad empezó a plantearse como un proceso metodológico, donde en su fase más importante se aplican una

serie de criterios para la priorización de áreas. Uno de esos criterios ha sido el valor estimado de conservar diversidad biológica, expresado de diferentes maneras, por ejemplo la riqueza de taxa o directamente la diversidad, la singularidad, o la representatividad. La planeación de este tipo de actividades empezó a migrar de una aproximación totalmente intuitiva a una algorítmica en la cual los criterios considerados fueron aplicándose mecánicamente hasta eliminar todo concepto intuitivo. Esta migración fue soportada por la aparición de computadores cada vez más sofisticados al alcance de los investigadores y al surgimiento de los Sistemas de Información Geográfica (SIG) (Justus y Sarkar, 2002).

El diseño de redes de áreas protegidas para la conservación de la biodiversidad buscan maximizar el número de especies conservadas en el menor área posible (Kati *et al.*, 2004). Un aspecto clave en estos diseños es el uso del principio de complementariedad, el cual asegura que las áreas seleccionadas para integrar una red complementan aquellas seleccionadas previamente. El diccionario de la lengua española define complemento como “cosa, cualidad o circunstancia que se añade a otra para hacerla íntegra o perfecta”. Esta es la idea principal del concepto de complementariedad y que se ilustra en la Fig 45.

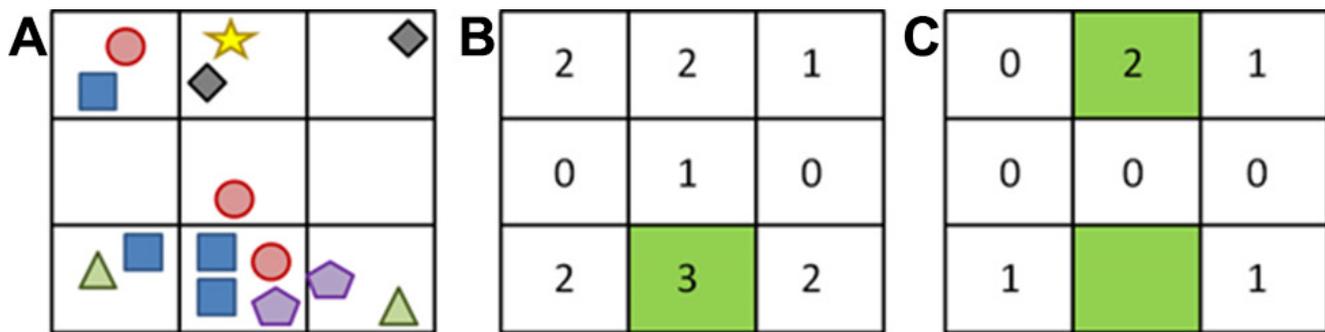


Figura 45. Cómo opera el principio de complementariedad. A. Seis formas diferentes representando taxa distribuidos en nueve celdas. B. Conteo por celda del número de taxa diferentes que ocurren en cada una, seleccionando la de mayor riqueza (tres taxa). C. La celda que cubre más taxa diferentes y complementarias a la celda seleccionada en el paso B es la resaltada en verde con el número “2” (dos taxa diferentes entre sí y complementarias). En este paso se estaría aplicando el principio de complementariedad.

La primera aproximación hacia el uso de la complementariedad para el diseño de redes de áreas protegidas fue realizado en 1983 en Tasmania, Australia. Allí se planteó que procedimientos no iterativos podrían no ser adecuados para la representación de la mayor cantidad de taxa en el menor número posible de sitios. Esto planteo la necesidad de reemplazar el uso del criterio de riqueza por un procedimiento iterativo que contemple la complementariedad entre las áreas. Posteriormente fueron apareciendo gradualmente algoritmos para procedimientos iterativos de selección de áreas usando como base la complementariedad (Justus y Sarkar, 2002).

En 1990 Rebelo y Siegfried (1990) desarrollaron un algoritmo para la designación de sitios para la protección de la vegetación en la región florística del Cabo (Suráfrica). Este algoritmo inicia con la selección de una zona en forma de celda basada en endemismo y riqueza (llamada en adelante “celda núcleo”), lo cual se logra sobreponiendo la dis-

tribución de los taxa objetivo sobre una grilla de celdas que cubren el territorio. Entonces los taxa cubiertos por la celda núcleo son retirados de la distribución total de taxa. Así se cierra el primer paso del algoritmo. Posteriormente aquella celda con el mayor número de taxa diferentes entre sí, las cuales no están presentes en la celda núcleo es seleccionada y se denomina “celda núcleo secundaria”. Los taxa presentes en esta celda también son retirados de la distribución total, dando por terminado el segundo paso. Paso a paso el algoritmo termina cuando se selecciona una celda para el último taxón restante no cubierto en el paso anterior. En el caso de producirse ‘empates’ (mismo número de taxa complementarios en dos o más celdas), se puede usar diferentes estrategias para el desempate: selección al azar, donde la riqueza de taxa sea mayor (considerando taxa complementarios y no complementarios), o donde se den mayor número de taxa endémicas, etc. Este procedimiento iterativo se le conoce comúnmente como el algoritmo de complementariedad de Rebelo y es ampliamente usado en la detección de áreas prioritarias para conservación, así como la evaluación de redes de áreas protegidas ya constituidas.

Si bien el algoritmo de complementariedad de Rebelo ha sido planteado originalmente sobre grillas que dividen el territorio en celdas de igual tamaño, para el caso de evaluaciones de redes de áreas ya constituidas, el mismo puede ser implementado sobre áreas de diversas formas.

Un análisis de complementariedad puede ser usado también para determinar áreas prioritarias para la recolección de germoplasma para ser conservado *ex situ*, una utilización particularmente útil para especies parientes silvestres de cultivadas. En este caso la determinación de zonas de alta riqueza de taxa y complementarias entre sí llevaría a recolectar en áreas relativamente pequeñas con un alto rendimiento en términos de taxa recolectados.

13.3 Concepto de cobertura

El concepto de cobertura de la protección de redes de áreas protegidas sobre la biodiversidad es mucho más simple e intuitivo que el de complementariedad. Basado en los preceptos de número y riqueza de taxa conservados, un análisis de cobertura sencillamente busca evaluar de forma rápida el grado de protección que ofrece una red de áreas protegidas y cada área que la integra.

De esta manera Parra-Quijano *et al.* (2003), determinaron mediante un análisis SIG el grado de cobertura que ofrecía los lugares de interés comunitario de la red Natura 2000 sobre seis especies del género *Lupinus* (algunas cultivadas y otras parientes silvestres) que ocurren en España Peninsular e Islas Baleares. De acuerdo a este estudio, un tercio de las poblaciones de estas especies de *Lupinus* están potencialmente protegidas por esta red de áreas y se determinaron cuáles eran las áreas protegidas que más número de especies diferentes conservaba y que especie estaban mejor conservadas que otras. En principio esta red no fue creada para la conservación de especies cultivadas nativas ni sus parientes silvestres, y por ello se entiende el bajo nivel de cobertura y que sólo para el 1.47 % de áreas de esta red se tiene constancia de presencia de especies en su interior.

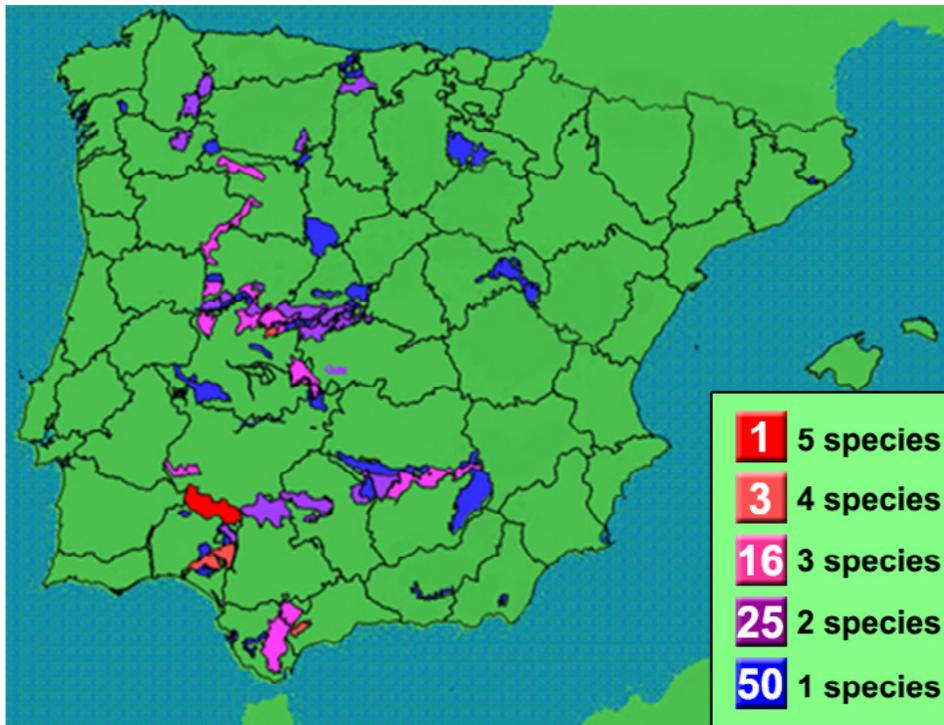


Figura 46. Mapa resultado del análisis de cobertura para seis especies de *Lupinus* en España usando los sitios de interés comunitario como red de áreas protegidas a evaluar. Aquí se muestran las áreas que mayor riqueza de especies conservan.

13.4. Cómo funciona Complementa

La herramienta para análisis de complementariedad y evaluaciones de cobertura Complementa requiere que el usuario suministre los datos de presencia para las taxa que quiere analizar. Los datos de presencia pueden corresponder a recolecciones de germoplasma o poblaciones referenciadas en literatura, bases de datos sobre biodiversidad, herbarios, etc. El formato requerido para introducir esta información, independientemente de su procedencia es el de pasaporte FAO-Bioersity 2012 con la adición de dos nuevos campos (de tipo opcional y sólo útiles para ciertos tipos de análisis de complementariedad). Sin embargo, a diferencia de otras herramientas, el formato de 'pasaporte' Complementa tiene un número de campos obligatorios bastante reducido. Los campos obligatorios son:

- ACCENUMB
- GENUS
- SPECIES
- SUBTAXA
- DECLATITUDE
- DECLONGITUDE

A medida que el análisis se hace más complejo, la tabla requiere más información. El análisis de complementariedad normal que realiza Complementa tiene como objeto una serie de taxa (géneros, especies y subespecies o variedades) y determina sitios de alta riqueza para éstos taxa y que sean complementarios entre sí.

13.4.1 ¿Análisis taxonómicos o taxonómicos-ecogeográficos?

La herramienta Complementa usualmente realiza análisis considerando sólo el aspecto taxonómico contenido en la tabla 'pasaporte' usada para ingresar la información de presencia de los taxa. Sin embargo, Complementa también puede llevar a cabo análisis de complementariedad para una combinación entre taxa y los escenarios adaptativos donde ocurre cada población (i.e. categorías de un mapa ELC). Esta variante requiere que el usuario haya desarrollado previamente un mapa ELC (herramienta ELC mapas) y luego haber usado la herramienta Representa con los sitios de ocurrencia que desea analizar con Complementa y con este mapa ELC. Uno de los resultados de la herramienta Representa es la tabla 'Genbank_ELC.txt'. Esta tabla contiene el campo llamado 'BGcat' que corresponde a la categoría del mapa ELC extraída según las coordenadas. Para poder realizar el análisis de complementariedad combinando taxa y escenario ecogeográfico, es necesario que la información del campo 'BGcat' pase al campo 'BG_ELC' del formato de 'pasaporte' de Complementa, teniendo cuidado de copiar la información entre las tablas correctamente (usando un campo de identificación común como 'ACCENUMB'). De esta manera Complementa podrá combinar la información y el análisis detectará áreas de alta riqueza y complementarias desde un punto de vista taxonómico y adaptativo (abiótico) a la vez.

13.4.2 ¿Celdas o áreas?

Complementa puede realizar análisis de complementariedad sobre el territorio dividido en celdas (grilla) de igual tamaño o sobre áreas pre-existentes. La idea recurrente cuando se elige realizar el análisis sobre celdas, es la de realizar un estudio de tipo exploratorio para detectar en donde se da esa doble condición (riqueza + complementariedad). Por otra parte, cuando el análisis se realiza sobre redes de áreas protegidas ya existentes, lo que se busca usualmente es dar un valor añadido a algunas de las áreas de la red debido a su capacidad de protección de una mayor biodiversidad en un menor espacio. En el segundo caso, el análisis no mostraría necesariamente las zonas dentro del territorio objetivo con mayor riqueza y complementariedad entre sí.

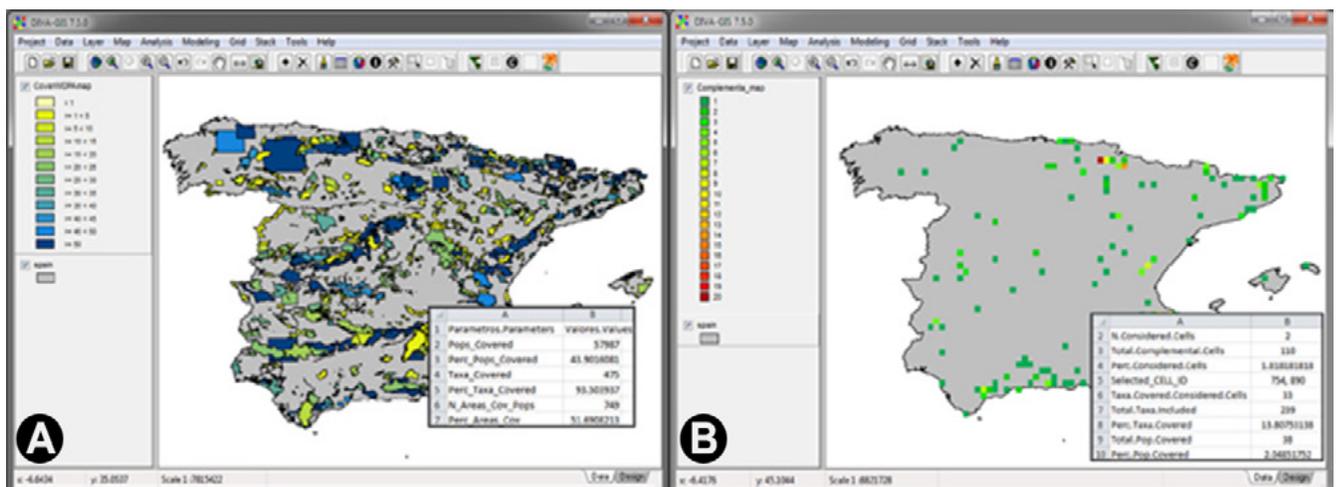


Figura 47. Mapas resultado de un análisis de complementariedad. A. Para una red de áreas protegidas, B. Para celdas.

13.4.3 Usando complementa para recolección optimizada de germoplasma

Si el análisis de complementariedad es usado para determinar sitios de recolección es necesario contar con información de faltantes ecogeográficos de la colección. Esto implica que será necesario la utilización de ELC mapas (para la obtención del mapa ELC) y Representa (con datos de recolecciones de la colección de germoplasma objetivo y de fuentes externas) para detectar áreas de alta riqueza y complementarias desde un punto de vista taxonómico y donde además ocurren poblaciones consideradas faltantes ecogeográficos en la colección de germoplasma. Para este tipo de análisis de complementariedad, el usuario deberá aportar como datos de 'pasaporte', solo los faltantes ecogeográficos contenidos en la tabla "Tabla_Fuentes_Externas_clasificadas_ExternalSourcesClassified.txt" resultado de la herramienta Representa. Por tanto, el contenido de la tabla resultado de Representa debe ser trasladado correctamente a la tabla de tipo 'pasaporte' que se usa en Complementa, cuidando de trasladar también la información de la columna "Tipo_falt/Gap_type" que aparece en la tabla de Representa, en la columna GAPTYPE que aparece en 'pasaporte' de Complementa. Esto indica que la columna GAPTYPE del formato de 'pasaporte' exclusivo de Complementa sólo se deberá rellenar cuando se use Complementa para recolecciones optimizadas y en esta columna sólo pueden aparecer los valores indicados en la Tabla 1 o valores NA.

Posteriormente, en el momento de llenar la información sobre los parámetros que requiere Complementa, algunos de ellos preguntarán al usuario cómo usar la información que aparece en la columna GAPTYPE, como por ejemplo cual es el valor umbral bajo el cual se considera que una población de fuentes externas se considera vacío o faltante ecogeográfico prioritario, o si desea excluir del análisis de complementariedad casos en GAPTYPE con valores NA (que no han sido analizados con la herramienta Representa).

13.4.4 Usando redes de áreas protegidas de la Base de Datos Mundial o redes de áreas aportadas por el usuario

Cuando un usuario ha decidido que desea realizar un análisis de complementariedad sobre áreas, Complementa le permite usar dos tipos de redes de áreas para tal fin:

13.4.4.1 Redes de áreas de la Base de Datos Mundial de Áreas Protegidas (World Database on Protected Areas -WDPA). Se trata de la base de datos mundial más completa sobre áreas terrestres y marinas protegidas, un proyecto de la IUCN y la UNEP, y que pueden visualizarse fácilmente desde la interfaz del proyecto Protected Planet (<http://www.protectedplanet.net/>). Dado que CAPFITOGEN3 se diseñó con el fin de facilitar los análisis a cualquier usuario, donde los conocimientos requeridos para operar las herramientas sean mínimos, se ha descargado esta información de cobertura global desde este portal y se ha dispuesto por países, recortándose e incluyéndose sólo áreas protegidas terrestres o las fracciones terrestres de áreas protegidas marinas. Para que la herramienta Complementa pueda usar estas áreas para un país o región determinada en modo local, el usuario debe seguir los siguientes pasos:

- Visitar la página web del Programa CAPFITOGEN donde se describe cómo descargar las herramientas y la información acompañante: <http://www.capfitogen.net/es/acceso/informacioncomplementa>
- En esta página se muestra un vínculo que lleva a un servicio de descarga de información ajustada y recortada de la WDPA. Hacer click sobre él.

- Aparecerá una lista de archivos con extensión “.RData” y nombres de países y regiones. El usuario debe descargar el archivo del país objetivo. Haciendo click sobre su nombre, la descarga iniciará.
- Guarde el archivo descargado en la carpeta “wdpa” que aparece en la estructura de carpetas y archivos CAPFITOGEN3 (usualmente X:/CAPFITOGEN3/wdpa siendo X la letra del disco).
- El usuario ya puede seleccionar el país o región para la cual ha descargado la información sobre áreas WDPA en el parámetro “pais” al momento de ejecutar la herramienta Complementa. Nunca se debe seleccionar un país o región para la cual no se cuente con un archivo RData en la carpeta “wdpa”.

Para Complementa en modo on server, el sistema ya tiene cargados todos los archivos .RData para países y regiones y por ello en el parámetro “pais” donde se pregunta por el nombre del país o región para el cual se cuenta con áreas WDPA, este sólo debe seleccionarse de una lista desplegable.

13.4.4.2 Redes de áreas protegidas suministradas por el usuario. En este caso, el usuario aporta una capa SIG con polígonos que representan las áreas protegidas que serán evaluadas. Para ambos modos (local y on server) el archivo debe tener las siguientes características:

- Debe ser una capa tipo “shapefile” (formato vectorial) de polígonos
- Este “shapefile” debe estar constituido por al menos tres archivos. Utilizando como ejemplo el nombre de ‘AreasProt’, el shapefile al menos debe tener los siguientes archivos: “AreasProt.shp”, “AreasProt.dbf” y “AreasProt.shx”. Sin embargo, un “shapefile” completo contendrá además de los anteriores, los siguientes archivos (no esenciales para Complementa): “AreasProt.sbn”, “AreasProt.sbx” y “AreasProt.prj”. El “shapefile” debe estar en el sistema de coordenadas Geodésico Mundial de 1984, más conocido como WGS84 (World Geodetic System 84), o geográfico Latitud-Longitud.
- Previo al uso del “shapefile”, el usuario debe comprobar que la georeferenciación de la capa es correcta, sobrepóniéndola sobre el “shapefile” de los límites del país (que se encuentran en la carpeta “MapasApoyo_BoundariesMaps” del conjunto de archivos y carpetas CAPFITOGEN). Para ello, el usuario se puede valer del programa DIVA-GIS (<http://www.diva-gis.org>).
- Un “shapefile” de polígonos contiene una tabla con información organizada (en columnas) para cada polígono (filas). El “shapefile” de áreas protegidas del usuario debe incluir en su tabla una columna de identificación de cada polígono, un código que puede ser numérico, alfabético o ambos, y que identifica inequívocamente cada polígono (área protegida). Identifique el nombre de esa columna de identificación con la ayuda del programa DIVA-GIS, por ejemplo. En su momento, la interfaz de la herramienta Complementa preguntará al usuario el nombre del “shapefile” y el nombre de la columna de identificación.

La siguiente instrucción debe ser tenida en cuenta para el modo local:

- Los archivos correspondientes al “shapefile” del usuario se deben copiar en la carpeta “wdpa” del conjunto de archivos y carpetas CAPFITOGEN.

Y la siguiente instrucción debe ser tenida en cuenta para el modo on server:

- Los archivos correspondientes al “shapefile” del usuario se deben subir a la zona de archivos de usuario y resultados, particularmente en la carpeta “uploads”.

13.4.5 Análisis de cobertura

La herramienta Complementa puede utilizar la misma información que el usuario introduzca para realizar el análisis de complementariedad para áreas (no disponible para análisis de celdas) para obtener un análisis de cobertura y así poder realizar una evaluación más completa de la red de áreas protegidas.

13.5. Utilización de la herramienta Complementa

Los usuarios pueden revisar algunas publicaciones donde se ha utilizado Complementa, sola o en conjunto con otras herramientas CAPFITOGEN como: Phillips *et al.* (2016), García *et al.* (2017), Rubio-Teso *et al.* (2019) o Mponya *et al.* (2021). Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Complementa, se deben especificar una serie de parámetros por parte del usuario.

13.5.1 Parámetros iniciales definidos por el usuario

13.5.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentra la herramienta CAPFITOGEN. Nota: use / antes que . Por ejemplo F:/, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

13.5.1.2 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta 'Pasaporte' pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama 'tabla.txt' y se encuentra en una carpeta llamada 'Stuberosum' que se encuentra dentro de 'Pasaporte', entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para Complementa, esta tabla debe tener unos campos adicionales como BG_ELC (categoría del mapa ELC asignada por Representa) y GAPATYPE (tipo de vacío o faltante ecogeográfico determinado por Representa). Se recomienda que el usuario disponga de los archivos Excel especialmente preparados para Complementa con las columnas requeridas, los cuales se encuentran en <https://drive.google.com/drive/folders/1xCnllZgzW0uDeClDvcxbADv9H583xzpn?usp=sharing>

13.5.1.3 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

13.5.1.4 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

13.5.1.5 Parámetro: gapttype

Explicación: Sólo aplica para casos de utilización del análisis de complementariedad en priorización de zonas de recolección (dentro de un diseño optimizado de recolección de germoplasma). Seleccione esta opción si su tabla de entrada de datos de presencia tiene valores utilizables en el campo GAPATYPE (producido por Representa) y los desea utilizar para determinar faltantes o vacíos ecogeográficos en su colección.

13.5.1.6 Parámetro: gaptresh

Explicación: Solo aplica si gapttype ha sido señalado como TRUE (✓ en modo on server). Indique el valor de la columna GAPATYPE que desea utilizar como umbral para designar una presencia como faltante o vacío ecogeográfico en su colección. Las presencias con valor de GAPATYPE por debajo de este umbral entrarán en el análisis de complementariedad.

13.5.1.7 Parámetro: gapna

Explicación: Solo aplica si gapttype ha sido señalado como TRUE (✓ en modo on server). Si usa los valores de la columna GAPATYPE para determinar faltantes o vacíos ecogeográficos y dicha columna contiene valores NA (datos de presencia no evaluados por la herramienta Representa), indique si desea incluir (escribir "include") o excluir (escribir "exclude") las presencias con dichos valores.

13.5.1.8 Parámetro: duplicat

Explicación: Indica (TRUE=sí FALSE=no para modo local, ✓=sí □=no en modo on server) si se eliminarán o no los registros de la misma GENUS/SPECIES/SUBTAXA debido a que se encuentran cerca espacialmente unos de otros.

13.5.1.9 Parámetro: distdup

Explicación: Solo aplica si duplicat ha sido señalado como TRUE (✓ en modo on server). Indique numéricamente la distancia (en km) dentro de la cual dos sitios de recolección u observación se considerarán la misma población (duplicado espacial). En ocasiones dos observadores o recolectores han registrado dos puntos que pertenecen a la misma población, pero no lo saben. Sólo hasta cuando se mapean los dos puntos es posible determinar que pertenecen a la misma población por lo cerca que esta un punto del otro, esto es lo que hace distdup. El parámetro distdup no puede ser menor de cero y tampoco puede ser un número exageradamente alto pues haría que todos los puntos son en realidad duplicados y el análisis se haría para un sólo punto, produciendo un error.

13.5.1.10 Parámetro: celdas

Explicación: Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea realizar el análisis de complementariedad sobre un conjunto de celdas (grilla) que cubre la extensión de la distribución de los datos de presencia.

13.5.1.11 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la generación del mapa de complementariedad por celdas. Note que 1x1 km ofrece mayor resolución, pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes

13.5.1.12 Parámetro: nceldas

Explicación: Solo aplica si celdas ha sido señalado como TRUE (✓ en modo on server). El análisis de complementariedad para celdas determina un ranking de celdas de alto valor de cobertura (que incluyen en su área el mayor número de taxa diferentes) y complementarias entre sí. Además, Complementa puede producir algunas estadísticas adicionales para un subconjunto de celdas del ranking. Determine el número de celdas que serán incluidas en el análisis adicional. La selección inicia desde la celda de más alto valor de cobertura y continúa hacia abajo en el ranking.

13.5.1.13 Parámetro: areas

Explicación: Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea realizar el análisis de complementariedad sobre un conjunto áreas prediseñadas.

13.5.1.14 Parámetro: WDPA

Explicación: Solo aplica si areas ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción si desea realizar el análisis de complementariedad para áreas usando polígonos de la Base de Datos Mundial de Áreas Protegidas (WDPA en sus siglas en Inglés).

13.5.1.15 Parámetro: país

Explicación: Solo aplica si WDPA ha sido señalado como TRUE (✓ en modo on server). Seleccione el país para el cual quiere realizar el análisis de complementariedad usando áreas de la WDPA. Verifique que para el país que selecciona hay un archivo “nombre del país.RData” en la carpeta “wdpa” del conjunto de carpetas que componen la estructura de carpetas y archivos CAPFITOGEN3 (modo local) que previamente se han descargado desde <https://drive.google.com/drive/folders/1xCnllZgzW0uDeClDvcxbADv9H583xzpn?usp=sharing> mientras que en modo on server sólo tiene que seleccionar el país o región de la lista desplegable.

13.5.1.16 Parámetro: propio

Explicación: Solo aplica si areas ha sido señalado como TRUE (✓ en modo on server) y WDPA como FALSE (☐ en modo on server). Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea realizar el análisis de complementariedad para áreas/polígonos usando un shapefile de polígonos propio. Previamente el usuario deberá copiar y pegar el shapefile en la carpeta “wdpa” del conjunto de carpetas y archivos que componen CAPFITOGEN.

Recuerde que un shapefile debe estar conformado como mínimo por tres archivos con el mismo nombre y con las extensiones .shp, .dbf y .shx. Adicionalmente el shapefile debe estar en el sistema de coordenadas geográficas latitud-longitud (datum WGS84).

13.5.1.17 Parámetro: nombre

Explicación: Solo aplica si propio ha sido señalado como TRUE (✓ en modo on server). Escriba el nombre exacto que tiene el shapefile propio que contiene los polígonos sobre los que desea realizar el análisis de complementariedad. NO añada ni puntos ni el nombre de la extensión. Por ejemplo, si el shapefile (conformado por varios archivos) tiene nombres como “areasprotegidas.shp”, “areasprotegidas.dbf” o “areasprotegidas.shx”, lo único que debe escribir en este campo es “areasprotegidas”.

13.5.1.18 Parámetro: campo

Explicación: Solo aplica si propio ha sido señalado como TRUE (✓ en modo on server). Escriba el nombre exacto que tiene el campo de identificación de cada polígono que contiene el archivo shapefile propio. Si tiene alguna duda sobre el nombre, abra el shapefile en el programa DIVA-GIS e inspeccione la tabla que acompaña al mapa de polígonos.

13.5.1.19 Parámetro: nareas

Explicación: Solo aplica si areas ha sido señalado como TRUE (✓ en modo on server). El análisis de complementariedad para áreas (ya sean de la WDPA o propias) determina un ranking de áreas de alto valor de cobertura (que incluyen dentro de sus límites el mayor número de taxa diferentes y complementarias entre sí). Además Complementa puede producir algunas estadísticas adicionales para un subconjunto de áreas del ranking. Determine el número de áreas

que serán incluidas en el análisis adicional. La selección inicia desde el área de más alto valor de cobertura y continúa hacia abajo en el ranking.

13.5.1.20 Parámetro: **coveran**

Explicación: Solo aplica si areas ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea obtener un análisis de cobertura (además del análisis de complementariedad) para las áreas WDPA, propias o ambas (según se haya seleccionado los parámetros “WDPA” o “propio”).

13.5.1.21 Parámetro: **niveltax**

Explicación: Indique el nivel taxonómico (opciones: “genus”, “species” o “subtaxa”) para el cual desea realizar el análisis de complementariedad. Esta selección afectará el análisis ya sea sobre celdas de una grilla o sobre áreas WDPA o propias. Tenga en cuenta que la información taxonómica proviene de los campos “GENUS”, “SPECIES” o “SUBTAXA” de la tabla de pasaporte o presencias.

13.5.1.22 Parámetro: **datanatax**

Explicación: Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea incluir en el análisis las presencias con valor NA para el campo taxonómico objetivo (“GENUS”, “SPECIES” o “SUBTAXA” de acuerdo al parámetro “niveltax”).

13.5.1.23 Parámetro: **mapaelcf**

Explicación: Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si la tabla de presencias contiene información en el campo “BG_ELC” y la desea combinar con la información taxonómica para realizar un análisis de complementariedad de tipo taxonómico-ecogeográfico.

13.5.1.24 Parámetro: **mapaelc**

Explicación: Solo aplica si mapaelcf ha sido señalado como TRUE (✓ en modo on server). Indique el nombre del archivo (modo local) o suba el archivo al servidor (modo on server) que contiene el mapa ELC producto de la aplicación de la herramienta ELCmapas, el cual debe encontrarse en la carpeta CAPFITOGEN3/ELCmapas (modo local). Este mapa, en formato DIVA-GIS (extensión .grd, como lo produce ELCmapas) debe escribirse con la extensión. Así, si el nombre del mapa es “mapa_elc_spain”, ud. debe escribir (modo local) o buscar el archivo (modo on server) “mapa_elc_spain.grd”.

13.5.1.25 Parámetro: **datanaelc**

Explicación: Solo aplica si mapaelcf ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea incluir en el análisis de complementariedad presencias con valor NA en el campo “BG_ELC”. Seleccionando esta opción NA en “BG_ELC” será considerado como otra categoría del mapa ELC.

13.5.1.26 Parámetro: data0elc

Explicación: Solo aplica si mapaelcf ha sido señalado como TRUE (✓ en modo on server). Seleccione esta opción (TRUE en modo local, ✓ en modo on server) si desea incluir en el análisis de complementariedad presencias con valor 0 en el campo “BG_ELC”. Esto quiere decir, que la categoría ecogeográfica 0 del mapa ELC se considerará como un escenario adaptativo.

13.5.1.27 Parámetro: resultados (sólo para modo local)

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

13.6. Resultados de Complementa

Después de definir todos los parámetros y rutas (para modo local) que requiere Complementa, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, Complementa guardará los resultados en la ruta y carpeta que se haya determinado en “resultados” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server.

Independientemente del modo usado (local o on server), según los análisis solicitados, se producirán resultados en carpetas para un análisis de complementariedad sobre áreas de usuario (usando un “shapefile” suministrado por el usuario), sobre áreas WDPA o sobre celdas, llamadas “AnálisisAreasProt_ProtectedAreasAnalysis”, “AnálisisWDPA_WDPAA-nalysis”, y “AnálisisCeldas_CellAnalysis” respectivamente. Si adicionalmente el usuario solicitó un análisis de cobertura dentro de su análisis de complementariedad por áreas, encontrará una carpeta llamada “CoverageAnalysis” dentro de las carpetas resultado de los análisis por áreas. A continuación, se describen los resultados que se guardan en cada carpeta.

13.6.1 Carpeta “AnálisisAreasProt ProtectedAreasAnalysis”

Tabla ‘Tabla_Estadisticas_Stats_Table_Complementa.txt/.xls’: Esta tabla contiene algunos interesantes datos y estadísticas descriptivas del resultado del análisis de complementariedad como el número total de áreas complementarias, el número de áreas consideradas (dato que el usuario suministra en el parámetro “nareas”), porcentaje de taxa cubiertos por estas áreas complementarias consideradas, etc.

Tabla ‘Tabla_final_Final_Table_Complementa.txt/.xls’: Esta área presenta en forma de ranking (de mayor a menor) las áreas con mayor riqueza y complementarias entre sí. El campo “arealID” corresponde al código de cada área del campo de identificación presente en el “shapefile”. Si se ha usado áreas WDPA para el análisis de complementariedad, esta columna tendrá el título “wdpaid” y el código de identificación corresponde con el de la base de datos mundial

de áreas protegidas. En cualquier de los dos casos, frente al código de identificación se muestra el número de taxa diferentes y complementarios que contiene dicha área.

Tabla ‘Tabla_Pob_en_fuera_AreasProt_Table_Pop_in_out_ProtAreas.txt/.xls’: Esta tabla muestra la lista total de sitios de ocurrencia analizados, indicando cuáles de ellos se encuentran dentro o fuera de la red de áreas protegidas utilizada. Para su identificación, la tabla contiene el código ACCENUMB e información taxonómica y de categoría ecogeográfica (sólo en el caso que el usuario la haya suministrado en el formato “pasaporte”). Para conocer si un sitio de ocurrencia se encuentra dentro o fuera de un área protegida, en la parte derecha de esta tabla se encuentran los campos con información de identificación de áreas protegidas; si estos campos aparecen con información NA, esto indica que el sitio está fuera de la red de áreas. Por el contrario, si aparece información de identificación y demás características de alguna área protegida, esto indicaría que el sitio de ocurrencia está situado dentro de los límites del área protegida allí descrita.

Tabla ‘Tabla_Pob_en_Top_Areas_Table_Pop_in_Top_Areas.txt/.xls’: Esta tabla corresponde a una lista de sitios de ocurrencia que se encuentran dentro de áreas protegidas complementarias seleccionadas, es decir, aquellas áreas de la parte alta de la tabla ‘Tabla_final_Final_Table_Complementa.txt’ y cuyo número corresponde al que el usuario suministra en el parámetro “nareas”.

Tabla ‘Tabla_Table_Base.txt/.xls’: Esta tabla corresponde al insumo básico para la obtención del análisis de complementariedad y muestra con un valor de “1” si un taxón está presente en un área protegida (identificada por su código) o “0” si está ausente.

Mapa ‘ComplementAreas.shp’: Este mapa tipo “shapefile” de polígonos muestra las áreas protegidas complementarias de la tabla ‘Tabla_final_Final_Table_Complementa.txt’ el cual puede ser visualizado en DIVA-GIS. La tabla acompañante de este “shapefile” contiene dos campos de alta importancia: el campo de identificación de las áreas, proveniente del “shapefile” suministrado por el usuario (con el mismo nombre de campo que el “shapefile” original) y el campo N_Taxa que indica el número de taxa complementarios que contiene cada área. Esta información coincide con el contenido de la tabla ‘Tabla_final_Final_Table_Complementa.txt’.

Mapa ‘final_analyzed_points.shp’: Este mapa muestra todos los sitios de presencia de los taxa analizados en forma de “shapefile” de puntos, el cual se puede visualizar en DIVA-GIS. La tabla asociada a este “shapefile” contiene campos clave para la identificación de los sitios de ocurrencia (campo ACCENUMB proveniente de la tabla de ‘pasaporte’ de entrada), información taxonómica, coordenadas, unidad taxonómica o taxonómica-ecogeográfica de referencia para el análisis de complementariedad (campo CHAIN) e información sobre al área protegida en la que estaría presente la población.

13.6.2 Carpeta “AnálisisWDPA WDPAAnalysis”

Tabla ‘Tabla_Estadisticas_Stats_Table_Complementa.txt/.xls’: corresponde a la tabla ‘Tabla_Estadisticas_Stats_Table_Complementa.txt’ del apartado anterior.

Tabla ‘Tabla_final_Final_Table_Complementa.txt/.xls’: corresponde a la tabla ‘Tabla_final_Final_Table_Complementa.txt’ del apartado anterior.

Tabla ‘Tabla_Pob_en_fuera_WDPA_Table_Pop_in_out_WDPA.txt/.xls’: Contiene información equivalente a la tabla ‘Tabla_Pob_en_fuera_AreasProt_Table_Pop_in_out_ProtAreas.txt’ del apartado anterior.

Tabla ‘Tabla_Pob_en_Top_WDPA_Table_Pop_in_Top_WDPA.txt/.xls’: corresponde a la tabla ‘Tabla_Pob_en_Top_Areas_Table_Pop_in_Top_Areas.txt’ del apartado anterior.

Tabla ‘Tabla_Table_Base.txt’: corresponde a la tabla ‘Tabla_Table_Base.txt/.xls’ del apartado anterior.

Mapa ‘ComplementaryWDPAmapping.shp’: corresponde al mapa ‘ComplementAreas.shp’ descrito en el apartado anterior.

Mapa ‘final_analyzed_points.shp’: corresponde al mapa ‘final_analyzed_points.shp’ descrito en el apartado anterior.

Mapa ‘TotalWDPAmapping.shp’: Este es un mapa tipo “shapefile” de polígonos que muestra la totalidad de áreas protegidas terrestres y marinas (considerando de éstas últimas sólo su parte terrestre si lo hubiera) de la base de datos WDPA para el país o región objetivo y que fue utilizado en el análisis de complementariedad.

13.6.3 Carpeta “AnálisisCeldas CellAnalysis”

Tabla ‘Datos_por_Celda_Data_by_CELL.txt/.xls’: Esta tabla es una lista de todas las celdas que contienen sitios de ocurrencia de los taxa (o combinación taxón-categoría ecogeográfica) analizados, indicando frente al campo de identificación de celda, el número de taxa (o combinación taxón-categoría ecogeográfica) ocurriendo dentro de la celda.

Tabla ‘Datos_por_Taxa_ELC_Data_by_Taxa_ELC.txt/.xls’: Esta tabla es una lista de todos los taxa (o combinación taxón-categoría ecogeográfica) que ocurren dentro de alguna celda, indicando frente al campo de identificación del taxa (o combinación taxón-categoría ecogeográfica) el número de celdas diferentes en las que ocurre.

Tabla ‘Tabla_Estadisticas_Stats_Table_Complementa.txt/.xls’: Esta tabla es equivalente a la tabla ‘Tabla_Estadisticas_Stats_Table_Complementa.txt’ de los dos apartados anteriores, sólo que su referencia son celdas de la grilla y no áreas protegidas.

Tabla ‘Tabla_final_Final_Table_Complementa.txt/.xls’: Esta tabla es equivalente a la tabla ‘Tabla_final_Final_Table_Complementa.txt’ de los dos apartados anteriores, sólo que su referencia son las celdas de la grilla y no áreas protegidas.

Tabla ‘Tabla_Pob_en_Top_Celdas_Table_Pop_in_Top_Cells.txt/.xls’: Esta tabla es equivalente a la tabla ‘Tabla_Pob_en_Top_WDPA_Table_Pop_in_Top_WDPA.txt’ o ‘Tabla_Pob_en_Top_Areas_Table_Pop_in_Top_Areas.txt/.xls’

de los dos apartados anteriores, sólo que su referencia son celdas de la grilla y no áreas protegidas.

Tabla ‘Tabla_Table_Base.txt’: Esta tabla es equivalente a la tabla ‘Tabla_Table_Base.txt/.xls’ de los dos apartados anteriores, solo que su referencia son las celdas de la grilla y no áreas protegidas.

Mapa ‘Complementa_map.grd/.tif’: Mapa tipo ráster, el cual se puede visualizar en DIVA-GIS (.grd) o en otros software (.tif), y que muestra el resultado final del análisis de complementariedad por celdas contenido en la tabla ‘Tabla_final_Final_Table_Complementa.txt/.xls’.

Mapa ‘final_analyzed_points.shp’: corresponde al mapa ‘final_analyzed_points.shp’ descrito en los dos apartados anteriores.

13.6.4 Carpeta “CoverageAnalysis”

Esta carpeta aparecerá dentro de las carpetas “AnálisisAreasProt_ProtectedAreasAnalysis” o “AnálisisWDPA_WDPAA-analysis” en caso de que el usuario haya solicitado este tipo de análisis. Su contenido en términos de tablas y mapas es el mismo, independientemente del origen del mapa de la red de áreas protegidas analizada. El único aspecto que cambia es el nombre de dos archivos, uno de mapa y otro de tabla, donde para el caso de áreas protegidas suministradas por el usuario, los nombres de los archivos resultado contendrán los términos AREAS o AreasProt, y para el caso de áreas de la base de datos mundial de áreas protegidas, contendrán los términos WDPA o AreasWDPA respectivamente. En la presente descripción se usará los nombres de archivos del caso de áreas suministradas por usuario.

Tabla ‘Tabla_AreasProt_Cubriendo_ProtAreas_Covering_Table.txt/.xls’: Esta tabla muestra la lista total de áreas protegidas con su código de identificación y en los campos “N_pops” y “N_Diff_Taxa” indica el número de poblaciones (o sitios de ocurrencia de la tabla pasaporte) que ocurren en cada área y el número de taxa (o combinación taxón-categoría ecogeográfica) diferentes respectivamente.

Tabla ‘Tabla_Estadísticas_Cobertura_Stats_Table_Coverage.txt/.xls’: Estadísticas simples sobre el cubrimiento de la red de áreas protegidas sobre los taxa de interés, como número de poblaciones (Pops_Covered) o taxa (Taxa_Covered) cubiertas en total y en porcentaje (Perc_Pops_Covered y Perc_Taxa_Covered, respectivamente) o el número de áreas cubriendo poblaciones y su porcentaje (N_Areas_Cov_Pops y Perc_Areas_Cov).

Tabla ‘Tabla_Taxa_Cubiertos_Table_Taxa_Covered.txt/.xls’: Corresponde a una lista similar a la de la tabla ‘Tabla_AreasProt_Cubriendo_ProtAreas_Covering_Table.txt’ pero centrándose en este caso en los taxa (o combinación taxón-categoría ecogeográfica). Para ello contiene un primer campo de identificación de los taxa o de las combinaciones taxa-ELC (Taxa_TaxaELC), campos que indican el número total de poblaciones (Total_N_Pops) o poblaciones cubiertas por algún área protegida (Pops_in_AREAS) y su porcentaje respecto al total de poblaciones incluidas en la tabla pasaporte (Perc_Pops_in).

Mapa ‘CoverAREASmap.shp’: Corresponde a un mapa tipo “shapefile” de polígonos que se puede visualizar en DI-

VA-GIS, que contiene en su tabla acompañante toda la información disponible de las áreas protegidas y en su parte derecha dos campos clave, “N_pops” que indica el número de poblaciones cubiertas por el área en cuestión y “N_Diff_Tax”, indicando el número de taxa o combinaciones taxón-categoría ecogeográfica diferentes ocurriendo en cada área.

13.7. Referencias

FAO. 1997. Plan de Acción Mundial para la Conservación y Utilización Sostenible de los Recursos Fitogenéticos para la Alimentación y la Agricultura y la Declaración de Leipzig. Rome, Italy. 64p.

FAO. 2012. Segundo Plan de Acción Mundial para los Recursos Fitogenéticos para la Alimentación y la Agricultura. Rome, Italy. 104p.

García, R. M., Parra-Quijano, M., Iriondo, J. M. 2017. A multispecies collecting strategy for crop wild relatives based on complementary areas with a high density of ecogeographical gaps. *Crop Science*, 57(3), 1059-1069.

Justus, J., Sarkar, S. 2002. The principle of complementarity in the design of reserve networks to conserve biodiversity: a preliminary history. *Journal of Biosciences*, 27(4): 421-435.

Kati, V., Devillers, P., Dufrêne, M., Legakis, A., Vokou, D., Lebrun, P. 2004. Hotspots, complementarity or representativeness? Designing optimal small-scale reserves for biodiversity conservation. *Biological Conservation*, 120(4): 471-480.

Mponya, N. K., Chanyenga, T., Brehm, J. M., Maxted, N. 2020. *In situ* and *ex situ* conservation gap analyses of crop wild relatives from Malawi. *Genetic Resources and Crop Evolution*, 68: 759-771.

Rebelo, A. G., Siegfried, W. R. 1990. Protection of fynbos vegetation: ideal and real-world options. *Biological Conservation*, 54(1): 15-31.

Parra-Quijano, M., Draper, D., Iriondo, J. 2003. Assessing *in situ* conservation of *Lupinus* spp. In Spain through GIS. *Crop Wild Relative*, 1: 8-9.

Phillips, J., Asdal, Å., Magos Brehm, J., Rasmussen, M., Maxted, N. 2016. *In situ* and *ex situ* diversity analysis of priority crop wild relatives in Norway. *Diversity and Distributions*, 22(11), 1112-1126.

Rubio Teso, M. L., Iriondo, J. M. 2019. *In situ* Conservation Assessment of Forage and Fodder CWR in Spain Using Phytosociological Associations. *Sustainability*, 11(21), 5882.

ESTACIÓN EXPERIMENTAL CENTRAL DE LA AMAZONÍA



Taller Nacional, Joya de los Sachas, Ecuador, Diciembre 2015.



14 | Herramienta Bfuture

14.1. Información (capas) bioclimática proyectada a futuro

La información sobre cómo puede ser el clima (precipitaciones y temperaturas) en un escenario de cambio climático futuro puede ser muy importante en el momento de analizar las perspectivas de adaptación de una especie, y fijar si es necesario, algunas medidas de protección.

Los modelos de distribución de especies utilizan capas SIG de información ambiental, incluyendo información climática, a manera de variables independientes o predictores, para proyectar la distribución potencial de una especie. Por tanto, cualquier cambio en las variables ambientales, derivará en un cambio de la proyección de la distribución de la especie. Si tenemos disponibles las variables climáticas de futuro, en forma de capas SIG, producidas a su vez por modelos que proyectan principalmente los cambios climáticos debidos a la acumulación de gases de invernadero, podremos predecir cómo será la distribución potencial futura de especies de nuestro interés.

Afortunadamente del trabajo de obtención de estas capas SIG se han encargado varios equipos de investigadores, algunos de ellos haciendo disponibles de forma gratuita esta información para que otros investigadores la puedan utilizar en estudios del impacto del cambio climático. Es el caso del proyecto worldclim (<http://www.worldclim.org>) liderado por Robert Hijmans, que en base a capas climáticas de presente, ha producido capas proyectadas de futuro usando modelos climáticos globales (o modelos generales de circulación), bajo diferentes escenarios de emisión de gases invernadero y para diferentes periodos de tiempo.

Los modelos climáticos globales (GCM en inglés) representan los procesos que ocurrirán en el océano, atmósfera, criosfera y superficie terrestre para simular la respuesta climática al incremento de los gases de invernadero, de acuerdo al Panel Intergubernamental sobre Cambio Climático (Intergovernmental Panel on Climate Change – IPCC, https://www.ipcc-data.org/guidelines/pages/gcm_guide.html). Los modelos climáticos globales disponibles en el portal worldclim se corresponden con aquellos incluidos en el proyecto CIMP5 (quinta fase del Coupled Model Inter-comparison Project), se listan en la Tabla 2.

Con la entrada de worldclim 2.1 (versión actual y disponible desde 2017), se han proyectado de nuevo GCM para la versión CIMP6. Los nuevos GCM CIMP6 disponibles desde el portal worldclim son BCC-CSM2-MR, CNRM-CM6-1, CNRM-ESM2-1, CanESM5, GFDL-ESM4, IPSL-CM6A-LR, MIROC-ES2L, MIROC6 y MRI-ESM2-0.

Los escenarios de emisiones representativas (Representative Concentration Pathways – RCPs en inglés) son cuatro trayectorias de emisión de gases de invernadero adoptadas por el IPCC en 2014, definidas por la trayectoria de su forzamiento radiativo total (medidas acumuladas de emisiones humanas de gases de invernadero de todas las fuentes expresadas en Watts por metro cuadrado) nivelado para el año 2100. Fueron seleccionadas estos cuatro escenarios ya que representan un rango amplio de posibles respuestas climáticas basadas en revisiones bibliográficas. El RCP2.6 por ejemplo, asume un forzamiento radiativo de 2.6 W/m², sería el escenario más optimista. RCP4.5 y RCP6.0 serían escenarios intermedios y RCP8.5 corresponde al escenario más pesimista. Así se proyecta que el incremento de temperatura promedio global para el periodo 2081-2100 relativa a 1986-2005 estaría en los siguientes rangos (determinados por las modelizaciones CIMP5) 0.3 a 1.7°C para RCP2.6, 1.1 a 2.6°C para RCP4.5, 1.4 a 3.1°C para RCP6.0 y 2.6 a 4.8°C para RCP8.5 (IPCC, 2013).

Tabla 2. Modelos climáticos globales (modelos generales de circulación - GCM) disponibles en worldclim (<http://www.worldclim.org>) para CIMP5 en forma de capas SIG.

Model (code)	Institution or Modeling centre	Terms of use
ACCESS1-0	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia) y BOM (Bureau of Meteorology, Australia)	Unrestricted
BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration	Unrestricted
CCSM4	Community Climate System Model, version 4	Unrestricted
CESM1-CAM5-1-FV2	National Science Foundation, Department of Energy, National Center for Atmospheric Research	Unrestricted
CNRM-CM5	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	Unrestricted
GFDL-CM3	Geophysical Fluid Dynamics Laboratory	Unrestricted
GFDL-ESM2G		
GISS-E2-R	NASA Goddard Institute for Space Studies	Unrestricted
HadGEM2-AO	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)	Unrestricted
HadGEM2-CC		
HadGEM2-ES		
INMCM4	Institute for Numerical Mathematics	Unrestricted
IPSL-CM5A-LR	Institut Pierre-Simon Laplace	Unrestricted
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	Non-commercial only
MIROC-ESM		
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Non-commercial only
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	Unrestricted
MRI-CGCM3	Meteorological Research Institute	Non-commercial only
NorESM1-M	Norwegian Climate Centre	Unrestricted

Fuente: Página web del Coupled Model Intercomparison Project.

Worldclim también ajusta los RCP para su nueva versión 2.1 proyectada sobre GCM CIMP6, introduciendo el concepto Shared Socioeconomic Pathways (SSP), dado que los RCP cumplieron 15 de años de uso y los escenarios de cambio climático han cambiado desde entonces. Cada uno de los nueve SSP que se han configurado, representa una proyección socioeconómica y un entorno político diferentes para el futuro. Sin embargo, se han priorizado cinco SSP en el (Sixth Assessment Report - AR6) del Panel Intergubernamental del Cambio Climático (IPCC), cuatro de los cuales están proyectados en worldclim a saber: SSP1-2.6 (SSP126) de la ruta socioeconómica 1 de “sostenibilidad” o de bajos desafíos de mitigación y adaptación, donde se asume para 2100 2.6 W m^2 en fuerza radiativa (radiative forcing), SSP2-4.5 (SSP245) de la ruta socioeconómica “en medio del camino” o desafíos intermedios de mitigación y adaptación, que para 2100 asume una fuerza radiativa de 4.5 W m^2 , SSP3-7.0 (SSP370) de la ruta socioeconómica “fragmentación / rivalidad regional” o altos desafíos de mitigación y adaptación, que para 2100 asume una fuerza radiativa de 7 W m^2 , y finalmente SSP5-8.5 (SSP585) de la ruta socioeconómica “desarrollo convencional o alimentado por combustibles fósiles” o altos desafíos de mitigación y baja adaptación, asumiendo para 2100 una fuerza radiativa de 8.5 W m^2 (Kebede *et al.*, 2018; Meinshausen *et al.*, 2020).

Las proyecciones también varían de acuerdo al periodo de tiempo que se desea estimar. El incremento de temperaturas no será el mismo si consideramos el año 2050 que el año 2100 asumiendo incrementos constantes de emisiones de gases invernadero. En el caso de worldclim, las proyecciones para las proyecciones en la versión 2.1 disponibles son 2030 (2021-2040), 2050 (2041-2060), 2070 (2061-2080) y 2090 (2081-2100).

14.2. Cómo funciona Bfuture

La herramienta Bfuture usado en modo local, puede trabajar de dos maneras para hacer disponible al usuario información climática proyectada a futuro en forma de capas SIG, listas para ser usadas por herramientas como Modela.

La primera es que el usuario haga la descarga del portal worldclim por su cuenta y el archivo .zip descargado lo sitúe en la carpeta ‘rdatamapsf’, luego la herramienta se encargará de descomprimir, transformar el formato original worldclim al formato aceptado por CAPFITOGEN y opcionalmente recortarlo a medida de la distribución de los sitios de presencia / sitios de recolección o usando los contornos de un país o región. En esta opción, el usuario requerirá conexión a internet en el momento en que descarga la información desde el portal de worldclim, no cuando ejecuta la herramienta Bfuture.

Bajo la segunda opción, Bfuture es capaz de hacer la descarga de la información climática proyectada a futuro desde worldclim directamente, siempre y cuando el usuario tenga conexión a internet durante el proceso de utilización de la herramienta. Luego, como en el caso anterior, Bfuture adaptará el contenido a los requerimientos de CAPFITOGEN y si así lo solicita el usuario, hará un recorte de las capas descargadas. Tanto por descarga directa cómo descargando por Bfuture, la cobertura de las capas descargadas es global.

En modo on server Bfuture sólo trabaja mediante la transferencia de archivos desde worldclim hasta el servidor de CAPFITOGEN3, por tanto si el usuario considera usar este modo, no debe descargar previamente las capas desde el portal worldclim.

En cualquier caso, antes de utilizar la herramienta Bfuture, se le recomienda al usuario tener decidido que modelo global de clima (GCM), escenario SSP y periodo de tiempo desea utilizar. Para ello puede consultar la página web de worldclim (<http://www.worldclim.org>) y abundante literatura científica sobre estudios del impacto de cambio climático en la biodiversidad donde se han utilizado este tipo de capas de información. También debe definir el tipo de información o capas que desea utilizar. Las capas están organizadas en cuatro grupos: 'bioclimatic_indices' que corresponde a las 19 variables 'bioclimáticas' ('bioclim' según aparecen en el portal worldclim), 'monthly_tot_prec' o las 12 variables de precipitación total mensual, 'monthly_min_temp' o las 12 variables de temperatura mínima mensual y 'monthly_max_tem' para las 12 variables de temperatura máxima mensual. Bfuture le permite seleccionar varios o todos los conjuntos de variables climáticas a la vez. Es importante anotar que worldclim no ofrece capas mensuales de temperatura promedio proyectadas a futuro y Bfuture no las crea, por lo que estas variables no estarán disponibles ni en Bfuture ni en herramientas como Modela, en el caso que se solicite hacer una modelización con datos de futuro.

También el usuario debe determinar previamente el tamaño de celda (resolución) que desea para las capas que descargará. Worldclim ofrece cuatro tamaños de celda: ~1x1, 5x5, 10x10 y 20x20 km (30 arc-segundos, 2.5, 5 y 10 arc-minutos respectivamente). Estas resoluciones de descarga están disponibles para Worldclim versión 1.0 y CIMP5, mientras que para la versión 2.1, sólo están disponibles las resoluciones 5x5, 10x10 y 20x20 km (2.5, 5 y 10 arc-minutos respectivamente) para enero de 2021. Se espera que en un futuro cercano la resolución más alta (1x1 km o 30 arc-segundos) esté disponible para Worldclim 2.1 y CIMP6.

Una vez definido los cinco parámetros (GCM, SSP, periodo temporal y conjunto(s) de variables y resolución), tiene los elementos necesarios para usar Bfuture tanto en modo local como modo on server.

Tenga en cuenta que el proceso de descarga o transferencia de capas de información, sea a través de Bfuture o descargando directamente de worldclim, puede requerir mucho tiempo (de acuerdo a la velocidad de la conexión de internet) y capacidad de almacenaje en el disco duro de destino (modo local), especialmente cuando las resoluciones son altas (1x1 km por ejemplo). Un conjunto de capas bioclim de cualquier GCM ocupa alrededor de 3.4 Gb, luego si desea descárgalo rápidamente y almacenarlo sin problemas de espacio, requerirá alta velocidad y suficiente espacio en el disco duro (modo local) o en la zona de archivos de usuario y resultados en modo on server. Además, tenga en cuenta que este archivo de 3.4 Gb luego debe ser descomprimido. El contenido descomprimido ocupará un espacio similar (3.5 Gb), luego se requiere cerca de 7 Gb disponibles en el disco duro por un instante en el proceso. Cuando es Bfuture quien realiza la descarga desde worldclim, sea a través del modo local o modo on server, el archivo .zip descargado será eliminado al final del proceso.

Cuando el usuario opte por recortar las capas globales usando el contorno de un país (parámetro croplayer=TRUE y paiscrop=TRUE), es requisito tener disponible en la carpeta 'rdatamaps', una carpeta con el nombre del país y dentro de ella otra carpeta con la resolución deseada y dentro las capas ecogeográficas para el país que se usará como molde. Por ejemplo, si el usuario quiere recortar la información descargada de futuro con cobertura global y resolución 5x5 km para Portugal, debe asegurarse que tiene capas (archivos .tif) en la ruta CAPFITOGEN/rdatamaps/portugal/5x5/.

El recorte por la distribución de datos de presencia / accesiones es muy similar al proceso que realiza la herramienta rLayer.

14.3. Utilización de la herramienta Bfuture

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Bfuture, se deben especificar una serie de parámetros por parte del usuario.

14.3.1 Parámetros iniciales definidos por el usuario

14.3.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

14.3.1.2 Parámetro: local (sólo para modo local)

Explicación: Seleccione esta opción (TRUE) si la información bioclimática proyectada a futuro con cubrimiento global ya fue descargada y el archivo .zip se encuentra disponible en la carpeta 'rdatamapsf' del conjunto de carpetas y archivos dentro de la carpeta CAPFITOGEN2. Si no marca esta casilla, la información bioclimática de futuro será descargada desde worldclim automáticamente, por lo cual requerirá conexión a internet.

14.3.1.3 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución, pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, sub-continentes o continentes. Si ya descargo el archivo de datos bioclimáticos de futuro desde el portal de worldclim (sólo modo local), note que el valor resol1 corresponde al término que aparece después de wc2.1_, al inicio del nombre del archivo. Por ejemplo, si su archivo es wc2.1_10m_tmin_BCC-CSM2-MR_ssp126_2021-2040.zip, el la resolución sería 10m, es decir, 10 minutos o resol1="celdas 20x20 km aprox (10 arc-min)". Para que Bfuture pueda extraer y ajustar correctamente la información del archivo .zip, el valor resol1 aquí seleccionado debe coincidir con el valor indicado resolución del archivo .zip (10m= "celdas 20x20 km aprox (10 arc-min)").

14.3.1.4 Parámetro: ssp

Explicación: Seleccione el escenario de ruta socioeconómica compartida (SSP en inglés) para el cual Ud. desea descargar o transferir información (modo local o modo on server) o para la que ya descargó en archivo .zip desde el portal worldclim (sólo para modo local). Hay cuatro SSP disponibles para CIMP6 en worldclim: 126 (ssp1, 2.6), 245 (ssp2, 4.5), 370 (ssp3, 7.0) y 585 (ssp5, 8.5). Si ya descargo el archivo de datos bioclimáticos de futuro desde el portal de worldclim, note que el valor ssp corresponde al número que aparece después de las letras ssp. Por ejemplo, si su archivo es wc2.1_10m_tmin_BCC-CSM2-MR_ssp126_2021-2040.zip, el ssp sería 126. Para que Bfuture pueda extraer

y ajustar correctamente la información del archivo .zip, el valor ssp aquí seleccionado debe coincidir con el valor ssp del archivo .zip.

14.3.1.4 Parámetro: gcm

Explicación: Seleccione el modelo de clima global (GCM en inglés, también conocidos como modelos generales de circulación) para el cual Ud. desea descargar información o ya ha descargado un archivo .zip desde el portal de worldclim. Si ya descargo el archivo de datos bioclimáticos de futuro desde el portal de worldclim (sólo para modo local), note que el valor gcm corresponde al término que va después del tipo de capas solicitado en el extenso nombre del archivo.zip, por ejemplo, si su archivo es wc2.1_10m_tmin_BCC-CSM2-MR_ssp126_2021-2040.zip, el nombre del gcm viene después de las capas solicitadas (parámetro varset, en este caso temperaturas mínimas o tmin_), es decir, "BCC-CSM2-MR". Para que Bfuture pueda encontrar correctamente la información del archivo descargado, la opción gcm aquí seleccionada debe coincidir con el valor gcm del archivo .zip.

14.3.1.5 Parámetro: proy

Explicación: Seleccione el periodo de tiempo para el cual desea descargar o ha descargado información bioclimática de futuro proyectada. La opción 50 representa el valor/año 2050 (promedio entre 2014 y 2060) y la opción 70 representa el valor/año 2070 (promedio entre 2061 y 2080). Si ya descargo el archivo de datos bioclimáticos de futuro desde el portal de worldclim (sólo para modo local), note que el valor 'proy' corresponde al intervalo que aparece en la parte final del nombre del archivo.zip descargado desde worldclim. Usanod el mismo ejemplo anterior, si su archivo es wc2.1_10m_tmin_BCC-CSM2-MR_ssp126_2021-2040.zip, el valor proy sería 2030 porque el intervalo del archivo es "2021-2040". Para que Bfuture pueda encontrar correctamente la información del archivo descargado, la opción proy aquí seleccionada debe coincidir con el valor de intervalo de tiempo del archivo .zip.

14.3.1.6 Parámetro: varset

Explicación: Seleccione el tipo de variables bioclimáticas proyectadas a futuro que desea transferir (modo local o on server) o que ya ha descargado en forma de archivo .zip desde el portal de worldclim. Si ya descargo el archivo de datos bioclimáticos de futuro desde el portal de worldclim (sólo para modo local), note que la opción 'varset' corresponde al término siguiente que define la resolución. Usando el ejemplo anterior, si su archivo es wc2.1_10m_tmin_BCC-CSM2-MR_ssp126_2021-2040.zip, el valor varset sería 'tmin' que corresponde a temperatura mínima ("monthly_min_temp"). Para que Bfuture pueda extraer y ajustar correctamente la información del archivo descargado, la opción varset aquí seleccionada debe coincidir con el código varset del archivo .zip.

14.3.1.7 Parámetro: croplayer

Explicación: Seleccione esta opción (TRUE en modo local o en modo on server) si desea recortar la información descargada (capas) ya sea por el contorno de un país o por la extensión de sus datos de distribución de especie(s). La opción FALSE (modo local) o dejar sin marcar esta opción (en modo on server) indicará a la herramienta que la información descargada mantendrá su cubrimiento original (global). Si la resolución es alta ('resol1' con valores de

“Celdas 1x1 km aprox (30 arc-seg)” o “Celdas 5x5 km aprox (2.5 arc-min)”), requerirá de bastante espacio en su disco duro (modo local) o en su área de usuario (modo on server) para conservar capas de cubrimiento global.

14.3.1.8 Parámetro: paiscrop

Explicación: Solo aplica si croplayer ha sido señalado como TRUE (✓ en modo on server). Marque esta opción (TRUE en modo local o ✓ en modo on server) si desea recortar las capas de información descargadas por el contorno de un país o región. Dejar sin marcar esta opción, indica que usted recortará las capas de información usando sus datos de presencia o distribución (en el formato de pasaporte CAPFITOGEN).

14.3.1.9 Parámetro: pais

Explicación: Solo aplica si paiscrop ha sido señalado como TRUE (✓ en modo on server). Seleccione el país/región por el cual desea recortar las capas descargadas. Nota: La resolución seleccionada en el parámetro “resol1” debe estar disponible para el país o región aquí seleccionada según la tabla que encuentra en este link <http://www.capfitogen.net/es/herramientas/cobertura/>. Si no hay concordancia entre “pais” y “resol1” de acuerdo a la tabla de disponibilidad, un error se producirá. Además Ud. deberá tener este país/región y resolución disponible con datos de presente en la carpeta ‘rdatamaps’ (sólo aplica para modo local).

14.3.1.10 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. La extensión geográfica de las ocurrencias incluidas en esta tabla será usada para recortar las capas SIG descargadas.

14.3.1.11 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro ‘pasaporte’) la tabla resultado de GEOQUAL denominada “PasaporteOriginalEvaluadoGEOQUAL.txt” con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

14.3.1.12 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

14.3.1.13 Parámetro: uname

Explicación: Solo aplica si croplayer ha sido señalado como TRUE (✓ en modo on server) y paiscrop como FALSE (□ en modo on server). Seleccione uno de los siguientes nombres para el nuevo conjunto de capas recortadas por los datos de distribución (usando la tabla indicada en 'pasaporte').

14.4. Resultados de Bfuture

Una vez finalizado el proceso de transferencia de información (si no se descargó directamente desde worldclim), descompresión, adaptación al formato CAPFITOGEN y (opcionalmente) recorte de capas, el resultado son las capas de futuro listas para ser usadas, dispuestas en la carpeta 'rdatamapsf' (modo local) o en la zona de archivos de usuario y resultados (modo on server). Dentro de 'rdatamapsf' se guarda el archivo 'Parametros.Parameters.Bfuture.txt' que contiene la lista de parámetros utilizados y aparecerá otra carpeta cuyo nombre varía de acuerdo a la opción de recorte, así:

- a) Sin recorte: aparecerá la carpeta 'world'
- b) Recortando por distribución de datos de presencia / accesiones: una carpeta llamada 'user1', 'user2' o 'user3' (modo local) o el nombre que haya elegido (modo on server) de acuerdo a lo seleccionado en el parámetro 'uname'.
- c) Recortando por la extensión de un país o región: una carpeta con el nombre del país o región

Dentro de esta carpeta 'world', 'userX' o 'pais/región' encontrará otra carpeta con un nombre compuesto por: a) Nombre del GCM, b) código del SSP y c) periodo de tiempo de la proyección. Por ejemplo

Cada elemento estará separado entre sí por un guion bajo. Dentro de esta carpeta encontrará otra cuyo nombre indica la resolución (parámetro 'resol1'), por tanto, los posibles nombres de carpetas son '1x1', '5x5', '10x10' y '20x20'. Dentro de esta carpeta encontrará las capas climáticas de futuro (.tif) ya adaptadas y opcionalmente cortadas. Por ejemplo, para Ecuador, solicitando a Bfuture el modelo MIROC6, un SSP5 8.5, el periodo 2090, un conjunto de variables "bioclim" y una resolución 5x5 km, la ruta que lleva a la primera nueva capa climática proyectada a futuro ('bio 1' o temperatura media anual) sería: "X:\CAPFITOGEN3\rdatamapsf\ecuador\MIROC6_585_2090\5x5\bio_1.tif"

14.5. Referencias

Kebede, A. S., Nicholls, R. J., Allan, A., Arto, I., Cazcarro, I., Fernandes, J. A., Hill, C.T., Hutton, C.W., Kay, S., Lázár, A.N., Macadam, I., Palmer, M., Suckall, N., Tompkins, E.L., Vincent, K., Whitehead, P. W. (2018). Applying the global RCP-SSP-SPA scenario framework at sub-national scale: A multi-scale and participatory scenario approach. *Science of the Total Environment*, 635, 659-672.

IPCC, 2013. Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. In: Stocker, T.F., Qin, D., G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Meinshausen, M., Nicholls, Z. R., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., van den Berg, M., Velders, G.J.M., Vollmer, M.K., Wang, R. H. 2020. The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571-3605.



Taller Nacional, Tlalnepantla, México, Mayo 2015.



15 | Herramienta Modela

15.1. Modelos de distribución de especies (SDM)

Un modelo de distribución de especies (SDM en inglés) es una predicción de la ocurrencia o ausencia de una especie dentro de un marco espacial basada en una modelización que busca determinar un patrón entre la ausencia o presencia de la especie y una serie de factores ambientales bajo ciertas hipótesis (Guisan y Zimmermann, 2000). Dado que la predicción se realiza sobre un espacio geográfico, los SIG son los socios apropiados de la técnica SDM para la visualización de resultados.

La capacidad de mostrar distribuciones potenciales ha convertido a los SDM en una herramienta muy útil en la toma de decisiones para la conservación de la agrobiodiversidad (Parra-Quijano *et al.* 2012b), habiéndose utilizado en estudios de:

- Recolecciones optimizadas (Parra-Quijano *et al.*, 2012c).
- Apoyo en la recolección de germoplasma de características especiales (Jarvis *et al.*, 2005)
- Detección de faltantes (gaps) en la conservación *ex situ* (Ramírez *et al.*, 2010).
- Impacto del cambio climático en la distribución de especies silvestres parientes de cultivadas (Jarvis *et al.*, 2008)
- Determinación de sitios apropiados para el establecimiento de reservas genéticas para silvestres parientes de cultivos (Parra-Quijano *et al.*, 2012a)

En los últimos años esta metodología se ha venido usando con frecuencia en la determinación del grado de riesgo al que puede estar sometida una especie vegetal por cuenta del cambio climático. Una explicación extensa y apropiada de los SDM o modelos de nicho ecológico y su aplicación en la conservación de la agrobiodiversidad se puede leer en el “Manual de Capacitación en Análisis Espacial de Diversidad y Distribución de Plantas” (Scheldeman y van Zonneveld, 2011) descargable desde <http://goo.gl/XSiflm>, en las páginas 139 y 140. Vale la pena resaltar tres condiciones que este documento señala para poder aplicar las técnicas de modelación y obtener resultados confiables:

- a) En teoría un SDM sólo debería utilizarse en casos de especies silvestres, donde la adaptación a los sitios donde se encuentran las poblaciones es altamente específica. La domesticación reduce gradualmente la adaptación específica, ampliando la distribución original del ancestro silvestre, llegando a lograr que especies de un continente puedan prosperar en otros. Sin embargo hay que considerar que las formas semicultivadas (que no han sido terminadas de domesticar) o las variedades locales ancestrales, cultivadas por cientos de años en sitios particulares, aun podrían llegar mantener niveles de adaptación específica suficiente para permitir que un SDM produzca predicciones confiables. En cualquier caso ésta última consideración puede estar sometida a debate. Para el caso de especies cultivadas, los investigadores pueden considerar el uso de modelos mecánicos simples como EcoCrop (Ramírez-Villegas *et al.*, 2013).
- b) Para predecir la distribución de una especie mediante un modelo, además de los datos de presencia (o presencia/ausencia), se requieren variables ambientales predictoras de tipo abiótico, con las cuales se establece el patrón (el modelo como tal) y se proyecta sobre ellas espacialmente, ya que se trata de capas SIG. Las variables predictoras usadas en la modelización deben ser determinantes en términos adaptativos y de distribución de la especie.
- c) Los datos de presencia de la especie que alimentarán el modelo no provienen de sitios cuyo ambiente haya sido artificialmente adaptado para que la especie sobreviva. En el caso de una especie cultivada, los datos de

presencia no debería provenir de plantaciones donde se haya alterado de manera significativa las condiciones ambientales originales del sitio para permitir la siembra (regadío permanente, estructuras tipo invernadero, grandes alteraciones químicas o físicas del suelo, etc.). Para especies o formas silvestres, los sitios de presencia no deben corresponder a traslocaciones artificiales que pueden ocurrir en sitios como jardines, arboretos, repoblaciones (con germoplasma no local) o cultivos (no como especie arvense sino como especie principal cultivada). Resulta evidente la importancia que tiene la calidad de la georreferenciación de los sitios de presencia o presencia/ausencia y la confiabilidad en la información contenida en las capas de los predictores para la obtención de un modelo y una predicción de distribución de alta precisión.

15.2. Selección de variables para modelizar – Herramienta SelecVIF

Los resultados de procesos de modelización de distribución de especies dependen de muchos factores, como por ejemplo de la calidad de los datos de presencia o ausencia de poblaciones (variables dependientes), el uso apropiado de algoritmos de modelización o la forma en que se usan las variables ecogeográficas o ambientales (Sillero y Barbosa, 2021). Otro factor muy importante es el conjunto de variables ecogeográficas (variables independientes), y su capacidad de servir como predictoras de la distribución debida a la adaptación abiótica de la especie objetivo. CAPFITOGEN3 ofrece una herramienta que ayuda a sus usuarios a determinar variables con capacidad para identificar variables de alta importancia a la hora de crear grupos ecogeográficos (vía Random Forest). SelecVar también puede filtrar variables que de forma lineal que sólo aportarían redundancia (vía determinación de correlación bivariada), incluyendo así en los procesos posteriores (como creación de mapas ELC o modelización de distribución de especies) sólo variables que aportan información única, al menos desde un punto de vista lineal.

Con la idea de aportar más análisis discriminatorios de variables para procesos de modelización de la distribución de especies, se incluye una herramienta denominada SelecVIF para determinar el valor del Factor de Inflación de Varianza o Variance Inflation Factor – VIF (para mayor detalle, ver https://en.wikipedia.org/wiki/Variance_inflation_factor) de un conjunto de variables ecogeográficas que el usuario desee evaluar, con el fin de detectar problemas de multicolinealidad y así poder filtrar aquellas variables que presenten valores más altos de VIF. Este proceso se aplica en remoción de variables para técnicas de regresión (Akinwande *et al.*, 2015) y muy comúnmente en procesos de modelización de la distribución de especies (Guisan *et al.*, 2002).

La herramienta SelecVIF permite realizar una selección automática de un conjunto de variables de baja multicolinealidad, mediante la eliminación paso a paso de aquellas cuyos valores VIF sobrepasen un umbral definido por el usuario, determinándose en cada paso de nuevo los valores VIF de las variables remanentes, proceso que se repite hasta que se cumpla la condición establecido (valores VIF bajo el umbral). Por esto SelecVIF es un complemento de Modela, la cual se aplicaría previamente, permitiendo tener un criterio de inclusión o exclusión de variables. Sin embargo, la selección de variables puede también basarse en la utilización de SelecVar o en ambas (selecVar y SelecVIF secuencialmente).

Como la determinación de VIF requiere generar un modelo, los parámetros de SelecVIF y Modela son bastante similares. En la sección de este capítulo donde se describen los parámetros, se indicarán aquellos que también son requeridos por SelecVIF.

15.3. Datos de sólo presencia o presencia/ausencia

Los modelos establecen patrones de relación entre la presencia y ausencia de una especie y las condiciones ambientales donde ello ocurre. Es importante notar que en general, un modelo requiere información sobre donde ocurre y donde no ocurre la especie. En la práctica y en particular, en los estudios sobre agrobiodiversidad, no se disponen de datos de ausencia, sólo donde las poblaciones/individuos se han observado o muestreado, es decir, datos de sólo presencia.

Este hecho conduce a la necesidad de generar pseudo-ausencias, o sitios donde no se tiene constancia de presencia y se asume la ausencia de la especie. Así, con presencias reales y pseudo-ausencias se obtiene y evalúa un modelo. Si bien existen algoritmos de modelización capaces de funcionar con datos de sólo presencia de la especie, la generación de pseudo-ausencias permite la utilización de cualquier algoritmo.

15.3.1 Métodos de generación de pseudo-ausencias

Desde la introducción de las pseudo-ausencias como sustituto de verdaderas ausencias, se han propuesto múltiples técnicas para su obtención, siempre persiguiendo el objetivo de obtener respuestas con la mayor evaluación posible. Varios autores (Chefaoui y Lobo, 2008; Phillips *et al.*, 2009; Barbet-Massin *et al.*, 2012) han producido estudios comparativos donde se prueba el efecto de diferentes tipos de pseudo-ausencias en la modelización de especies.

La distribución espacial de las pseudo-ausencias es un factor a tener en cuenta, ya que afecta la respuesta de los modelos, en particular cuando se extrapola la predicción, es decir, cuando la proyección sobrepasa la extensión de los puntos de presencia (van der Wal *et al.*, 2009).

La herramienta Modela pone a disposición cuatro formas de obtención de pseudo-ausencias:

- a) Random: Las pseudo-ausencias se seleccionan aleatoriamente de cualquier celda que forme parte del marco de trabajo que no esté ocupada por un sitio de presencia. Hijmans y Elith (2015) describen esta opción como puntos de fondo, o 'background', diferenciándola de los demás tipos de pseudo-ausencias, ya que de acuerdo a estos autores estos puntos de fondo establecen el escenario ambiental del estudio, mientras los puntos de presencia determinan las condiciones donde la especie ocurre con mayor probabilidad. En contraste, los demás tipos de pseudo-ausencias intentan determinar donde ocurrirían las ausencias. Según los mismos autores, esta opción es preferible ya que su concepto requiere menos supuestos y tiene coherencia estadística respecto al solape entre puntos de presencia y de fondo.

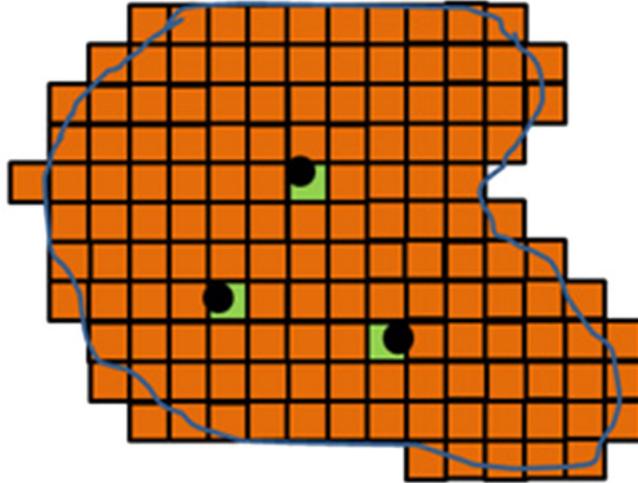


Figura 48. Selección de pseudo ausencias tipo Random o 'background'. En este ejemplo, las pseudo-ausencias serán seleccionadas de las celdas naranjas que cubren el marco de trabajo donde no ocurren datos de presencia (puntos negros).

b) SRE: Las pseudo-ausencias serán extraídas del área no cubierta por la alta probabilidad de ocurrencia dictaminada por un modelo de envoltorio de rango de superficie, en otras palabras, un modelo BIOCLIM. Para determinar la alta probabilidad, esta opción requiere un valor de cuantil que la defina. En cuanto el valor de cuantil se acerca a uno, el envoltorio (área apropiada para encontrar la especie) ocupa más espacio, reduciendo el área de selección de pseudo-ausencias. Esta opción puede conducir a evaluaciones de modelo engañosas (sobre-estimación positiva).

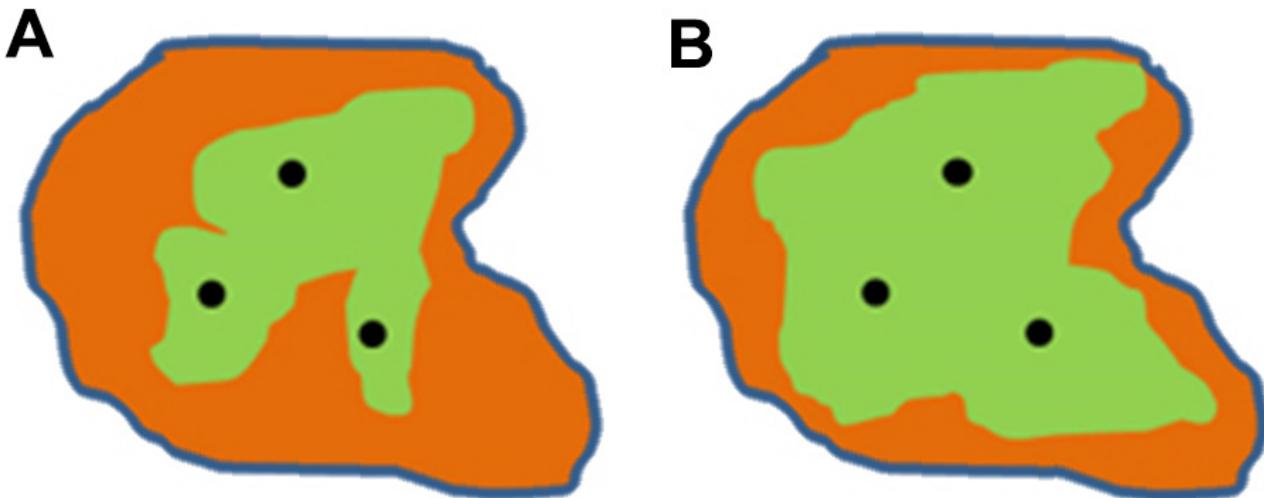


Figura 49. Selección de pseudo ausencias SRE. En este ejemplo, las pseudo-ausencias serán seleccionadas (aleatoriamente) dentro de las áreas naranjas de baja probabilidad de ocurrencia de acuerdo al modelo Bioclim (alta probabilidad en color verde). En el caso A, se habría señalado un cuantil más bajo que en el caso B (cuantil más cercano a 1). El incremento en el valor de cuantil reduce al área seleccionable.

- c) **Disk**: Esta opción crea un disco espacial alrededor de cada punto de presencia, que se puede situar más o menos cerca de dicho punto de acuerdo al criterio del usuario. Para crear el disco, el usuario determina el valor de distancia mínima al punto de presencia, así como también el valor de distancia máxima. Esta variante sería una generalización de algunos métodos de obtención de pseudo-ausencias por distancia a puntos de presencia sugeridos por Barbet-Massin *et al.* (2012).

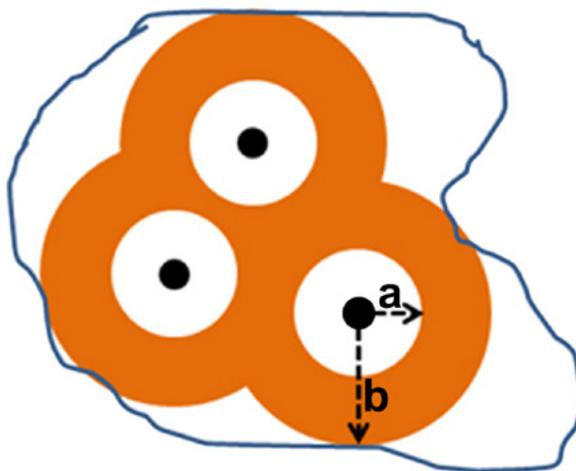


Figura 50. Selección de pseudo ausencias Disk. En este ejemplo, las pseudo-ausencias serán seleccionadas (aleatoriamente) dentro de los discos de color naranja creados mediante la aplicación de una distancia mínima (a) y una distancia máxima (b) desde los puntos de presencia.

- d) **ELC**: Esta opción de obtención de pseudo-ausencias se propone como novedad dentro del ámbito de los SDM por parte de la nueva herramienta Modela. Parte de un mapa ELC previamente obtenido para la especie objetivo (mediante el uso de la herramienta ELCmapas, Capítulo 6). Estos mapas no generan probabilidades, simplemente dividen el territorio en unidades (categorías) de alta similitud ambiental. Luego de la superposición de los puntos de presencia sobre el mapa ELC, es posible determinar la frecuencia de ocurrencia de la especie sobre cada categoría de mapa. Estas frecuencias se pueden agrupar en cuartiles, como lo hace la herramienta Representa, de la siguiente manera: 1. cuartil de alta frecuencia (>0.75), cuartil de media-alta frecuencia (0.5-0.75), cuartil de media baja frecuencia (0.25-0.5) y cuartil de baja frecuencia (<0.25). Adicionalmente se considera el grupo de categorías donde la especie no ocurre (frecuencia nula). En base a esta clasificación de las categorías por frecuencias, la herramienta Modela le permite al usuario seleccionar el (o los) grupo(s) de categorías de donde desea obtener las pseudo-ausencias. Esta aproximación evitaría la sobre-estimación positiva que conlleva el método SRE. Se recomienda que los usuarios de Modela utilicen para la generación pseudo ausencias por esta vía, un mapa ELC de igual extensión y tamaño de celda (resolución) que las capas ecogeográficas con las que se realizará la modelización.



Figura 51. Selección de pseudo ausencias ELC. En este ejemplo, las pseudo-ausencias serán seleccionadas de los grupos de categorías de ELC que indique el usuario. Así las pseudo-ausencias pueden provenir de celdas (seleccionadas aleatoriamente) color verde (grupo 'alta frecuencia'), azul (grupo 'media alta frecuencia'), fucsia (grupo 'media baja frecuencia'), amarillo (grupo 'baja frecuencia') o gris (frecuencia nula). Modela permite elegir varios grupos simultáneamente.

15.4. Algoritmos de modelización

Un algoritmo de modelización es el conjunto de pasos matemáticos necesarios para determinar la existencia de un modelo o patrón entre la variable dependiente (en este caso, la presencia o presencia ausencia de una especie) y las variables independientes (predictores ecogeográficos). Existen muchas alternativas de algoritmos para modelizar la distribución potencial de las especies. Guisan y Zimmermann (2000), Guisan y Thuiller (2005) o Elith y Leathwick (2009), entre otros, ofrecen una explicación detallada de las bases teóricas de la modelización de la distribución de especies, los algoritmos disponibles para modelizar, reglas de elección de algoritmos y los tipos de datos apropiados para cada tipo de algoritmo (presencia/ausencia o sólo presencia, predictores cuantitativos, semicuantitativos, cualitativos).

Recomendamos a los usuarios CAPFITOGEN que deseen usar la herramienta Modela, una revisión previa de los algoritmos disponibles (parámetros 'dismodel' y 'modelos') en la literatura científica. Los algoritmos disponibles en Modela se detallan en la Tabla 3. La misma tabla contiene algunas referencias que pueden consultar los usuarios sobre la aplicación de cada algoritmo en la modelización de la distribución de especies.

También hay abundantes referencias de comparativas de eficiencia entre algoritmos que indican que la elección del algoritmo apropiado depende de varios factores que van desde el tipo de datos de entrada, el marco de trabajo, la resolución de los predictores hasta la naturaleza de la especie en cuestión (ver como ejemplo Guisan y Thuiller, 2005).

Tabla 3. Algoritmos de modelización disponibles en *Modela*, provenientes del uso de los paquetes 'dismo' y 'biomod2' de R

Código algoritmo	Paquete de R	Nombre extendido	Referencia de aplicación en SDM
Domain	dismo	Domain	Carpenter <i>et al.</i> , 1993
Bioclim / SRE	dismo	Bioclim / Surface Range Envelope	Booth <i>et al.</i> , 2014
GLM	biomod2	Generalized Linear Model	Guisan <i>et al.</i> , 2002
GBM	biomod2	Generalized Boosting Model (Boosted Regression Trees)	Elith y Leathwick, 2008
GAM	biomod2	Generalized Additive Model	Guisan <i>et al.</i> , 2002
CTA	biomod2	Classification Tree Analysis	Thuiller <i>et al.</i> , 2003
ANN	biomod2	Artificial Neural Network	Özesmi y Özesmi, 1999
FDA	biomod2	Flexible Discriminant Analysis	Maiorano <i>et al.</i> , 2013
MARS	biomod2	Multiple Adaptive Regression Splines	Mateo <i>et al.</i> , 2010
RF	biomod2	Random Forest	Bradter <i>et al.</i> , 2013
Maxent	dismo y biomod2	Maximum Entropy	Elith <i>et al.</i> , 2011

15.5. Evaluación de modelos

La evaluación de la exactitud de la predicción de los modelos está basada en dos aspectos, la capacidad de discriminación (habilidad de distinguir correctamente entre zonas ocupadas y no ocupadas) y la confiabilidad (concordancia entre la predicción de ocurrencia y la proporción observada de sitios ocupados por la especie) (Pearce y Ferrier, 2000). Existe una serie de índices usados para evaluar ambos aspectos, aplicados a respuestas binarias o continuas transformadas en binarias (binarizadas) mediante el uso de un valor de corte, denominado umbral. Estos índices se denominan 'umbral-dependientes' y requieren una tabla o matriz de confusión (Liu *et al.*, 2009) como la que describe la Tabla 4.

Tabla 4. Tabla de confusión usada para determinar índices umbral-dependientes.

		Predicción	
		Presencia	Ausencia
Observación	Presencia	a (verdadero positivo)	b (falso positivo)
	Ausencia	c (falso negativo)	d (verdadero negativo)

La capacidad predictiva del modelo se puede calcular mediante evaluadores que usan los elementos de la matriz de confusión. La obtención de la matriz sólo es posible mediante la reserva de una parte de las presencias y ausencias o pseudo-ausencias, las cuales no participan en el proceso de modelización y se denomina ‘conjunto de prueba’ o ‘test set’. La parte de las presencias y ausencias o pseudo-ausencias con las que se obtiene el modelo se denomina ‘conjunto de entrenamiento’ o ‘train set’. En la herramienta Modela, los usuarios determinan el porcentaje de datos que se usan como ‘train set’ y ‘test set’ (parámetros ‘datadiv’ o ‘datadiv2’).

Los evaluadores usan algunas relaciones de los elementos de la matriz de confusión como:

- Sensibilidad = $a / (a+c)$
- Tasa de omisión = $c / (a+c)$
- Especificidad = $d / (b+d)$

Como ejemplo de la utilización de estas relaciones por parte de los evaluadores, la característica operativa del receptor o ROC (siglas en inglés) es un evaluador que enfrenta gráficamente la sensibilidad vs 1-especificidad.

Los evaluadores disponibles en Modela son los siguientes:

- ROC (Relative Operating Characteristic), ver Fawcett (2004).
- Kappa de Cohen, ver Cohen (1960) o Wood (2007).
- TSS (True skill statistic) ver Allouche *et al.* (2006)
- FAR (False alarm ratio) aplicado y explicado por Barnes *et al.* (2007), Barnes *et al.* (2009) y Roeber *et al.* (2009) sobre estudios meteorológicos.
- SR (Success ratio) es igual a 1-FAR, ver en Roeber *et al.* (2009)
- ACCURACY (fraction correct) ver explicación en <http://goo.gl/jEL02V>.
- BIAS (Bias score or frequency bias) ver en Roeber *et al.* (2009)
- POD (Probability of detection) ver en Roeber *et al.* (2009)
- CSI (Critical success index) ver en Roeber *et al.* (2009)
- ETS (Equitable threat score) ver explicación en <http://goo.gl/jEL02V>.

Explicaciones de cada evaluador y algunos ejemplos de su función en la verificación de predicciones climáticas se pueden observar en la web <http://goo.gl/jEL02V>. Con el fin de elegir el evaluador correcto, se recomienda también la revisión de publicaciones que profundizan en la utilización de estos en la medida de exactitud de las predicciones de

distribución de especies (Liu *et al.* 2009) y algunos que realizan comparativas e incluso críticas a ciertos evaluadores como los trabajos de Pearce *et al.* (2000), Allouche *et al.* (2006) o Lobo *et al.* (2008).

15.6. Ensamblaje

Dado que la técnica de modelización de la distribución de especies tiene siempre algún grado de incertidumbre (mayor o menor de acuerdo a los resultados de la evaluación), recientemente se ha introducido la idea de obtener modelos “consenso”, producto del ensamblaje de diversos modelos individuales (Thuiller, 2004). Marmion *et al.* (2009) introduce varias técnicas de “ensamblaje” mientras que Thuiller *et al.* (2009) ofrece una explicación de cómo funciona el ensamblaje de las predicciones o proyecciones del proceso de modelización mediante el uso del paquete R ‘biomod’. Dado que Modela utiliza el paquete biomod2 (evolución del paquete biomod) para realizar ensamblaje de modelos, los tipos de ensamblaje disponibles son:

- “PA_dataset+repet”
- “PA_dataset+algo”
- “PA_dataset”
- “algo”
- “all”

donde “PA_dataset” indica que el conjunto de pseudo-ausencias va a ser utilizado como factor de ensamblaje. Hay que tener en cuenta que se pueden crear varios conjuntos de pseudo-ausencias (PA) de igual tamaño, pero constituidos por diferentes localizaciones ya que cualquier método de obtención de pseudo-ausencia realiza selección aleatoria, dentro de una zona particular o todo el marco de trabajo.

Así, “repet” indica que se utilizará la repetición de un modelo como factor de ensamblaje. Modela permite obtener repeticiones de un mismo modelo (algoritmo) el cual puede ser evaluado en base a diferentes criterios. Como se debe realizar una partición de los datos para permitir la evaluación, cada partición genera conjuntos nuevos de presencias y pseudo-ausencias que a su vez generan respuestas diferentes, razón por la cual “repet” puede ser utilizado como factor de ensamblaje.

El término “algo” indica que se utilizará el(los) algoritmo(s) seleccionado(s) para modelizar como factor de ensamblaje.

Y finalmente, “all” indica que “PA_dataset”, “repet” y “algo” serán utilizadas a la vez para ensamblar los modelos. Modela permite al usuario ensamblar todos los modelos producidos o sólo aquellos que hayan superado determinado valor de evaluación.

Para entender cómo funciona y que implica cada opción de ensamblaje, se recomienda revisar el documento “Ensemble Modelling: the different available ways to build ensemble forecasts” (ensamblaje de modelos: las diferentes formas disponibles de ensamblar predicciones) que se puede descargar desde <https://www.capfitogen.net/es/EnsembleModelingAssembly.pdf>.

15.7. Predicciones para condiciones bioclimáticas de futuro

Con la idea de cuantificar el posible impacto del cambio climático en ciertas especies, se ha propuesto proyectar modelos sobre capas SIG que representan futuros escenarios climáticos provenientes de modelos climáticos globales (GCM) como los que se ilustran en la Tabla 2. De esta manera se establece un patrón (a través del modelo) entre los datos de presencia/pseudo-ausencia o presencia/ausencia y las variables ecogeográficas (las bioclimáticas correspondiendo a condiciones de presente), el cual se proyecta en capas de variables bioclimáticas de futuro. En caso de haberse utilizado edáficas y geofísicas (obviamente de presente), para el establecimiento del modelo, las mismas capas serán incluidas en la proyección de futuro. La obtención de estas proyecciones de futuro han permitido, por ejemplo, determinar el impacto del cambio climático en especies silvestres parientes de cultivadas de cacahuete (*Arachis*), papa (*Solanum*) y caupí (*Vigna*), en América Latina y África (Jarvis *et al.*, 2008). Este tipo de estudios requieren de comparaciones entre las proyecciones a presente y futuro, comparación posible mediante el uso de la herramienta Mcompare.

15.8. Subcolecciones FIGS por modelización (calibración)

En el capítulo para la herramienta FIGS_R se describe la metodología para la obtención de conjuntos de germoplasma (subcolecciones) seleccionados por el alto potencial de contener características de interés para los mejoradores. La herramienta Modela permite la obtención de subcolecciones FIGS por la vía denominada “calibración”. Para ello, Modela realiza un proceso de modelización asumiendo datos de presencia/ausencia, donde el conjunto de datos (tabla) de presencia se entiende como las accesiones caracterizadas o evaluados que presentan el rasgo de interés y el conjunto (tabla) de ausencia incluye las accesiones que se conoce no contienen el rasgo. En el momento de la proyección del patrón obtenido en la modelización, ésta se realiza sobre otro conjunto de accesiones, las cuales no han sido evaluadas por el rasgo de interés. Modela también proyecta sobre capas SIG para obtener un mapa de probabilidad de aparición del rasgo.

15.9. Cómo funciona Modela

Para comprender el funcionamiento de Modela y su complejidad, la Figura 52 ilustra las relaciones y procesos conducentes a un mapa de predicción de la distribución de una especie. La Figura 53 muestra el proceso y los elementos involucrados en la obtención de subcolecciones FIGS por el método de calibración.

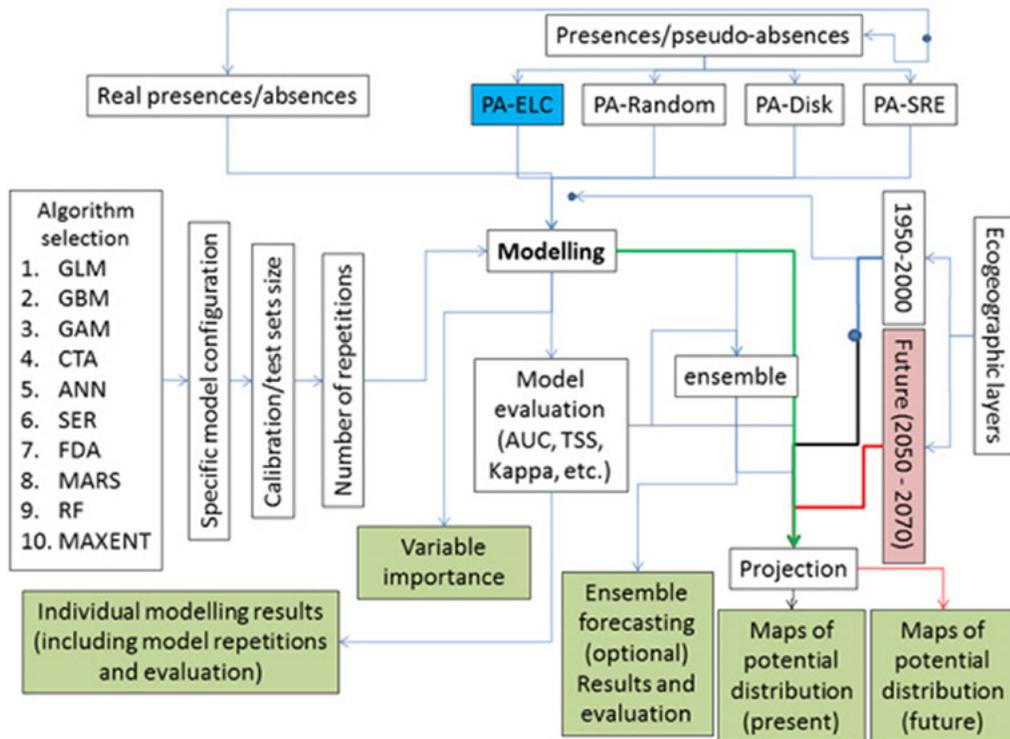


Figura 52. Diagrama que ilustra el proceso de generación de modelos y proyecciones de distribución de especies por parte de la herramienta Modela.

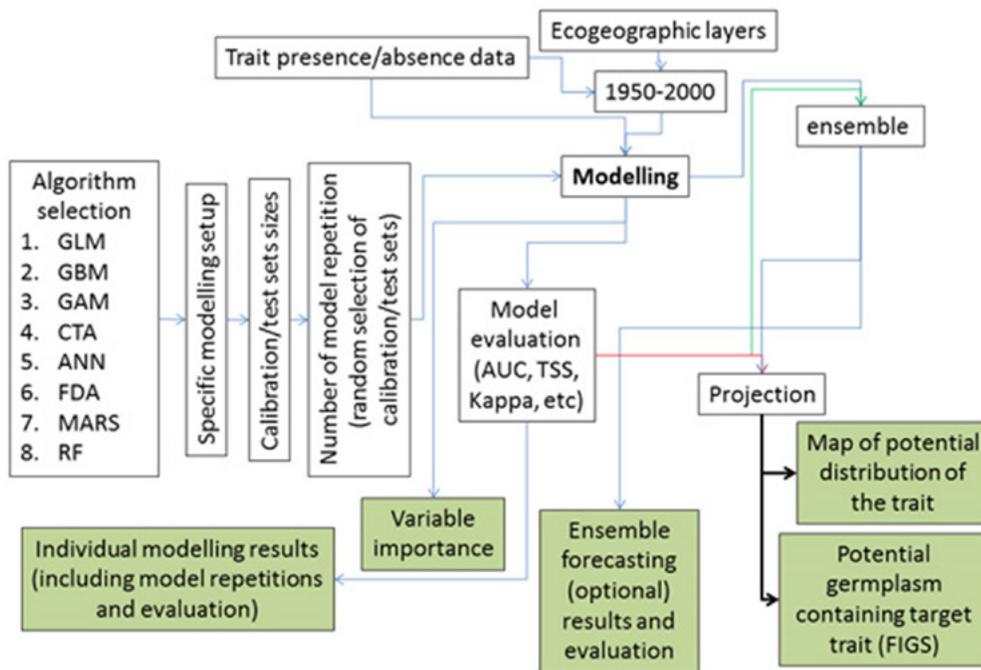


Figura 53. Diagrama que ilustra el proceso de obtención de subcolecciones FIGS por parte de la herramienta Modela.

15.9.1. Utilización de Modela y SelecVIF mediante una tabla de parámetros predefinidos

Es importante señalar una característica diferenciadora de Modela y SelecVIF, y es la forma en cómo se configuran los parámetros. Si el usuario lo desea, puede seleccionar las opciones de cada parámetro con la ayuda de una plantilla facilitada y sin el acoso del tiempo. A través de esta opción, habilitada en ambos modos de utilización (modo local y modo on server) Modela y SelecVIF puede configurarse antes de abrir la interfaz de la herramienta mediante una tabla en Excel (archivo “TableParameters.xlsx” para Modela o “TableParametersSelecVIF.xlsx” para SelecVIF) que se puede descargar desde https://drive.google.com/drive/folders/16Fl9V9STi2iZ4vUHDkO_rhbXLk-7i99e?usp=sharing o se encuentra en la carpeta “ModelOptions” (ruta CAPFITOGEN3/ModelOptions). Estos archivos Excel contienen una hoja de cálculo llamada “ParameterTable” que contiene todos los parámetros que controlan el uso de Modela. En esta tabla, el usuario especifica el valor (columna “Valor/Value”) de cada parámetro de Modela o SelecVIF. Las celdas de la columna “Valor/Value” están configuradas para ofrecer a los usuarios las opciones posibles o válidas para cada parámetro. La Figura 54 muestra la configuración de la tabla “ParameterTable” del archivo “TableParameters.xlsx”. En ella se aprecia como se va rellenando los campos de la columna “Valor (indicate here your option)” por parte del usuario. Los colores que aparecen en la columna “Parameter” indican la relación que hay entre los diferentes parámetros, así campos con mismo color tienen cierto grado de dependencia o subordinación. La columna “Is this parameter required?” va indicando al usuario, a medida que va configurando cada parámetro, si debe o no proporcionar información para los siguientes parámetros (en orden descendiente). Así, por ejemplo, el usuario en el ejemplo de la Fig. 54 determina el valor TRUE en el parámetro “geoqual”, y en ese momento, en la columna C (“Is this parameter required?”), frente al parámetro “totalqual” cambia el mensaje de “totalqual value is not necessary” a “totalqual value must be provided”. La columna “Notes” describe el tipo de dato requerido para cada parámetro y alguna información extra de interés.

Parameter	Value (indicate here your option)	Is this parameter required?	Notes
1 ruta	C:/CAPFITOGEN3	Ok	para indicar la ruta use sólo este slash/ to in
2 pais	Argentina	Ok	Seleccione un país/región / Select one country
3 pasaporte	Pasaporte.txt	Ok	nombre incluye la extensión .txt, el archivo de
4 geoqual	TRUE	Ok	TRUE = Si yes, FALSE = No
5 totalqual		90 totalqual value must be provided	número entre 0 y 100 number from 0 to 100
6 distdup		1 Ok	número (por defecto 0) number (0 by default)
7 temp	present	Ok	Seleccione una opción Select one option
8 gcm	present	gcm (CIMP6) cell must be filled	Seleccione un modelo global climático de futu
9 rcp/ssp	future	rcp/ssp value must be provided	Seleccione un escenario SSP de emisiones Se
10 proy		proy value must be provided	Seleccione un periodo de proyección para dat
11 resol1		resol1 value is required	Seleccione una opción Select one option
12 variabv[1]		Please select one variable for variabv[1]	Seleccione una variable ecogeográfica a ser ut
13 variabv[2]		Please select one variable for variabv[2]	Seleccione una variable ecogeográfica (no sel
14 variabv[3]		Please select one variable for variabv[3]	Seleccione una variable ecogeográfica (no sel
15 variabv[4]		Select additional variable for variabv[4] if necessary	Seleccione una variable ecogeográfica (no sel
16 variabv[5]		Select additional variable for variabv[5] if necessary	Seleccione una variable ecogeográfica (no sel
17 variabv[6]		Select additional variable for variabv[6] if necessary	Seleccione una variable ecogeográfica (no sel
18 variabv[7]		Select additional variable for variabv[7] if necessary	Seleccione una variable ecogeográfica (no sel
19 variabv[8]		Select additional variable for variabv[8] if necessary	Seleccione una variable ecogeográfica (no sel
20 variabv[9]		Select additional variable for variabv[9] if necessary	Seleccione una variable ecogeográfica (no sel
21 variabv[10]		Select additional variable for variabv[10] if necessar	Seleccione una variable ecogeográfica (no sel
22 variabv[11]		Select additional variable for variabv[11] if necessar	Seleccione una variable ecogeográfica (no sel
23 variabv[12]		Select additional variable for variabv[12] if necessar	Seleccione una variable ecogeográfica (no sel
24 variabv[13]		Select additional variable for variabv[13] if necessar	Seleccione una variable ecogeográfica (no sel
25 variabv[13]		Select additional variable for variabv[13] if necessar	Seleccione una variable ecogeográfica (no sel

Figura 54. Tabla contenida en el archivo “TableParameters.xlsx”.

Algunos parámetros de selección múltiple, como “variabv” que se muestra parcialmente en la Figura 54, se repiten varias veces permitiendo al usuario seleccionar varias opciones. En el caso de “variabv” aparecen 20 repeticiones (numeradas como [1], [2], [3], etc.) para poder seleccionar hasta 20 variables ecogeográficas con las cuales modelizar. El hecho de disponer de varias repeticiones no implica que se deba seleccionar opciones para todas ellas. Si por ejemplo sólo se modelizará con tres variables, se seleccionarán variables sólo en las primeras tres repeticiones (variabv[1], variabv[2] y variabv[3]) y las demás se dejarían en blanco. Justamente para entender los códigos de variables ecogeográficas que aparecen para las repeticiones del parámetro “variabv”, se ha dispuesto dentro del archivo “TableParameters.xlsx” una segunda hoja de cálculo llamada “Variables_Codes”. Esta tabla le permite al usuario conectar cada código de “variabv” con su descripción en inglés y español.

Una vez se ha terminado de rellenar todos los parámetros requeridos según el caso y tipo de análisis que se desea realizar, indicando las opciones hasta el último parámetro, “resultados”, para el modo local. En modo on server, la definición del parámetro “ruta” y “resultados” en el formato Excel, no genera ninguna acción, pues los resultados siempre se guardarán en la zona de archivos de usuario y resultados. Finalmente, con el formato en Excel ya terminado se recomienda guardarlo en ese mismo formato (.xlsx) y para que pueda ser leído por la herramienta, la tabla “ParameterTable” se debe salvar en formato texto delimitado por tabulaciones (.txt). En modo local, el archivo de texto con la tabla de parámetros se debe guardar en la carpeta “ModelOptions”. En ambos modos aparecerá un parámetro en común, denominado “partablename”. En modo on server, el usuario debe subir el archivo de la tabla en formato texto delimitado por tabulaciones para “partablename”. En modo local, en el script de parámetros se debe indicar el nombre del archivo.

15.9.2. Configuración de algoritmos

Los algoritmos que se pueden usar con el paquete biomod2 (parámetro “modelby” opción “biomod”) pueden ser configurados en detalle, para adecuar su funcionamiento a los requerimientos del caso o necesidades del usuario.

Para configurar las opciones/parámetros de un algoritmo, se deben seguir los siguientes pasos:

- Abrir el archivo “ModelOptions_OpcionesModelos.xlsx” que se encuentra en la carpeta “ModelOptions” (ruta CAPFITOGEN2/ModelOptions).
- Este archivo Excel tiene 10 hojas de cálculo, cada una con el nombre de un algoritmo. En la hoja de cálculo de cada algoritmo se encuentra una tabla con las opciones/parámetros que se pueden especificar para el algoritmo en cuestión. La columna “Option” muestra el nombre original de la opción/parámetro configurable y la columna “Value (Default)” muestra el valor por defecto que tiene esa opción/parámetro (valor que sería utilizado por Modela en caso de que en el parámetro “modelopc” se indique FALSE). Modifique las opciones/parámetros de la columna “Value (Default)” según las necesidades o requerimientos. Algunos campos de esta columna presentan menús desplegables con las opciones permitidas para facilitar la modificación (Figura 49). Se recomienda a los usuarios, que antes de modificar cualquier opción/parámetro de algoritmo, se haya estudiado en detalle las consecuencias del cambio, evitando así errores de modelización (por utilizar valores fuera de rango) o resultados inesperados. Todos los posibles cambios disponibles están de acuerdo a la función “BIOMOD.Model.Options” del paquete biomod2. Por esta razón, es altamente aconsejable revisar las explicaciones sobre las

opciones/parámetros de cada algoritmo incluidas en el manual de referencia del paquete biomod2 (descargable desde <https://goo.gl/ogj359>).

	A	B
1	Option	Value (default)
2	distribution	bernoulli
3	n. trees	bernoulli
4	interaction.depth	huberized
5	n.minobsinnode	multinomial
6	shrinkage	adaboost
7	bag.fraction	poisson
8	train.fraction	coxph
9	cv.folds	quantile
10	keep.data	parwise
11	verbose	1
12	perf.method	3
		FALSE
		FALSE
		CV

Figura 55. Tabla para la modificación de opciones/parámetros del algoritmo GBM incluida en el archivo Excel “ModelOptions_OpcionesModelos.xlsx”. Se muestra como para la opción “distribution” se despliega un menú con las opciones disponibles.

- Salve individualmente las tablas de opciones/parámetros de cada algoritmo donde haya realizado un cambio y guárdela en formato de texto delimitado por tabulaciones en la carpeta “ModelOptions” (ruta CAPFITOGEN2/ModelOptions). Tenga cuidado de no reemplazar ninguno de los archivos de texto allí presentes, nombrados con los códigos de cada algoritmo más el número 1 (por ejemplo “GLM1.txt”), pues estos serán utilizados en el caso de que se de las siguientes tres condiciones simultáneamente: 1) haber seleccionado el algoritmo en cuestión (en el parámetro “modelos”), 2) haber indicado en el parámetro “modelopc” que utilizará modificaciones de opciones/parámetros para alguno de los algoritmos seleccionados, y 3) para alguno de los algoritmos no realizará cambios (justo para los algoritmos ‘sin cambios’ son necesarios los archivos nombrados con los códigos de cada algoritmo más el número 1).
- Indique si para al menos uno de los algoritmos que desea usar para modelizar va a utilizar opciones/parámetros específicos indicando el valor TRUE (o haciendo “click” en el cuadro de selección en la interfaz) para el parámetro “modelopc”.
- Indique para cuales algoritmos en particular desea introducir cambios en sus opciones/parámetros, indicando el valor TRUE (o haciendo “click” en el cuadro de selección en la interfaz) en los parámetros de Modela “GLM”, “GBM”, “GAM”, “CTA”, “ANN”, “SRE”, “FDA”, “MARS”, “RF” o “MAXENT” según sea el caso.

- Indique el nombre del archivo de texto que contiene la tabla de opciones/parámetros modificada en los parámetros “GLMopt”, “GBMopt”, “GAMopt”, “CTAopt”, “ANNopt”, “SREopt”, “FDAopt”, “MARSopt”, “RFopt” o “MAXENTopt”.
- Proceda a rellenar el resto de los campos según los criterios de su análisis.

15.9.3. Proyección de modelos sobre capas bioclimáticas de futuro

Gracias a que la herramienta Bfuture hace disponible información bioclimática proyectada a futuro en forma de capas SIG para su uso en otras herramientas, Modela es capaz de realizar modelización con datos ecogeográficos de presente y proyectar los modelos en capas bioclimáticas de futuro (junto con capas edáficas y/o geofísicas de presente, si así fuera el caso). Para hacer que Modela proyecte sobre capas bioclimáticas de futuro, sólo debe indicar el valor “future” en el parámetro “temp”. Como consecuencia se debe indicar el modelo climático global (parámetro “gcm”), la vía de concentración representativa (parámetro “rcp”) y el periodo de proyección de las capas de futuro (parámetro “proy”) de las capas proyectadas a futuro que se quieren utilizar. Tenga en cuenta que, para proyectar sobre capas de futuro, se requiere:

- Contar con las mismas capas para presente y futuro para el país/región o marco de trabajo personalizado (“user1”, “user2” o “user3”) y resolución (tamaño de celda). El usuario puede verificar si dispone de las mismas capas para presente y futuro. Por ejemplo, si el país objetivo es “Bolivia” y la resolución con la que se desea trabajar es 5x5 km, y para las capas futuro ha descargado y adaptado (mediante el uso de Bfuture) capas para el modelo climático global “GFDL-ESM2G”, rcp4.5 y proyectada a 2050, el usuario debe disponer de capas (archivos con el nombre de la variable y extensión .gnd y .gri) en la siguiente ruta para condiciones de presente: CAPFITOGEN2/rdatamaps/bolivia/5x5/ y en la siguiente ruta para condiciones de futuro: CAPFITOGEN2/rdatamapsf/Bolivia/GFDL-ESM2G_45_50/5x5/
- No seleccionar en el parámetro “variabv” ninguna variable de temperatura promedio mensual (Temp prom1, Temp prom 2, ... Temp prom 12).
- Finalmente Modela generará predicciones de presencia para los dos escenarios bioclimáticos (presente y futuro) no sólo para futuro cuando se indique el valor “future” en el parámetro “temp”.

15.9.4. Mantenimiento de la proporción presencias/ausencias o presencias/pseudo ausencias en los “train” y “test” sets

La evaluación de los modelos obtenidos por Modela es opcional cuando se utiliza el paquete dismo para modelizar (parámetro “modelby” opción “dismo”) pero cuando se modeliza por biomod2 (parámetro “modelby” opción “biomod”), la evaluación se realiza por defecto. En cualquier caso, si se evalúa el o los modelos, se debe determinar la proporción de presencias/ausencias o presencias/pseudo ausencias destinadas a conformar el “train set” (parámetro “datadiv2” para modelización por dismo, parámetro “datadiv” para modelización por biomod). Cuando se utiliza el paquete biomod2 para modelizar, la función “BIOMOD_Modeling” asigna aleatoriamente las presencias y ausencias (o presencias y pseudo ausencias) que van a componer el “test” y el “train” set, sin reparar en la proporción final de presencias/ausencias (o presencias/pseudo ausencias) en cada conjunto. Modela, mediante el parámetro “modprop”

le permite al usuario elegir si desea la selección totalmente al azar que hace biomod2 de presencias y ausencias (o presencias y pseudo ausencias) o si prefiere mantener la proporción de presencias y ausencias (o presencias y pseudo ausencias) original de los datos aportados por el usuario en el “test” y el “train” set. Utilizando la opción de mantener la proporcionalidad (parámetro “modprop” opción TRUE), Modela seleccionará aleatoriamente las presencias y ausencias (o pseudo ausencias) necesarias para guardar la proporcionalidad original en el “test” y el “train” set.

15.9.5. Pasos y requerimientos en Modela

A continuación, se describen los pasos y requerimientos para los diferentes tipos de análisis que se pueden llevar a cabo con Modela.

15.9.5.1 Modelización usando datos de presencia-ausencia

Para este análisis se requiere la siguiente entrada de datos:

- a)** Tabla de datos de presencia en formato pasaporte CAPFITOGEN (parámetro “pasaporte”). Opcionalmente estos datos de presencia podrían haberse evaluado por la calidad de su georreferenciación (usando previamente la herramienta GEOQUAL).
- b)** Tabla de datos de ausencia en formato pasaporte CAPFITOGEN (parámetro “pasaporteb”).

Y los pasos a seguir para este tipo de análisis son:

- c)** Indicar las opciones para los parámetros “ruta”, “país”, “distdup” y “resol1”. La resolución de las capas ecogeográficas (parámetro “resol1”) será la resolución de los mapas de predicción de los modelos.
- d)** Si se va a proyectar sobre capas bioclimáticas de futuro (parámetro “temp” opción “future”), especificar los parámetros “gcm”, “rcp” y “proy”.
- e)** Seleccionar las variables/capas ecogeográficas con las que se va a modelizar (parámetros “variabv”, “latitud” y “longitud”).
- f)** Seleccionar para el parámetro “ausencia” la opción “absence”.
- g)** Seleccionar para el parámetro “modelby” la opción “biomod” (ya que la opción “dismo” ofrece sólo algoritmos para datos de “sólo presencia”). También se debe indicar un identificador (una palabra preferiblemente corta) para el proceso de modelización por “biomod” (parámetro “modelid”).
- h)** Seleccionar el o los algoritmos a utilizar (parámetro “modelos”) y si es el caso, utilizar opciones/parámetros específicos para los algoritmos (parámetro “modelopc”) contenidos en archivos de texto individuales. Para el algoritmo que se desee usar opciones/parámetros específicos, se debe seleccionar TRUE para el parámetro con el código del algoritmo y luego introducir el nombre de la tabla en formato de texto que contiene las especificaciones para dicho algoritmo. Por ejemplo, si se desea utilizar opciones específicas para un algoritmo GLM, se debe usar la opción TRUE en el parámetro “GLM” y posteriormente escribir el nombre de la tabla de texto en el parámetro “GLMopt”.
- i)** Determine el número de repeticiones de evaluación (parámetro “modrep”), la proporción en la partición de los datos de presencia/ausencia para el “test” y el “train” set (parámetro “datadiv”), la prevalencia de las ausencias sobre las presencias o viceversa (parámetro “preval”, para más detalles ver el manual de referencia de bio-

mod2) y el número de permutaciones si está interesado en estimar la importancia de los predictores. Posteriormente indique cuales evaluadores desea utilizar (parámetro “testparam”), si desea re-escalar la respuesta de los diferentes algoritmos a una escala 0-1000 (parámetro “reescal”), si desea para cada algoritmo obtener una respuesta usando la totalidad de los datos, es decir, sin hacer partición en “test” y “train” set (parámetro “modcomp”) y si desea proporcionalidad de presencias/ausencias en los “test” y “train” set (parámetro “modprop”).

- j) A continuación debe señalar si desea proyectar (i.e. obtener un mapa de predicción) para todos los modelos obtenidos o sólo para aquellos que superen un valor umbral para determinados evaluadores (parámetro “proyosome”). Si selecciona proyectar sólo los modelos que superen ese umbral, indique cuál de los evaluadores seleccionados en el parámetro “testparam” será el evaluador primario (parámetro “proytest1”) y su umbral (parámetro “proytest1u”). Opcionalmente puede utilizar un parámetro secundario (parámetro “proyosome2” opción TRUE) para realizar un segundo filtrado de modelos indicando cuál evaluador usar (parámetro “proytest2”) y su respectivo umbral (parámetro “proytest2u”).
- k) Adicionalmente el usuario puede solicitar la obtención de un mapa de predicción binarizada (parámetro “binar”) de uno de los modelos, utilizando uno de los evaluadores seleccionados en el parámetro “testparam” (indicándolo en el parámetro “binarmet”) y usando como punto de corte el valor optimizado obtenido para dicho evaluador. Dado que los puntos de corte optimizados para todos los evaluadores utilizados quedan a disposición del usuario en una de las tablas resultado, es posible binarizar el resto de los mapas de predicción en DIVA-GIS, utilizando la opción “Reclass” que se encuentra en la pestaña “Grid”.
- l) El usuario puede solicitar la obtención de un mapa “máscara”, en el cual se señalan las áreas del marco de trabajo donde la predicción es incierta dado que los valores de las capas de las variables ecogeográficas (predictores) están fuera del rango usado para calibrar los modelos (parámetro “maskout”).
- m) Si desea ensamblar todos o algunos de los modelos producidos, marcar la opción TRUE en el parámetro “ensamb”. Si desea considerar todos los modelos obtenidos para su posible ensamblaje seleccione la opción “all” en el parámetro “mod2bens”, de lo contrario seleccionando “best” se considerarán sólo los modelos seleccionados para ser proyectados (aquellos que superaron valores umbrales para determinados evaluadores, parámetros proyosome, proytest1, proytest1u, proyosome2, proytest2, proytest2u). Debe especificar la forma en la que serán ensamblados los modelos (ver apartado 15.5) considerando las repeticiones, los conjuntos de pseudo ausencias y algoritmos empleados (parámetro “tipensam”). Si seleccionó “best” para el parámetro “mod2bens”, se debe definir cuáles modelos serán finalmente ensamblados, y para ello indicar si se utilizará cualquiera de los evaluadores seleccionados en el parámetro “testparam” (parámetro “ensamet” opción FALSE) o los evaluadores y umbrales utilizados para seleccionar los modelos a proyectar (parámetro “ensamet” opción TRUE). El usuario puede decidir si desea que el modelo ensamblado sea evaluado (parámetro “testparam2” opción TRUE) en cuyo caso debe especificar cuales evaluadores se utilizarán (parámetro “testparam3”, múltiples opciones). A continuación, debe señalar con alguno de los siguientes parámetros, el método para unificar los modelos a ensamblar: “probmean”, “probcv”, “probci” (si elige este método, debe indicar también el parámetro “probalfa”), “probmedian”, “probca” o “probmw” (si elige este método, debe indicar también el parámetro “probmwd”).
- n) Indique la ruta donde quiere que se guarden los resultados (parámetro “resultados”).

15.9.5.2 Modelización usando datos de presencia-pseudo ausencia

Para este análisis se requiere la siguiente entrada de datos:

- a) Tabla de datos de presencia en formato pasaporte CAPFITOGEN (parámetro “pasaporte”). Opcionalmente estos datos de presencia podrían haberse evaluado por la calidad de su georreferenciación (usando previamente la herramienta GEOQUAL).

Y los pasos a seguir para este tipo de análisis son:

- a) Indicar las opciones para los parámetros “ruta”, “país”, “distdup” y “resol1”. La resolución de las capas ecogeográficas (parámetro “resol1”) será la resolución de los mapas de predicción de los modelos.
- b) Si se va a proyectar sobre capas bioclimáticas de futuro (parámetro “temp” opción “future”), especificar los parámetros “gcm”, “rcp” y “proy”.
- c) Seleccionar las variables/capas ecogeográficas con las que se va a modelizar (parámetros “variabv”, “latitud” y “longitud”).
- d) Seleccionar para el parámetro “ausencia” la opción “pseudo absence”.
- e) Indicar el número de conjuntos de pseudo ausencias que desea obtener (parámetro “pareps”) teniendo en cuenta que con cada una de ellas se realizará un proceso de modelización. Seleccionar también el tamaño de cada conjunto de pseudo ausencias (parámetro “pansel”) y la estrategia de generación de las pseudo ausencias (parámetro “pastrat”).
- f) Si eligió la estrategia “sre” en el parámetro “pastrat”, debe indicar el valor de cuantil en el parámetro “pasreq”. Si eligió la estrategia “disk”, debe indicar valores de distancia (en metros) de los parámetros “padiskmin” y “padiskmax”. Si eligió la estrategia “elc”, debe indicar el o los cuantiles de frecuencia para el parámetro “paelc” (ver apartado 15.2.1), el nombre del mapa ELC (parámetro “mapaelc”) y del archivo (.txt) que contiene la tabla de estadísticas del mapa ELC (parámetro “statelc”).
- g) Puede seleccionar para el parámetro “modelby” la opción “biomod” o “dismo”. Si selecciona “dismo” seleccione el algoritmo para modelizar vía “dismo” en el parámetro “dismodel” (permite seleccionar un solo algoritmo). Posteriormente indique si desea obtener un modelo con todos los datos de presencia/pseudo ausencia disponibles (parámetro “modcompl”) y si desea realizar una modelización con evaluación (parámetro “diseval” opción TRUE). Si se eligió realizar evaluación, se debe indicar el porcentaje de datos de presencia/pseudo ausencia que se usarán para conformar el “train” set (parámetro “datadiv2”). Finalmente indique la ruta donde guardar los resultados (parámetro “resultados”). Si selecciona la opción “biomod”, los pasos que debe seguir son idénticos a los descritos en el apartado 15.8.5.1 desde la sección h) hasta la n).

15.9.5.3 FIGS usando datos de presencia-ausencia de rasgo

Para este análisis se requiere la siguiente entrada de datos:

- a) Tabla con las accesiones que presentan el rasgo de interés en formato pasaporte CAPFITOGEN (parámetro “pasaporte”). Opcionalmente estos datos podrían haberse evaluado por la calidad de su georreferenciación (usando previamente la herramienta GEOQUAL).
- b) Tabla con las accesiones que NO presentan el rasgo de interés en formato pasaporte CAPFITOGEN (parámetro “pasaporteb”).
- c) Tabla con las accesiones no caracterizadas o evaluadas por el rasgo de interés en formato pasaporte CAPFITOGEN (parámetro “pasaportec”).

Los pasos a seguir para este tipo de análisis son los mismos que para el análisis “Modelización usando datos de presencia-ausencia” (apartado 15.8.5.1), con un requerimiento adicional: seleccionar la opción TRUE para el parámetro “figs”.

15.10. El caso del algoritmo Maxent

Tenga en cuenta que Maxent sólo se utilizará en casos de datos de sólo presencia (presencia/pseudo ausencia). **El ajuste para que el software de Maxent pueda ser utilizado se realiza a través del paquete “biomod” únicamente a partir de CAPFITOGEN3.** Esto implica que para utilizar Maxent, R se debe ‘poner en contacto’ con el software Java, proporcionarle toda la información y luego los resultados traerlos de nuevo a R, donde la programación de la herramienta hará que se guarden en la carpeta indicada en el parámetro “resultados” junto con el resto de las proyecciones de los algoritmos utilizados.

La opción de realizar modelos con el paquete dismo desde Modela queda deshabilitada desde CAPFITOGEN3. Si bien algunos parámetros de Modela, remanentes de versiones anteriores, aun aparece la opción “dismo”, esta no debe seleccionarse pues generaría un error de ejecución.

15.10.1. Requerimientos para que funcione Maxent

Dado que el paquete biomod2 puede realizar modelos Maxent si se ejecutan externamente desde su programa original que funciona con Java. Para instalar el software java requerido, seguir las siguientes instrucciones:

1. Instalar la siguiente versión de java development kit (jdk) desde: https://drive.google.com/file/d/0B3ZXTUGKr-jX_b01BUkplNzVTQTg/view?usp=sharing haciendo doble click en el archivo jdk-8u60-windows-x64.exe que se genera al descomprimir el archivo .zip descargado. En cada paso de la instalación se debe seguir las opciones “next” o “siguiente” hasta terminar instalación. Este archivo instalará una versión Java para Windows 64 bit.
2. En Windows ir a Configuración y buscar “variables de entorno” y seleccionar la opción “Editar las variables de entorno del sistema”. Esto abrirá una ventana emergente llamada “Propiedades del sistema”, buscar y hacer click en el botón “Variables de entorno” y allí crear (o editar, si ya están creadas) las siguientes variables en la ventana “Variables de usuario para pc”:
 - JAVA_HOME: Para esta variable establecer como su valor la ruta C:\Program Files\Java\jre1.8.0_60 o equivalente que lleve en su PC al programa Java Run Environment (o JRE) que acaba de instalar.
 - JRE_HOME: establecer la misma ruta que para JAVA_HOME
 - path: establecer la ruta a R 64 bit, como puede ser C:\Program Files\R\R-3.6.3\bin\x64
3. En la sección “Variables del sistema” hay una llamada “Path”, hacer click sobre ella y luego en el botón “Editar...”. Aparecerá una ventana llamada “Editar variable del sistema”, y en ella una lista de rutas. Crear una nueva ruta haciendo click en el botón “Nuevo”, pegar aquí la ruta que lleva a la carpeta bin de Java SE Development Kit (o JDK), la cual suele ser C:\ProgramFiles\Java\jdk1.8.0_60\bin (incluyendo la parte de “bin”). Una vez creada esta nueva ruta, esta debe posicionarse como la primera en la lista, localizándola allí con la ayuda del botón “Subir”.

4. Ejecutar la palabra java en el “ejecutador” de Windows (programa “Ejecutar” que está en Iniciar-Todos los programas-Accesorios). Aparecerá una ventana, con una caja de texto precedida de una indicación que dice “Abrir:”. Luego de escribir “java” en la caja y hacer click en “Aceptar”, se abrirá y cerrará rápidamente una ventana de color negro de fondo y no deberá aparecer ventana de error. Si es así, las variables de entorno establecidas en el paso 2 son correctas.

15.11. Utilización de la herramienta Modela

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Modela, se deben especificar una serie de parámetros por parte del usuario. El caso de Modela junto con el de la herramienta SelecVIF, representan una excepción en la forma en cómo se introducen los parámetros en la herramienta. Dado el gran número de parámetros que se deben configurar para estas herramientas, y que para muchos de los parámetros las opciones son limitadas y dependen de la configuración de otros parámetros. Por eso, como se comentó antes, el usuario cuenta con unos archivos Excel, con tablas que facilitan la configuración de los parámetros. A continuación, se listan todos los parámetros de Modela y se indican adicionalmente cuáles de ellos son compartidos con SelecVIF (el nombre del parámetro aparece acompañado por – SelecVIF). Posteriormente se listan algunos pocos parámetros que son exclusivos de SelecVar.

15.11.1 Parámetros iniciales definidos por el usuario para Modela

15.11.1.1 Parámetro: ruta (sólo se utiliza cuando se usa el modo local) - SelecVIF

Explicación: Ruta donde se ha copiado o se encuentra las herramientas CAPFITOGEN3. Evite que ésta ruta contenga espacios (por ejemplo ‘C:/Mis documentos’). Nota: use el signo / (slash) en la ruta antes que el signo \ (backslash). Por ejemplo, debe escribir F:/ o C:/CAPFITOGEN3 o D:/MisHerramientas/CAPFITOGEN3, etc.

15.11.1.2 Parámetro: país - SelecVIF

Explicación: Seleccione el país/región o recorte personalizado (user1, user2 o user3 producidos por las herramientas rLayer y Bfuture) que desea ser usado como área de trabajo. Si las ocurrencias se distribuyen en más de un país, considere seleccionar una región, subcontinente, continente o áreas personalizadas antes que un país en particular.

15.11.1.3 Parámetro: pasaporte - SelecVIF

Explicación: Escriba el nombre del archivo que contiene la tabla de datos de presencia/ocurrencias (en formato pasaporte CAPFITOGEN, texto delimitado por tabulaciones) sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. En caso de que desee

realizar un análisis FIGS, la tabla del parámetro “pasaporte” corresponde a la tabla de accesiones evaluadas que poseen el rasgo de interés.

15.11.1.4 Parámetro: **geoqual** - **SelecVIF**

Explicación: Seleccione esta opción (indicando TRUE) si los datos de presencia/pasaporte han sido analizados con la herramienta GEOQUAL y contienen las cinco columnas adicionales (SUITQUAL, LOCALQUAL, COORQUAL, TOTALQUAL y TOTALQUAL100) producto de dicha herramienta. Usualmente la tabla resultado de GEOQUAL se denomina PasaporteOriginalEvaluadoGEOQUAL.txt y puede ser usada como tabla de pasaporte en el punto anterior.

15.11.1.5 Parámetro: **totalqual** - **SelecVIF**

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

15.11.1.6 Parámetro: **distdup** - **SelecVIF**

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. consideraría que dos sitios de recolección en realidad representan una misma población (duplicado geográfico). El valor cero (valor mínimo y por defecto) excluye del análisis ocurrencias con idénticas coordenadas. En la medida que el valor de distancia aquí estipulado aumente, un mayor número de ocurrencias se considerarían duplicados geográficos.

15.11.1.7 Parámetro: **temp**

Explicación: Seleccione una opción de acuerdo a si la modelización la desea hacer sólo sobre datos bioclimáticos de presente (periodo 1950-2000, seleccione “present”) o sobre presente y futuro (proyectado a 2050 o 2070, seleccione “future”). Si sólo usará variables edáficas o geofísicas para modelizar, seleccione la opción “present”. Tenga en cuenta que si selecciona la opción “future”, debe tener disponible los datos de presente y futuro para el país, región o área personalizada, y los de futuro deben estar en la carpeta ‘rdatamapsf’, que es donde la herramienta Bfuture los produce.

15.11.1.8 Parámetro: **rcp/ssp**

Explicación: Sólo aplica si temp=future. Seleccione la vía de concentración representativa (RCP en inglés) de la información proyectada a futuro que tiene disponible para el país/región (parámetro país) en la carpeta “rdatamapsf”. RCP es en otras palabras, el escenario de concentración de gases de invernadero para sus datos bioclimáticos de futuro. Hay cuatro RCP disponibles de acuerdo con el IPCC: RCP2.6 (aquí señalado como 26), RCP4.5 (45), RCP6 (60) y RCP8.5 (85). Estos valores tienen que ver con los valores radiativos (+2.6, +4.5, +6.0, and +8.5 W/m², respectivamente) para el año 2100 en relación a los valores de la era pre-industrial. Acerca de SSP, hay cuatro disponibles: 126 (ssp1, 2.6), 245 (ssp2, 4.5), 370 (ssp3, 7.0) y 585 (ssp5, 8.5).

15.11.1.9 Parámetro: gcm

Explicación: Sólo aplica si temp=future. Seleccione el modelo de clima global (GCM en inglés, también conocidos como modelos generales de circulación) de la información proyectada a futuro que tiene disponible para el país/región (parámetro país) en la carpeta “rdatamapsf”. Estos modelos simulan el clima dependiendo en parte de la concentración de gases de invernadero (rcp/ssp). Si desea saber más de los GCM, puede visitar <http://goo.gl/4XhU6g> o <http://goo.gl/2VUuRP>.

15.11.1.10 Parámetro: proy

Explicación: Sólo aplica si temp=future. Seleccione el periodo de tiempo de la información proyectada a futuro que tiene disponible para el país/región (parámetro país) en la carpeta “rdatamapsf”. La opción 50 representa el valor/año 2050 (promedio entre 2014 y 2060) y la opción 70 representa el valor/año 2070 (promedio entre 2061 y 2080).

15.11.1.11 Parámetro: resol1 - SelecVIF

Explicación: Seleccione el nivel de resolución que desea utilizar para las variables predictoras para modelizar y los mapas resultado. Note que 1x1 km ofrece mayor resolución pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, subcontinentes, continentes o áreas personalizadas.

15.11.1.12 Parámetro: variabv - SelecVIF

Explicación: Seleccione las variables ecogeográficas a usar como predictores en la modelización. Si desea modelizar para presente y futuro (opción “future” en el parámetro “temp”) y usar variables bioclimáticas, tenga en cuenta que debe tener esas variables disponibles para futuro y para el área de trabajo (país, región o área personalizada) en la carpeta ‘rdatamapsf’. NOTA IMPORTANTE: No seleccione variables de temperatura media mensual (desde “Temp prom 1” hasta “Temp prom 12”) si desea proyectar los modelos a condiciones de futuro (opción “future” en parámetro “temp”), pues éstas no están disponibles desde worldclim y por tanto en la carpeta ‘rdatamapsf’.

15.11.1.13 Parámetro: latitud - SelecVIF

Explicación: Marque esta opción (indicando TRUE) si desea incluir la latitud como predictor en su(s) modelo(s).

15.11.1.14 Parámetro: longitud - SelecVIF

Explicación: Marque esta opción (indicando TRUE) si desea incluir la longitud como predictor en su(s) modelo(s).

15.11.1.15 Parámetro: genero - SelecVIF

Explicación: Escriba el nombre del género de la especie para la que desea modelar su distribución. Este campo es obli-

gatorio pero la palabra aquí introducida no tiene que coincidir exactamente con el género que aparece en el campo GENUS del formato de pasaporte usado en CAPFITOGEN. El término aquí introducido sólo servirá para dar nombre (junto con el parámetro “especie”) a algunas carpetas de resultado.

15.11.1.16 Parámetro: especie - SelecVIF

Explicación: Escriba el nombre del epíteto de la especie para la que desea modelar su distribución. Este campo es obligatorio pero la palabra aquí introducida no tiene que coincidir exactamente con la especie que aparece en el campo SPECIES del formato de pasaporte usado en CAPFITOGEN. El término aquí introducido sólo servirá para dar nombre (junto con el parámetro “genero”) a algunas carpetas de resultado.

15.11.1.17 Parámetro: ausencia - SelecVIF

Explicación: Seleccione el tipo de ausencias que utilizará en la modelización. La opción “absence” indica que Ud. utilizará datos de ausencia reales y los aportará en tablas independientes (parámetro “pasaporteb”). Si lo que desea es realizar una selección de germoplasma por FIGS, también debe seleccionar la opción “absence”. Si no cuenta con datos de presencia real para la modelización, seleccione la opción “pseudo absence”. Nota 1: si en el parámetro “modelby” se selecciona “dismo”, el parámetro “ausencia” será automáticamente “pseudoabsence”. Nota 2: si en “ausencia” se selecciona “absence”, en el parámetro “modelby” se selecciona “biomod” y en el parámetro “modelos” se selecciona “MAXENT” entre otros modelos, este algoritmo no se utilizará para modelar.

15.11.1.18 Parámetro: pasaporteb - SelecVIF

Explicación: Sólo aplica para “ausencia” = “absence”. Escriba el nombre del archivo que contiene la tabla de datos de ausencia (en formato pasaporte CAPFITOGEN, texto delimitado por tabulaciones) sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Si lo que desea es realizar un análisis FIGS, la tabla del parámetro “pasaporteb” corresponde a la tabla de accesiones evaluadas que no poseen el rasgo de interés.

15.11.1.19 Parámetro: figs

Explicación: Sólo aplica para “ausencia” = “absence”. Marque esta opción si desea realizar un análisis FIGS. Para ello aportará una tabla de accesiones evaluadas positivamente (parámetro “pasaporte”), una tabla de accesiones evaluadas negativamente (parámetro “pasaporteb”) y una tabla de accesiones no evaluadas donde se proyectará la modelización (parámetro “pasaportec”).

15.11.1.20 Parámetro: pasaportec

Explicación: Sólo aplica para “figs” marcado. Escriba el nombre del archivo que contiene la tabla de accesiones sin evaluar (en formato pasaporte CAPFITOGEN, texto delimitado por tabulaciones) sin olvidar incluir la extensión (.txt).

Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Esta tabla sólo es válida para análisis FIGS (parámetro "figs") y contiene las accesiones sobre las cuales se proyectarán los modelos.

15.11.1.21 Parámetro: pareps - SelecVIF

Explicación: Sólo aplica para "ausencia" = "pseudo absence". Especifique el número de conjunto de pseudo ausencias que desea generar para modelizar. Tenga en cuenta que si selecciona para el parámetro "modelby" la opción "dismo", sólo el primer conjunto de pseudo ausencias será utilizado (escoja el valor "pareps" = 1). Si selecciona "modelby" = "biomod", el número de procesos de modelización que se realizarán por cada algoritmo (parámetro "modelos") estarán determinados por la multiplicación de "pareps" por el parámetro "modrep". Considere que un número alto resultado de esta multiplicación puede tomar mucho tiempo de procesamiento.

15.11.1.22 Parámetro: pansel - SelecVIF

Explicación: Sólo aplica para "ausencia" = "pseudo absence". Especifique el número de pseudoausencias que desea generar por cada conjunto determinado en "pareps". Tenga en cuenta que el número aquí determinado será el número de pseudo ausencias que junto a las presencias (parámetro "pasaporte") serán usados para modelizar. Es importante considere un balance entre presencias y pseudo ausencias, para más información ver <http://goo.gl/4aMsDZ>.

15.11.1.23 Parámetro: pastrat - SelecVIF

Explicación: Sólo aplica para "ausencia" = "pseudo absence". Seleccione la estrategia de generación de pseudo ausencias que desea utilizar. Note que algunas estrategias podrían condicionar el número de pseudo ausencias (parámetro "pansel") a la baja, porque algunas estrategias reducen bastante el área posible para la selección de pseudo ausencias.

La opción "random" selecciona pseudo ausencias de cualquier celda de 'fondo' (celdas de un mapa que cubren el área de trabajo con valores ecogeográficos y donde no ocurren las presencias).

La opción "sr" significa que las pseudo ausencias se seleccionarán fuera de las condiciones ambientales para la especie definidas en términos amplios, creando para ello primero un modelo de envoltura ambiental definiendo la parte apropiada para la especie a través de un cuantil (parámetro "pasreq"). Las pseudo ausencias serán seleccionadas fuera de esta envoltura.

La opción "disk" crea áreas buffer circulares alrededor de los puntos de presencia (discos) con una distancia mínima (parámetro "padiskmin") y una distancia máxima (parámetro "padiskmax") respecto de cada punto. Las pseudo ausencias serán seleccionadas dentro del espacio conformado por los discos. La opción "elc" usa la frecuencia de ocurrencia de la especie sobre las categorías de un mapa ELC (producto de la herramienta ELCmapas). El usuario puede elegir de cuáles cuantiles de frecuencia (parámetro "paelc") se extraerán las pseudo ausencias.

15.11.1.24 Parámetro: padiskmin - SelecVIF

Explicación: Sólo aplica para “pastrat” = “disk”. Escriba la distancia (en metros) desde los sitios de presencia hasta el círculo interior de los discos de donde se seleccionarán las pseudo ausencias.

15.11.1.25 Parámetro: padiskmax - SelecVIF

Explicación: Sólo aplica para “pastrat” = “disk”. Escriba la distancia (en metros) desde los sitios de presencia hasta el círculo exterior de los discos de donde se seleccionarán las pseudo ausencias.

15.11.1.26 Parámetro: pasreq - SelecVIF

Explicación: Sólo aplica para “pastrat” = “sre”. Escriba el valor del cuantil el cual definirá el tamaño del “envoltorio” para la especie y el espacio fuera de él donde se seleccionará pseudo ausencias. Este valor debe ser superior a 0 e inferior a 0.5.

15.11.1.27 Parámetro: mapaelc - SelecVIF

Explicación: Sólo aplica para “pastrat” = “elc”. Indique el nombre del archivo que contiene el mapa ELC (producto de la aplicación de la herramienta ELCmapas), el cual debe encontrarse en la carpeta ELCmapas, la cual es una de las carpetas que componen el directorio CAPFITOGEN. Este mapa debe estar en formato DIVA-GIS (extensión .grd tal y como lo produce ELCmapas) y debe escribirse con la extensión. Así, si el nombre del mapa es “mapa_elc_spain”, ud. debe escribir “mapa_elc_spain.grd”.

15.11.1.28 Parámetro: statelc - SelecVIF

Explicación: Sólo aplica para “pastrat” = “elc”. Indique el nombre del archivo que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (esta herramienta suele nombrar a este archivo como “Estadist_ELC_” más el nombre del país o región). Al igual que el mapa ELC, este archivo debe estar situado en la carpeta ELCmapas. De igual forma el nombre debe escribirse junto con la extensión, en este caso al ser una tabla, la extensión es txt. Así si el archivo se llama “Estadist_ELC_spain” se debe escribir “Estadist_ELC_spain.txt”.

15.11.1.29 Parámetro: paelc

Explicación: Sólo aplica para “pastrat” = “elc”. Seleccione uno o varios cuantiles de frecuencia que utilizará para definir las áreas donde se generarán aleatoriamente las pseudo ausencias.

15.11.1.30 Parámetro: modelby (parámetro deshabilitado o que sólo debe indicar “biomod”)

Explicación: Sólo aplica la opción “biomod”. Esta usa el paquete biomod2 (<https://goo.gl/ogj359>). La implementación de biomod2 en CAPFITOGEN permite la utilización de nueve algoritmos (parámetro “modelos” = ‘GLM’, ‘GBM’, ‘GAM

’, ‘CTA’, ‘ANN’, ‘SRE’, ‘FDA’, ‘MARS’, ‘RF’) aptos para datos de presencia-ausencia y sólo presencia y de ‘MAXENT’, que sólo permite datos de sólo presencia (“ausencia” = “pseudo absence”). Los 10 modelos disponibles en “biomod” son configurables (parámetro “modelopc”) y se pueden ensamblar (parámetro “ensamb”). Usando la opción “biomod” es posible utilizar todos los conjuntos de pseudo ausencias generados (parámetro “pareps” puede ser =1 o >1).

15.11.1.31 Parámetro: **dismodel**

Explicación: Sólo aplica para “modelby” = “dismo”. Indique el algoritmo de modelización que desea utilizar dentro del paquete “dismo”.

15.11.1.32 Parámetro: **modcompl**

Explicación: Sólo aplica para “modelby” = “dismo”. Marque esta opción si desea obtener un modelo utilizando el 100% de sus datos de presencia.

15.11.1.33 Parámetro: **diseval**

Explicación: Sólo aplica para “modelby” = “dismo”. Marque esta opción si desea obtener modelos con una parte de datos de presencia (train set), para ser evaluados con los datos reservados (test set).

15.11.1.34 Parámetro: **datadiv2**

Explicación: Sólo aplica para “modelby” = “dismo”. Especifique un valor entre 1 y 100 que determine el porcentaje de datos de presencia destinados a la obtención del modelo (train). El porcentaje restante será destinado para el conjunto de evaluación (test).

15.11.1.35 Parámetro: **modelid**

Explicación: Sólo aplica para “modelby” = “biomod”. Especifique un nombre que identifique este proceso de modelización respecto a otros pasados o futuros.

15.11.1.36 Parámetro: **modelos**

Explicación: Sólo aplica para “modelby” = “biomod”. Seleccione él o los algoritmos con los que desea modelizar. La explicación de los códigos la puede encontrar en <https://goo.gl/ogj359> o en el manual de usuario CAPFITOGEN. Seleccionar ‘MAXENT’ implica que sus datos son de sólo presencia y que debe descargar y colocar apropiadamente un archivo java desde https://biodiversityinformatics.amnh.org/open_source/maxent/ (ver su manual de usuario).

15.11.1.37 Parámetro: **modelopc**

Explicación: Sólo aplica para “modelby” = “biomod”. Marque esta opción si desea utilizar modelizar utilizando pará-

metros particulares para los algoritmos seleccionados. Si marca esta opción, tenga presente que debe proveer archivos de texto (.txt) donde se especifican los parámetros particulares y se deben ubicar en la carpeta 'ModelOptions'. En caso de NO marcar esta opción, los modelos utilizarán los parámetros por defecto (éstos pueden observarse en el archivo Excel 'ModelOptions_OpcionesModelos.xlsx' ubicado en la carpeta 'ModelOptions').

15.11.1.38 Parámetro: GLM

Explicación: Sólo aplica cuando "modelopc" ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo GLM (Generalized Linear Model).

15.11.1.39 Parámetro: GLMopt

Explicación: Sólo aplica cuando "GLM" ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para GLM, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel 'ModelOptions_OpcionesModelos.xlsx'. El archivo de texto debe localizarse en la carpeta 'ModelOptions'.

15.11.1.40 Parámetro: GBM

Explicación: Sólo aplica cuando "modelopc" ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo GBM (Generalized Boosting Model).

15.11.1.41 Parámetro: GBMopt

Explicación: Sólo aplica cuando "GBM" ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para GBM, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel 'ModelOptions_OpcionesModelos.xlsx'. El archivo de texto debe localizarse en la carpeta 'ModelOptions'.

15.11.1.42 Parámetro: GAM

Explicación: Sólo aplica cuando "modelopc" ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo GAM (Generalized Additive Model).

15.11.1.43 Parámetro: GAMopt

Explicación: Sólo aplica cuando "GAM" ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para GAM, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel 'ModelOptions_OpcionesModelos.xlsx'. El archivo de texto debe localizarse en la carpeta 'ModelOptions'.

15.11.1.44 Parámetro: CTA

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo CTA (Classification Tree Analysis).

15.11.1.45 Parámetro: CTAopt

Explicación: Sólo aplica cuando “CTA” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para CTA, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.46 Parámetro: ANN

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo ANN (Artificial Neural Network).

15.11.1.47 Parámetro: ANNopt

Explicación: Sólo aplica cuando “ANN” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para ANN, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.48 Parámetro: SRE

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo SER (Surface Range Envelop, usualmente llamado BIOCLIM).

15.11.1.49 Parámetro: SREopt

Explicación: Sólo aplica cuando “SRE” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para SRE, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.50 Parámetro: FDA

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo FDA (Flexible Discriminant Analysis).

15.11.1.51 Parámetro: FDAopt

Explicación: Sólo aplica cuando “FDA” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para FDA, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.52 Parámetro: MARS

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo MARS (Multiple Adaptive Regression Splines).

15.11.1.53 Parámetro: MARSopt

Explicación: Sólo aplica cuando “MARS” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para MARS, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.54 Parámetro: RF

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo RF (Random Forest).

15.11.1.55 Parámetro: RFopt

Explicación: Sólo aplica cuando “RF” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para RF, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.56 Parámetro: MAXENT

Explicación: Sólo aplica cuando “modelopc” ha sido marcado. Marque esta opción si desea especificar los parámetros para el algoritmo MAXENT (Maximum Entropy).

15.11.1.57 Parámetro: MAXENTopt

Explicación: Sólo aplica cuando “MAXENT” ha sido marcado. Escriba el nombre del archivo de texto (incluyendo la extensión .txt) que contiene la lista de parámetros para MAXENT, algunos de ellos posiblemente modificados por Ud. Este archivo puede ser generado a partir del archivo excel ‘ModelOptions_OpcionesModelos.xlsx’. El archivo de texto

debe localizarse en la carpeta ‘ModelOptions’.

15.11.1.58 Parámetro: modrep

Explicación: Sólo aplica para “modelby” = “biomod”. Especifique el número de repeticiones de evaluación de los modelos. Considere que un alto número de “modrep” junto a un alto número de “pareps” y el uso de varios algoritmos (“modelos”) implica la generación de muchos modelos que puede tomar mucho tiempo y capacidad de cómputo.

15.11.1.59 Parámetro: datadiv

Explicación: Sólo aplica para “modelby” = “biomod”. Especifique un valor entre 1 y 100 que determine el porcentaje de datos de presencia destinados a la obtención del modelo (train). El porcentaje restante será destinado para el conjunto de evaluación (set).

15.11.1.60 Parámetro: preval

Explicación: Sólo aplica para “modelby” = “biomod”. Determine el valor de prevalencia. Determinar la prevalencia le permite asignar más o menos peso a ciertas observaciones. Si “preval” = NA cada observación tiene el mismo peso (independientemente del número de presencias o ausencias), esta es la opción por defecto. Si “preval” = 0.5, las ausencias tendrán el mismo peso que las presencias. Si “preval” se fija por debajo o por encima de 0.5, las ausencias o presencias tendrán más peso, respectivamente.

15.11.1.61 Parámetro: imporvar

Explicación: Sólo aplica para “modelby” = “biomod”. Especifique el número de permutaciones para estimar la importancia de las variables (predictores) en el proceso de modelización. Si el valor especificado aquí es 0, no se estimará la importancia de las variables.

15.11.1.62 Parámetro: testparam

Explicación: Sólo aplica para “modelby” = “biomod”. Seleccione uno o varios métodos de evaluación de modelos.

15.11.1.63 Parámetro: reescal

Explicación: Sólo aplica para “modelby” = “biomod”. Marque esta opción si desea re-escalar todas las predicciones de los algoritmos de modelización mediante un GLM binomial (escala de 0 a 1000, algo que requiere la herramienta Mcompare). Si no marca esta opción, la predicción estará en la escala que produce cada algoritmo.

15.11.1.64 Parámetro: modcomp

Explicación: Sólo aplica para “modelby” = “biomod”. Marque esta opción si desea obtener un modelo utilizando el

100% de sus datos de presencia.

15.11.1.65 Parámetro: modprop

Explicación: Sólo aplica para “modelby” = “biomod”. Marque esta opción si desea que en la formación de conjuntos de presencia/ausencia o presencia/pseudoausencia para entrenar el modelo (test set) y para evaluarlo (test set), el balance de presencias/ausencias o presencias/pseudo ausencias sea el mismo que ocurre en los datos originales. De lo contrario, los conjuntos de entrenamiento y evaluación se conformarán con presencias/ausencias o presencias/pseudo ausencias donde se puede o no mantener el balance original. En ambos casos la selección se hará de forma aleatoria.

15.11.1.66 Parámetro: proysome

Explicación: Sólo aplica para “modelby” = “biomod”. Marque esta opción si desea proyectar sólo los modelos que han superado un valor de umbral para un método de evaluación (ambos definidos por el usuario). No marcar esta opción indica que se proyectarán todos los modelos.

15.11.1.67 Parámetro: proytest1

Explicación: Sólo aplica cuando “proysome” ha sido marcado. Seleccione un método de evaluación primario por el cual quiere filtrar los modelos a ser proyectados mediante la fijación de un umbral (parámetro “proytests1u”).

15.11.1.68 Parámetro: proyest1u

Explicación: Sólo aplica cuando “proysome” ha sido marcado. Indique el valor de umbral para el método evaluador primario. Sólo los modelos que superen este valor serán proyectados. Tenga en cuenta que el valor de desempeño óptimo es usualmente 1.

15.11.1.69 Parámetro: proysome2

Explicación: Sólo aplica cuando “proysome” ha sido marcado. Marque esta opción si desea usar un segundo método de evaluación para filtrar de nuevo el grupo de modelos que serán proyectados.

15.11.1.70 Parámetro: proytest2

Explicación: Sólo aplica cuando “proysome2” ha sido marcado. Seleccione un método de evaluación secundario por el cual quiere volver a filtrar los modelos a ser proyectados usando un segundo umbral (parámetro “proytests2u”).

15.11.1.71 Parámetro: proytest2u

Explicación: Sólo aplica cuando “proysome2” ha sido marcado. Indique el valor de umbral para el método evaluador secundario. Sólo los modelos (ya previamente filtrados) que superen este segundo umbral serán proyectados. Tenga

en cuenta que el valor de desempeño óptimo es usualmente 1.

15.11.1.72 Parámetro: binar

Explicación: Sólo aplica para “modelby” = “biomod”. Marque ésta opción si desea obtener un mapa con la respuesta binarizada (valores 0 para baja probabilidad y 1 para alta probabilidad de presencia) en base a la binarización de uno de los métodos de evaluación utilizados (parámetro “testparam”).

15.11.1.73 Parámetro: binarmet

Explicación: Sólo aplica cuando “binar” ha sido marcado. Seleccione el método de evaluación cuyo proceso de binarización desea utilizar para generar el mapa de respuesta binarizada. Este método debe estar en la lista de los seleccionados en el parámetro “testparam”.

15.11.1.74 Parámetro: maskout

Explicación: Sólo aplica para “modelby” = “biomod”. Marque ésta opción si desea obtener un mapa máscara que identificará las áreas donde las predicciones son inciertas dado que los valores de los variables (predictores) allí están fuera del rango usado para calibrar los modelos.

15.11.1.75 Parámetro: ensamb

Explicación: Sólo aplica para “modelby” = “biomod”. Marque ésta opción si desea ensamblar modelos. Puede ensamblar todos los modelos producidos o sólo los mejor evaluados.

15.11.1.76 Parámetro: mod2bens

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Indique si desea ensamblar todos los modelos (opción “all”) o sólo los modelos mejor evaluados (“best”). Corresponde al parámetro “chosen.models” de la función BIOMOD_EnsembleModeling del paquete biomod2.

15.11.1.77 Parámetro: tipensam

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Seleccione la forma en que se ensamblarán los modelos. Para entender lo que implica cada opción por favor revise el documento “EnsembleModelingAssembly.pdf” que puede descargar desde <https://www.capfitogen.net/es/EnsembleModelingAssembly.pdf> o que encuentra en la carpeta “Documentación_Referencias/Modela”. Corresponde al parámetro “em.by” de la función BIOMOD_EnsembleModeling del paquete biomod2.

15.11.1.78 Parámetro: ensamet

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea aplicar los mismos evaluadores (y los mismos umbrales) para seleccionar los modelos a ser ensamblados que originalmente fueron usados para seleccionar los modelos a ser proyectados (parámetros “proytest1” y “proytest1u”, opcionalmente “proytest2” y “proytest2u”). Si no marca esta opción, serán usados todos los evaluadores disponibles. Los evaluadores y sus umbrales que esta opción permite utilizar serán asignados al parámetro “eval.metric” y “eval.metric.quality.threshold” (respectivamente) de la función BIOMOD_EnsembleModeling del paquete biomod2.

15.11.1.79 Parámetro: testparam2

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea evaluar los modelos ensamblados.

15.11.1.80 Parámetro: testparam3

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Seleccione uno o varios métodos de evaluación de modelos. Corresponde al parámetro “models.eval.meth” de la función BIOMOD_EnsembleModeling del paquete biomod2.

15.11.1.81 Parámetro: probmean

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea el promedio de las probabilidades de las predicciones (modelos).

15.11.1.82 Parámetro: probcv

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea el coeficiente de variación de las probabilidades de las predicciones (modelos).

15.11.1.83 Parámetro: probci

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea el intervalo de confianza sobre el promedio de las probabilidades de las predicciones (modelos).

15.11.1.84 Parámetro: probalfa

Explicación: Sólo aplica cuando “probci” ha sido marcado. Indique el nivel de significancia para estimar el intervalo de confianza.

15.11.1.85 Parámetro: probmedian

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea la mediana de las probabilidades de las predicciones (modelos).

15.11.1.86 Parámetro: probca

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea por medio de un comité de promedios. Para este método se usarán los modelos binarizados. El puntaje del comité de promedios será el promedio de las predicciones binarias a manera de una votación sencilla, cada modelo votando por la presencia o ausencia de la especie. Así para cada sitio, la suma de 1 es dividida por el número de modelos. De esta manera se da una medida de predicción y una de incertidumbre. Así valores cercanos a 0 o 1 están indicando que la mayoría de los modelos concuerdan en la ausencia o presencia de la especie. Un valor alrededor de 0.5 significa que la mitad de los modelos señalan la presencia y la mitad la ausencia.

15.11.1.87 Parámetro: probmw

Explicación: Sólo aplica cuando “ensamb” ha sido marcado. Marque esta opción si desea que el método de ensamblaje de modelos sea la suma ponderada de las probabilidades de las predicciones (modelos).

15.11.1.88 Parámetro: probmwd

Explicación: Sólo aplica cuando “probmw” ha sido marcado. Defina la importancia de los pesos. La opción proporcional (“proportional”) asigna pesos proporcionales a la evaluación que cada modelo recibió en el proceso de modelización.

15.11.1.89 Parámetro: resultados (sólo se utiliza cuando se usa el modo local) - SelecVIF

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / antes que \. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

15.11.2 Parámetros exclusivos SelecVIF**15.11.2.1 Parámetro: testNAs**

Explicación: Indique si desea (TRUE) o no (FALSE) remover las presencias o sitios de recolección que no extraen información, previo el proceso de modelización.

15.11.2.2 Parámetro: cutoff

Explicación: Introduzca un valor numérico que indique el umbral de VIF por encima del cual las variables serán descartadas.

15.11.2.3 Parámetro: dynamic

Explicación: Indique si desea (TRUE) o no (FALSE) realizar un filtrado de variables por VIF a través de un proceso paso

a paso, eliminando variables de alta VIF en cada uno y determinando de nuevo el VIF de las variables restantes.

15.12. Resultados de Modela

Después de definir todos los parámetros de Modela y hacer click en el botón “Analizar”, el proceso de generación de pseudo ausencias (en su caso), modelización, proyección de los modelos y ensamblaje (en su caso), puede tomar un tiempo variable, que puede ir de un par de minutos a horas. Grandes marcos de trabajo, altas resoluciones (tamaños de celda reducidos) y la utilización de muchos algoritmos y evaluadores puede ralentizar el proceso y/o producir demandas excesivas de recursos de la computadora.

Describir en detalle todos los posibles resultados de Modela, como se ha hace en otros capítulos para otras herramientas resultaría en un texto demasiado extenso y enrevesado. Por tanto, la descripción a continuación mostrará los patrones de cómo se organizan los resultados, cómo se componen los nombres de las carpetas y el tipo de tablas y mapas que pueden producirse.

En la carpeta de resultados, además del archivo de texto que muestra las opciones de los parámetros utilizados en la modelización “Parametros.Parameters.Modela.txt”, pueden aparecer una serie de carpetas donde se organizan los productos de Modela. Las carpetas dentro de resultados se deben básicamente al tipo de datos de presencia/ausencia (presencia/pseudoausencia o presencia/ausencia real) que se pueden observar en las Tabla 5, y las carpetas que se deben a las opciones de modelización (Tabla 6).

Tabla 5. Carpetas que se producen en la ruta especificada en el parámetro “resultados” como resultado del tipo de datos de presencia/ausencia utilizados. Las columnas “ausencia” y “pastrat”. Los términos en color rojo varían de acuerdo a lo especificado en los parámetros “genero” y “especie”.

Opciones “ausencia”	Opciones “figs” o “pastrat”	Carpetas	Código
absence	figs=FALSE	Real_Absences_ genero_especie	1
absence	figs=TRUE	Absences_NonEval_FIGS_ genero_especie	2
pseudo absence	sre	PseudoAbsences_SRE_ genero_especie	3
	disk	PseudoAbsences_Disk_ genero_especie	4
	elc	PseudoAbsences_ELC_ genero_especie	5
	random	PseudoAbsences_ genero_especie	6

Tabla 6. Carpetas que se producen en la ruta especificada en el parámetro “resultados” como resultado de la modelización con dismo (parámetro modelby = “biomod”). Los términos en color rojo varían de acuerdo a lo especificado en los parámetros “genero”, “especie” y “modelid”.

Opciones	Carpetas	Código
Básico	genero.especie	10
	Results_Evaluation_Models_modelid	11
	Results_Project_raw_modelid	12
imporvar > 0 (apartado 15.11.1.63)	Results_Variables_importance_modelid	13
maskout = TRUE (apartado 15.11.1.76)	Results_Clamping_mask_modelid	14
binar = TRUE (apartado 15.11.1.74)	Results_Project_binar_modelid	15
ensamb = TRUE (apartado 15.11.1.77)	Results_Ensemb_Individual_Models_modelid	16
ensamb = TRUE y mod2bens = best (apartados 15.11.1.77 y 15.11.1.78)	Results_Ensemb_Best_Model_modelid	17
ensamb = TRUE y testparam2= TRUE (apartados 15.11.1.77 y 15.11.1.81)	Results_Evaluation_Ensemble_modelid	18
temp = future (apartado 15.11.1.9)	Results_Project_raw_modelid_F	19
temp = future y maskout = TRUE, ensamb = TRUE, mod2bens = best, binar = TRUE	Results_Clamping_mask_modelid_F	20
	Results_Project_binar_modelid_F	21
	Results_Ensemb_Individual_Models_modelid_F	22
	Results_Ensemb_Best_Model_modelid_F	23
figs = TRUE y ensamb = TRUE (apartados 15.11.1.21 y 15.11.1.77)	Results_Emsemb_non_evaluated_germplasm	24

Gracias a estas tablas que muestran el origen de cada carpeta en términos de los parámetros que la originan, es posible asociar el nombre de cada carpeta a la naturaleza de su contenido. Para referirnos con mayor facilidad a las carpetas posteriormente, éstas serán citadas a través del código que aparece en la parte derecha de cada tabla.

Es importante aclarar que la carpeta “genero.especie” (que viene de los parámetros “genero” y “especie”) que aparece cuando se modeliza por biomod2 (parámetro modelby = “biomod”), contiene una serie de carpetas y archivos que produce el paquete biomod2 de R, útiles para técnicos familiarizados con el funcionamiento de este paquete. Sin embargo, a efectos de la utilización práctica de la herramienta Modela, no resulta de valor su explicación ya que los contenidos en su mayor parte sólo pueden ser visualizados en R.

15.12.1. Tablas de Modela

15.12.1.1. Tablas de presencias/ausencias o presencias/pseudo ausencias

La mayor parte de las tablas en los resultados de Modela están presentes en las carpetas 1 a 6, que son las carpetas relacionadas con los datos de presencia y ausencia o pseudo ausencia utilizados para la modelización. La Tabla 8 lista las posibles tablas que el usuario de Modela se puede encontrar en los resultados, carpetas 1 a 6.

Tabla 7. Archivos correspondientes a tablas relacionadas con el tipo de datos (presencias/ausencias o presencias/pseudo ausencias) utilizados para modelizar.

Opciones	Nombre archivo	Código carpeta	Descripción tablas
cualquier vía de modelización	input_data.txt	1,2,3,4,5,6	Tabla que muestra cifras sobre presencias y ausencias /pseudoausencias utilizadas en al modelización y las variables ecogeográficas
cualquier vía de modelización	extraction_table_PresAbs.txt extraction_table_PresAbs.xls extraction_table_Random.txt extraction_table_Random.xls extraction_table_Sre.txt extraction_table_Sre.xls extraction_table_Disk.txt extraction_table_Disk.xls extraction_table_ELC.txt extraction_table_ELC.xls	1,2,3,4,5,6	Tabla que contiene los valores extraídos de las diferentes capas ecogeográficas para los datos de presencia y ausencia o pseudo ausencia (marcados como 1 y 0 o NA respectivamente en la columna "inputs.data.species")
ausencia="pseudo absence" y pastrat="elc"	Genbank_ELC.txt Genbank_ELC.xls	5	Tabla que contiene todos los campos de la tabla de pasaporte y una columna adicional (BGcat) que contiene la categoría del mapa ELC extraída para cada sitio.
ausencia="pseudo absence" y pastrat="elc"	PATable.txt PATable.xls	5	Tabla que contiene información para identificar las pseudo ausencias generadas (columna resvar, valores = NA) que hacen parte de cada conjunto de repeticiones (columnas PA1, PA2, PA3,... PAn), TRUE=incluida, FALSE=excluida
ausencia="pseudo absence" y pastrat="elc"	Quartile_ELCmap_Classification.txt Quartile_ELCmap_Classification.xls	5	Tabla que contiene cifras, porcentajes, frecuencias y clases (cuartiles) en los que se reclasifica el mapa ELC, para permitir seleccionar áreas disponibles para la obtención de pseudo ausencias

15.12.1.2. Tablas de evaluación

Al modelizar por biomod2 (parámetro modelby = “biomod”), se obtendrá una sola tabla llamada “models_evaluation_results.txt” o “models_evaluation_results.xls” que aparecerá en las carpetas 11 y 18 si se ha optado por ensamblar modelos. Esta tabla contiene las siguientes columnas:

- Model.name: especifica cada modelo evaluado con un nombre compuesto, en el caso de parámetro ausencia = “pseudo absence” por: genero.especie_PAX_RUNY_ZZZ, siendo X el conjunto de pseudo-ausencias, Y la repetición del modelo, y ZZZ el algoritmo utilizado. Si se ha modelizado con datos de presencia y ausencia (parámetro ausencia = “absence”), el elemento RUNY se reemplaza por AllData.
- Eval.metric: especifica el evaluador.
- Testing.data: corresponde al valor alcanzado para el evaluador por el modelo.
- Cutoff: corresponde al valor de umbral optimizado para el evaluador.
- Sensitivity: parámetro de sensibilidad.
- Specificity: parámetro de especificidad.

15.12.1.3. Tabla de predicción FIGS

La tabla de predicción de la ocurrencia del rasgo aparece en la carpeta 24 y el archivo se llama “ensembled_prediction_nonevaluated_germplasm.txt” (o con terminación .xls para formato Excel). Esta tabla contiene las coordenadas de las accesiones sobre las que se ha proyectado los modelos (columnas “DECLONGITUDE” y “DECLATITUDE”) y el valor de predicción de cada modelo original o ensamblado, con columnas nombradas con los componentes de nombres especificados en el apartado 15.12.2.2. Los valores de predicción están en un rango de 0 a 1000, siendo los valores cercanos a 1000 los que indican mayor probabilidad de aparición del rasgo.

15.12.2. Mapas de Modela

Los mapas de predicción son los resultados de tipo práctico más importantes dentro de la herramienta Modela. Sin embargo, Modela también ofrece otros tipos de mapas que pueden ayudar al análisis de los resultados. Así, dentro de algunas carpetas especificadas en las Tablas 5 y 6, aparecerán mapas en formato ráster (extensiones .tif o .grd) y/o vectorial (shapefiles). Cualquier tipo de mapa producido por Modela puede ser visualizado e incluso modificado en el programa DIVA-GIS (<http://www.diva-gis.org>).

15.12.2.1. Mapas de presencias/ausencias o presencias/pseudo ausencias

Los mapas relacionados a las presencias/ausencias o presencias/pseudo ausencias se detallan en la Tabla 9.

Tabla 8. Archivos correspondientes a mapas que muestran la ubicación de las presencias/ausencias o presencias/pseudo ausencias y, en su caso, las áreas de donde se obtuvieron pseudo ausencias vía pastrat= "elc".

Opciones	Nombre archivo	Código carpeta	Descripción mapa
De acuerdo a las opciones seleccionadas en los parámetros ausencia y pastrat	Final_extraction_PresAbs.shp Final_extraction_Random.shp Final_extraction_Sre.shp Final_extraction_Disk.shp Final_extraction_ELC.shp	1,3,4,5,6	Mapa vectorial de puntos, muestra presencias y ausencias/pseudoausencias, señaladas con 1 y 0 en la columna "inputs_dat" respectivamente, y los valores extraídos para cada sitio de las capas ecogeográficas
ausencia= "pseudo absence" y pastrat="elc"	Presences_Pseudoabsences_ELC.shp	5	Mapa vectorial de puntos, muestra presencias y pseudo ausencias. En las columnas PA1, PA2, PA3.. hasta PAn siendo n el valor del parámetro "pareps" con valores "True" y "False" muestra la distribución de sitios en cada conjunto de pseudoausencias.
ausencia= "pseudo absence" y pastrat="elc"	mapa_Class_Sp.grd	5	Mapa ráster que muestra la clasificación en cuartiles de frecuencia del mapa ELC.
ausencia= "pseudo absence" y pastrat="elc"	PA_availab.grd	5	Mapa ráster que muestra las áreas disponibles (valor 1) para la obtención de los pseudo ausencias de acuerdo a lo especificado en el parámetro "paelc".
ausencia= "absence" y figs=TRUE	points_to_be_predicted.shp	2	Mapa vectorial que muestra la ubicación de las accesiones no evaluadas

15.12.2.2. Estructura de los nombres de los mapas

Los archivos que contienen los diferentes mapas de predicción, binarizados o no, proyectados sobre capas bioclimáticas de futuro o de presente, tienen nombres compuestos por elementos que indican inequívocamente lo que representa cada mapa.

A continuación se explica el origen o significado de los elementos que pueden componer la estructura de los nombres de los mapas ráster:

- **AllData:** sólo aparece cuando se ha indicado la opción "absence" en el parámetro ausencia, en las carpetas 12, 15, 16, 19, 21 o 22, e indica que el mapa proviene de un modelo que utilizó datos de presencia/ausencia.

- **PA:** usualmente aparece en las carpetas 12, 15, 19 y 21. En general, indica la utilización de pseudo ausencias, y si va acompañado de un número (por ejemplo PA1) indica que el mapa proviene de un modelo que utilizó uno de los conjuntos de pseudo ausencias (conjunto indicado por el número)
- **complete o full:** usualmente aparece en las carpetas 7, 8, 9, 12 o 19. indica que el mapa corresponde a un modelo obtenido con todos los datos disponibles de presencia/ausencia o presencia/pseudo ausencia.
- **F o Fut:** sólo aparece cuando se indica la opción “future” en el parámetro temp en las carpetas 7.2, 8.2, 9.2, 19, 20, 21, 22 o 23. Indica que el mapa corresponde a una proyección sobre capas bioclimáticas de futuro.
- **binar o binary:** este componente de nombre puede aparecer en las carpetas 7.2, 8.2, 9.2, 15 o 21. indica que el mapa corresponde a una proyección binarizada mediante la aplicación de un umbral. Si además viene acompañado con el nombre de un evaluador, se indica que el umbral utilizado proviene de este evaluador, por ejemplo “binaryROC” indica que el umbral optimizado para ROC fue utilizado para binarizar la proyección.
- **RUN:** sólo aparece cuando se modeliza por biomod2 y usualmente en las carpetas 12 y 19. Indica que el mapa corresponde a una proyección utilizando una de las repeticiones de modelización (determinadas en el parámetro “modrep”).
- **EMxxByXXX:** Sólo aparece cuando se ha ensamblado modelos (parámetro ensamb=TRUE) en las carpetas 16 y (si parámetro temp=“future”) 22. Indica la función mediante la cual se han ensamblado los modelos. Después de las letras EM aparecerá cualquiera de las siguientes opciones: mean, cv, ciInf y ciSup, median, ca y wmean, indicando la función de ensamblaje determinada en los parámetros probmean, provcv, probci, probmedian, probca y probmw respectivamente. Después de las letras By aparecerá cualquier de las siguientes opciones: KAPPA, TSS, ROC, FAR, SR, ACCURACY, BIAS, POD, CSI y ETS, indicando el evaluador utilizado.
- **Clamping_mask_map:** Aparece en las carpetas 14 y 20. Indica que se trata del mapa máscara que identifica zonas de predicción incierta que se origina por el parámetro maskout=TRUE.
- **GLM, GBM, GAM, CTA, ANN, SRE, FDA, MARS, RF o MAXENT:** Estos términos que indican el algoritmo utilizado para modelizar, aparecen cuando se modeliza vía biomod2 y suelen ser la parte final del nombre de los archivos de mapas en las carpetas 12 y 19 (mapas de proyección en escala 0-1000) y en la parte media de los archivos de mapa en las carpetas 15 y 21 (mapas de proyección binarizados).
- **mergedXXXX:** Sólo aparece cuando se ha ensamblado modelos (parámetro ensamb=TRUE) y lo hace en archivos de mapa en las carpetas 16 y (si parámetro temp=“future”) 22. Indica la estrategia que se usó para combinar los modelos y construir los modelos ensamblados que se ha establecido en el parámetro tipensam. Por ello, luego de la palabra “merged” aparecerá el código de la vía utilizada para ensamblar (aquí señalado con las letras XXXX), que puede ser: “Run” (se ensambló por las diferentes repeticiones de modelización señaladas en el parámetro “modrep”), “Algo” (se ensambló por los diferentes algoritmos utilizados, parámetro “modelos”) o “Data” (se ensambló por los modelos obtenidos de los diferentes conjuntos de pseudo ausencias o PA_dataset, parámetro “pareps”). Pueden aparecer a lo largo del nombre del mapa varios “mergedXXXX”, indicando que se ha ensamblado por una combinación de vías (opciones “all”, “PA_dataset+repet”, “PA_dataset+algo” en el parámetro “tipensam”). En el caso de haber seleccionado “all”, por ejemplo, en el nombre de los archivos de los mapas ensamblados aparecerá “mergedAlgo_mergedRun_mergedData”.
- **Best:** Sólo aparece cuando se ha indicado la opción “best” en el parámetro mod2bens, en las carpetas 17 y 23, e indica que se trata de los mapas de los modelos ensamblados que han superado un valor umbral para uno de los evaluadores utilizados.

15.12.3. Eliminación de archivos y carpetas temporales

Sólo aplica para utilización en modo local. Como consecuencia de la obtención de modelos y mapas (de mayor o menor tamaño en Mb o incluso Gb), se generan una serie de archivos temporales que pueden llegar a saturar su disco duro principal, usualmente la unidad C:\. Estos archivos temporales se acumulan en una ruta un tanto larga que inicia en la carpeta de usuarios (regularmente Windows le asigna el nombre de “Users”). Un ejemplo de esta ruta puede ser: C:\Users\pc\AppData\Local\Temp

En esa carpeta “Temp”, Windows almacena temporales de varios programas (software), por tanto aparecerán varias carpetas dentro con nombres de los programas. Una vez ha acabado todos los procesos en Modela o con cualquier otra herramienta, se le recomienda al usuario cerrar todas las aplicaciones relacionadas con CAPFITOGEN y dirigirse a esta carpeta de temporales, buscar una carpeta que refiere al software R, usualmente de nombre “R_raster_pc” y determinar su tamaño. Considere la eliminación de los contenidos de esta carpeta (mas no la eliminación de la propia carpeta) si el tamaño que ocupa es suficiente como para ralentizar el funcionamiento de su PC. Esta eliminación, mientras no se realice durante un proceso de Modela, de cualquier otra herramienta o de cualquier otra aplicación en R, no tendrá ningún efecto en su funcionamiento.

Finalmente, si se ha utilizado el algoritmo Maxent mediante la modelización vía biomod2 (parámetros modelby=”biomod” y modelos=”MAXENT”), el proceso deja una carpeta vacía con el nombre compuesto por el género (parámetro “genero”, apartado 15.11.1.17) y la especie (parámetro “especie”, apartado 15.11.1.18), en la ruta señalada en el parámetro “ruta” (apartado 15.11.1.3). Esta carpeta puede ser eliminada sin mayor inconveniente.

15.13. Referencias

Akinwande, M. O., Dikko, H. G., Samson, A. 2015. Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open Journal of Statistics*, 5(07), 754-767.

Allouche, O., Tsor, A., Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 5(7): 1223-1232.

Barbet-Massin, M., Jiguet, F., Albert, C. H., Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many?. *Methods in Ecology and Evolution*, 3(2), 327-338.

Barnes, L. R., Grunfest, E. C., Hayden, M. H., Schultz, D. M., Benight, C. 2007. False alarms and close calls: A conceptual model of warning accuracy. *Weather and Forecasting*, 22(5), 1140-1147.

Barnes, L. R., Schultz, D. M., Grunfest, E. C., Hayden, M. H., Benight, C. C. 2009. Corrigendum: False alarm rate or false alarm ratio?. *Weather and Forecasting*, 24(5), 1452-1454.

Booth, T. H., Nix, H. A., Busby, J. R., Hutchinson, M. F. 2014. BIOCLIM: the first species distribution modelling

package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, 20(1), 1-9.

Bradter, U., Kunin, W. E., Altringham, J. D., Thom, T. J., Benton, T. G. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. *Methods in Ecology and Evolution*, 4(2), 167-174.

Chefaoui, R. M., Lobo, J. M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological modelling*, 210(4), 478-486.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-40.

Elith, J., Leathwick, J. R., Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

Elith, J., Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43-57.

Fawcett, T. 2004. ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31, 1-38.

Fielding, A. H., Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, 24(01), 38-49.

Guisan, A., Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological modelling*, 135(2), 147-186.

Guisan, A., Edwards, T. C., Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, 157(2), 89-100.

Guisan, A., Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9), 993-1009.

Hijmans, R., Elith, J., 2015. Species distribution modeling with R. <https://goo.gl/p8beyk>

Jarvis, A.; Williams, K.; Williams, D.; Guarino, L.; Caballero, P.J. Mottram, G. 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet. Resour. Crop Evol.* 52:671-682.

Jarvis, A., Lane, A., Hijmans, R. J. 2008. The effect of climate change on crop wild relatives. *Agriculture, ecosystems & environment*, 126(1), 13-23.

Liu, C., White, M., Newell, G. 2009. Measuring the accuracy of species distribution models: a review. In *Proceedings 18th World IMACs/MODSIM Congress*. Cairns, Australia (pp. 4241-4247).

Lobo, J. M., Jiménez-Valverde, A., Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography*, 17(2), 145-151.

Maiorano, L., Cheddadi, R., Zimmermann, N. E., Pellissier, L., Petitpierre, B., Pottier, J., Guisan, A. 2013. Building the niche through time: using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*, 22(3), 302-317.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions*, 15(1), 59-69.

Mateo, R. G., Croat, T. B., Felicísimo, A. M., Munoz, J. 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions*, 16(1), 84-94.

Özesmi, S. L., Özesmi, U. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological modelling*, 116(1), 15-31.

Parra-Quijano, M., Iriondo, J.M., Frese, L., Torres, E.. 2012a. Spatial and ecogeographic approaches for selecting genetic reserves in Europe. En: Maxted, N., Dulloo, M.E., Ford-Lloyd, B.V., Frese, L., Iriondo, J., Pinheiro de Carvalho, M.A.A. (ed.) *Agrobiodiversity Conservation: securing the diversity of crop wild relatives and landraces*. CABI, Wallingford, UK

Parra-Quijano, M., Iriondo, J. M., Torres, E. 2012b. Applications of ecogeography and geographic information systems in conservation and utilization of plant genetic resources. *Spanish journal of agricultural research*. 2: 419-429.

Parra-Quijano, M., Iriondo, J. M., Torres, E. 2012c. Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodiversity and Conservation*, 21(1), 79-96.

Pearce, J., Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological modelling*, 133(3), 225-245.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197.

Ramirez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D., Guarino, L. 2010. A gap analysis methodology for collecting crop gene pools: a case study with Phaseolus beans. PLoS ONE 5(10), e13497. doi:10.1371/journal.pone.0013497.

Ramirez-Villegas, J., Jarvis, A., Läderach, P. 2013. Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. Agricultural and Forest Meteorology, 170, 67-78.

Roebber, P. J. 2009. Visualizing multiple measures of forecast quality. Weather and Forecasting, 24(2), 601-608.

Russell, J., van Zonneveld, M., Dawson, I. K., Booth, A., Waugh, R., Steffenson, B. 2014. Genetic diversity and ecological niche modelling of wild barley: refugia, large-scale post-lgm range expansion and limited mid-future climate threats. PloS one, 9(2), e86021.

Scheldeman, X., van Zonneveld, M. 2011. Manual de Capacitación en Análisis Espacial de Diversidad y Distribución de Plantas. Bioversity International, Roma, Italia.

Sillero, N., Barbosa, A. M. 2021. Common mistakes in ecological niche models. International Journal of Geographical Information Science, 35(2):213-226.

Thuiller, W., Araújo, M. B., Lavorel, S. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. Journal of Vegetation Science, 14(5), 669-680.

Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. Global Change Biology, 10(12), 2020-2027.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M. B. 2009. BIOMOD—a platform for ensemble forecasting of species distributions. Ecography, 32(3), 369-373.

VanDerWal, J., Shoo, L. P., Graham, C., Williams, S. E. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know?. Ecological modelling, 220(4), 589-594.

Wood, J. M. 2007. Understanding and Computing Cohen's Kappa: A Tutorial. WebPsychEmpiricist. URL: <http://wpe.info/>.



Taller Regional, México (capacitación online), Marzo 2021.



16 | Herramienta Mcompare

16.1. Comparación de predicciones presente-futuro

Scheldeman y van Zonneveld (2011) plantearon una interesante comparación entre proyecciones de distribución de especies usando predictores climáticos de presente y futuro dentro de un escenario de cambio climático. Para esta comparación, los predictores climáticos (capas SIG) de presente y futuro provenían de worldclim (<http://www.worldclim.org>).

Para la obtención de proyecciones de presente y futuro, se realiza un proceso de modelización con datos de presente para obtener un patrón. Posteriormente se proyecta el patrón sobre capas SIG de presente y sobre capas SIG compuestas por capas bioclimáticas de futuro y si fuera el caso, capas edáficas y geofísicas de presente, un proceso que lleva a cabo la herramienta Modela (parámetro temp, opción “future”). Las capas SIG que usa Modela para proyectar distribuciones potenciales de presente son las mismas que usan el resto de las herramientas CAPFITOGEN (para la parte bioclimática se corresponden a las variables de presente que ofrece worldclim). Las capas SIG de futuro son las que descarga y adapta la herramienta Bfuture, también provenientes de worldclim.

Una vez obtenidas las proyecciones, se realiza una comparación celda a celda, tratando de establecer la ocurrencia de cualquiera de las cuatro siguientes situaciones descritas por Scheldeman y van Zonneveld (2011) en su “Manual de Capacitación en Análisis Espacial de Diversidad y Distribución de Plantas”:

1. Reducción de la distribución: Denominado por Scheldeman y van Zonneveld como “Áreas de alto impacto”. Son celdas donde la especie tiene probabilidad de ocurrencia en condiciones de presente y la deja de tener en condiciones de futuro proyectado. Representan una amenaza en la persistencia de la especie dentro del escenario de cambio climático utilizado.
2. No presente - no futuro: Denominado por Scheldeman y van Zonneveld como “Áreas por fuera del nicho realizado”. Son celdas donde la especie no tiene probabilidad de ocurrencia ni en condiciones de presente ni en condiciones de futuro proyectado. Representan áreas que la especie no ha podido ni podrá ocupar asumiendo el escenario de cambio climático utilizado.
3. Mantenimiento de la distribución: Denominado por Scheldeman y van Zonneveld como “Áreas de bajo impacto”. Son celdas donde la especie tiene probabilidad de ocurrencia tanto en condiciones de presente como de futuro proyectado. Representan aquellas áreas que la especie es capaz de seguir ocupando a pesar del cambio climático, posiblemente gracias a su plasticidad y/o capacidad o rango de adaptación.
4. Aumento de la distribución: Denominado por Scheldeman y van Zonneveld como “Nuevas áreas adecuadas”. Son aquellas celdas donde la especie no tiene probabilidades de ocurrencia para condiciones de presente pero si las tiene para condiciones proyectadas de futuro. Representan zonas donde la especie no ocurría pero que podría ocupar (zona de expansión) debido a que el escenario de cambio climático utilizado es favorable o compatible con la capacidad adaptativa de la especie en estas zonas.

Esquemáticamente, la Figura 56 muestra cómo se realiza celda por celda la clasificación de las situaciones previamente descritas en base a la comparación de distribuciones predictivas sobre escenarios de presente y futuro.

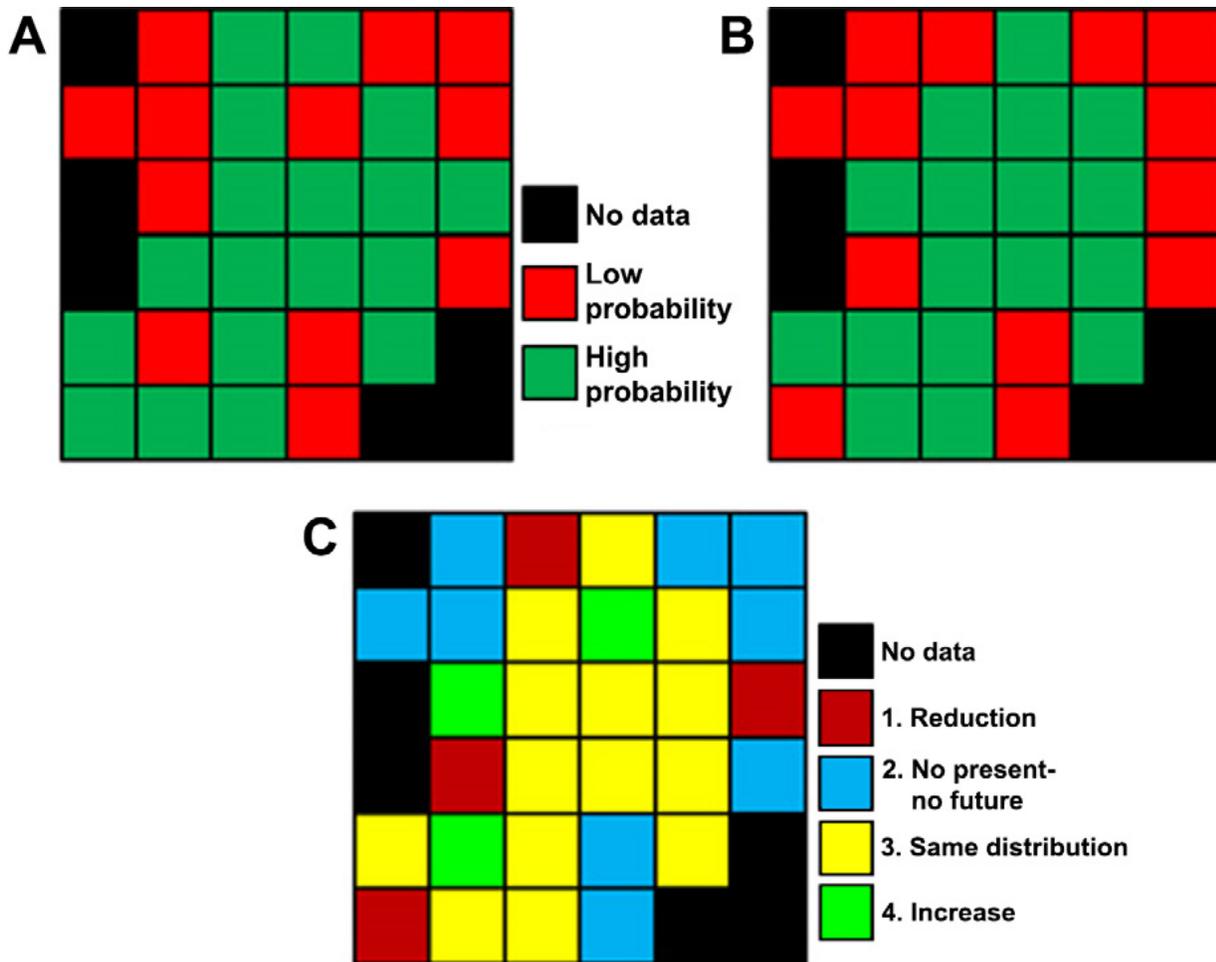


Figura 56. Comparación de dos capas de predicción, A) para condiciones de presente y B) para condiciones de futuro proyectado, de la cual se puede determinar las situaciones descritas en el apartado 16.1, para obtener una capa C) donde se clasifica cada celda de acuerdo a la situación.

16.2. Cómo funciona Mcompare

La herramienta Mcompare requiere de dos mapas de predicción de distribución, uno para presente y otro para un escenario de cambio climático a partir de variables bioclimáticas proyectadas a futuro. El usuario debe indicar a la herramienta Mcompare donde se encuentran y como se llaman los archivos que contienen los mapas. Si los mapas predictivos (presente o futuro) no tienen la proyección binarizada (ver en el capítulo 15 todo lo referente a la binarización) se debe indicar, y a continuación establecer un valor de corte para binarizar la proyección entre 0 y 1000. Para ello puede utilizar los valores de corte que proporcionan tablas de evaluación de modelos como AUC, Kappa o TSS. Adicionalmente el usuario indica el nombre del archivo de datos de presencia, el cual aporta la información necesaria para crear un mapa vectorial de puntos (shapefile) que incluye una columna llamada "Code Impac" resultado de la extracción del valor de celda del mapa de situaciones (1. reducción, 2. no presente-no futuro, 3. mantenimiento y 4. aumento) para cada sitio de presencia.

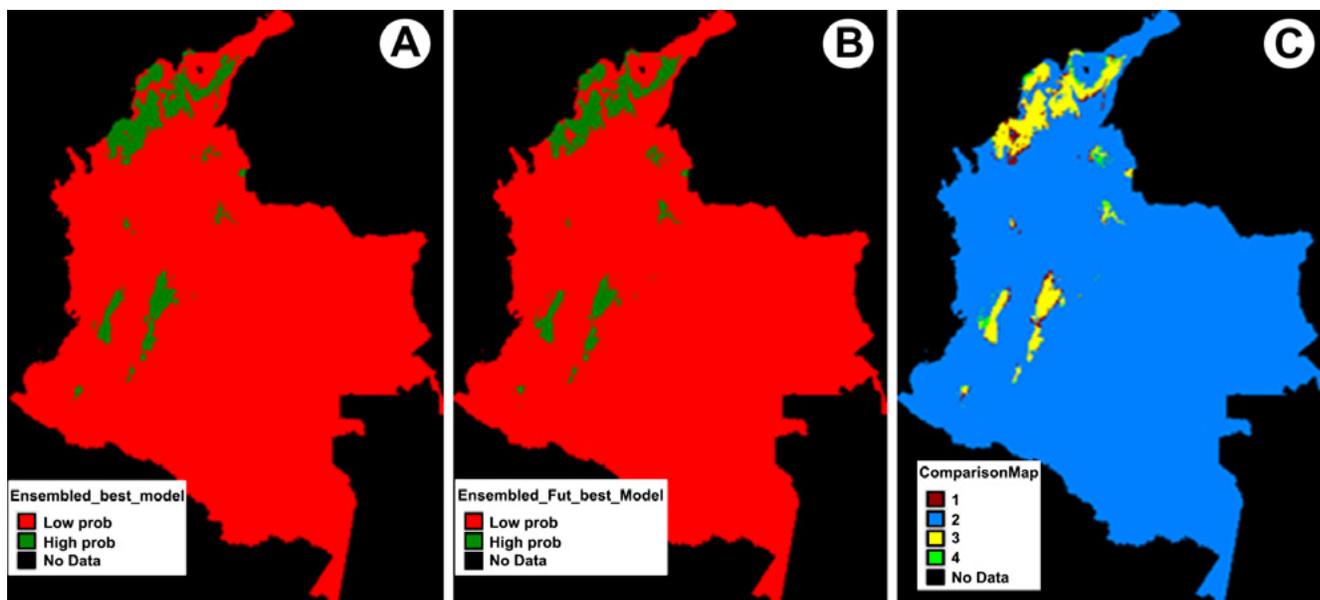


Figura 57. Resultados en forma de mapa de la herramienta Mcompare utilizando los mismos colores de la Figura 50. A) mapa de distribución potencial para condiciones de presente, B) para condiciones de futuro proyectado y C) mapa de la clasificación de las cuatro situaciones.

Para la realización de la comparación, es requisito indispensable que los mapas de predicción de la distribución potencial presente y futura deben coincidir exactamente en extensión (xmin, xmax, ymin e ymax) y tamaño de celda (1x1, 5x5, 10x10 o 20x20 km). Una mínima desviación entre los dos mapas en extensión o resolución de celdas, generará un aviso de error al momento de aplicar la herramienta.

16.3. Utilización de la herramienta Mcompare

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Mcompare, se deben especificar una serie de parámetros por parte del usuario.

16.3.1 Parámetros iniciales definidos por el usuario

16.3.1.1 Parámetro: (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

16.3.1.2 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta 'Pasaporte' pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama 'tabla.txt' y se encuentra en una carpeta llamada 'Stuberosum' que se encuentra dentro de 'Pasaporte', entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

16.3.1.3 Parámetro: geoqual

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50/51 columnas (y no las 45/46 del modelo de pasaporte básico). Para seleccionar esta opción use como tabla de entrada (parámetro 'pasaporte') la tabla resultado de GEOQUAL denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" con este u otro nombre siempre que corresponda a esta tabla. Seleccionar esta opción implica que se realizará un filtrado de datos de entrada, conservándose los sitios de recolección/presencia de mayor calidad en cuanto a su georreferenciación.

16.3.1.4 Parámetro: totalqual

Explicación: Solo aplica si geoqual ha sido señalado como TRUE (✓ en modo on server). Si su tabla de pasaporte ha sido previamente analizada con GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

16.3.1.5 Parámetro: disdup

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas. La determinación de la distancia depende de condiciones biológicas (flujo de genes) y espaciales (tamaños promedios de las poblaciones). Se trata de un parámetro específico para la especie objetivo y muchas veces será necesario consultar el concepto de un experto.

16.3.1.6 Parámetro: rutapresent (sólo aplica para modo local)

Explicación: Escriba aquí la ruta que lleva al mapa en formato ráster que contiene el modelo de distribución de la especie objetivo obtenido con datos bioclimáticos del presente. Evite que ésta ruta contenga espacios (por ejemplo 'C:/

Mis documentos’). Nota: use el signo slash (/) en la ruta antes que el signo backslash (\). Por ejemplo, debe escribir F:/ o C:/CAPFITOGEN o D:/MisHerramientas/CAPFITOGEN, etc.

16.3.1.7 Parámetro: **modelpresent**

Explicación: Escriba aquí el nombre del mapa en formato ráster que contiene el modelo de distribución de la especie objetivo obtenido con datos bioclimáticos del presente para el modo local. Se trata del archivo con extensión .tif. Al escribir el nombre no olvide añadir la extensión, por ejemplo ‘Presente_completo_Nombre_especie.tif’. Para el modo on server, seleccione el mapa y súbalo al servidor al hacer click en el botón frente al parámetro.

16.3.1.8 Parámetro: **binarizedp**

Explicación: Indique (con la opción TRUE en modo local o ✓ en modo on server) si la respuesta en el mapa de predicción de distribución de presente está binarizada (0 ausencia, 1 presencia).

16.3.1.9 Parámetro: **binarthresp**

Explicación: Sólo aplica cuando la opción ‘binarizedp’ se ha señalado como FALSE (☐ en modo on server). Escriba un número entre 1 y 999 que sirva como umbral para binarizar el mapa de predicción de distribución de presente. Sólo aplica para mapas de predicción con valores entre 0 y 1000. Valores por encima de la cifra aquí especificada, se asumirán como presencia, valores menores como ausencia.

16.3.1.10 Parámetro: **rutaigual (sólo aplica para modo local)**

Explicación: Indique con la opción TRUE si la ruta que lleva al modelo con datos bioclimáticos de presente es la misma que lleva al modelo con datos bioclimáticos de futuro.

16.3.1.11 Parámetro: **rutafuture (sólo aplica para modo local)**

Explicación: Sólo aplica cuando la opción ‘rutaigual’ se ha señalado como FALSE. Si las rutas a los modelos presente y futuro son diferentes, escriba aquí la ruta que lleva al mapa en formato ráster que contiene el modelo de distribución de la especie objetivo obtenido con datos bioclimáticos del futuro. Evite que ésta ruta contenga espacios (por ejemplo ‘C:/Mis documentos’). Nota: use el signo slash (/) en la ruta antes que el signo backslash (\). Por ejemplo, debe escribir F:/ o C:/CAPFITOGEN o D:/MisHerramientas/CAPFITOGEN, etc.

16.3.1.12 Parámetro: **modelfuture**

Explicación: Escriba aquí el nombre del mapa en formato ráster que contiene el modelo de distribución de la especie objetivo obtenido con datos bioclimáticos del presente (modo local). Se trata del archivo con extensión .tif. Al escribir el nombre no olvide añadir la extensión, por ejemplo ‘Futuro_completo_Nombre_especie.tif’. Para el modo on server, seleccione el mapa y súbalo al servidor al hacer click en el botón frente al parámetro.

16.3.1.13 Parámetro: **binarizedf**

Explicación: Indique (con la opción TRUE en modo local o ✓ en modo on server) si la respuesta en el mapa de predicción de distribución de futuro está binarizada (0 ausencia, 1 presencia).

16.3.1.14 Parámetro: **binarthresf**

Explicación: Sólo aplica cuando la opción 'binarizedf' se ha señalado como FALSE (□ en modo on server). Escriba un número entre 1 y 999 que sirva como umbral para binarizar el mapa de predicción de distribución de futuro. Sólo aplica para mapas de predicción con valores entre 0 y 1000. Valores por encima de la cifra aquí especificada, se asumirán como presencia, valores menores como ausencia.

16.3.1.15 Parámetro: **resultados**

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / antes que \. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

16.4. Resultados de Mcompare

Después de definir todos los parámetros y rutas (para modo local) que requiere Mcompare, al hacer clic en el botón "Run" (modo local en RStudio) o "Iniciar" (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, Mcompare guardará los resultados en la ruta y carpeta que se haya determinado en "resultados" (en el modo local), o en la zona de archivos de usuario y resultados en modo on server.

La tabla "Comparison_stats.txt/xls" contiene una tabla con las cuatro posibles situaciones descritas en el apartado 16.1 y la cantidad de celdas para cada situación en el mapa "ComparisonMap.tif" en la columna "Freq".

La tabla "PresenceData_Classification.txt/xls" contiene una tabla con una estructura de columnas idéntica a la tabla "Comparison_stats.txt/xls", pero en este caso la clasificación en las cuatro posibles situaciones no corresponde a las celdas del mapa "ComparisonMap.tif" sino de los sitios de presencia, representados en el mapa "Current_PresenceData_Classified.shp".

El mapa ráster "ComparisonMap.tif" (o "ComparisonMap.grd" compatible con DIVA-GIS) muestra el marco de trabajo clasificado según las cuatro posibles situaciones descritas en el apartado 16.1.

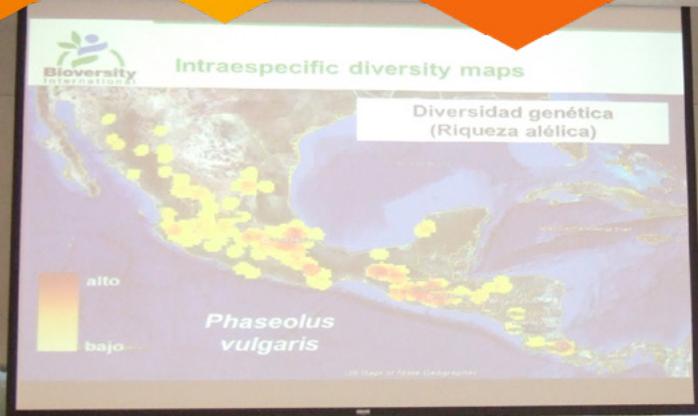
El mapa vectorial de puntos "Current_PresenceData_Classified.shp" muestra los sitios de presencia contenidos en la tabla indicada en el parámetro "pasaporte", y en la tabla asociada el shapefile se incluye una columna denominada

“Code Impac” con los siguientes códigos para cada situación: 1 = reducción, 2 = no presente-no futuro, 3 = mantenimiento y 4 = aumento.

Los dos tipos de mapas pueden ser visualizados y editados (colores y valores en las leyendas) en el software DIVA-GIS.

16.5. Referencias

Scheldeman, X., van Zonneveld, M. 2011. Manual de Capacitación en Análisis Espacial de Diversidad y Distribución de Plantas. Bioversity International, Roma, Italia.



1963

50 años de normas
promovidas
de responsabilidad
ambiental

A vertical banner on the right side of the image commemorates the 50th anniversary of environmental standards. It features the year '1963' at the top and the text '50 años de normas promovidas de responsabilidad ambiental' (50 years of promoted environmental responsibility standards) written vertically.





17 | Herramienta Tzones

17.1. Base conceptual

Las especies exhiben una amplia variación espacial y temporal en los patrones locales de adaptación genética, que están relacionados con una serie de factores bióticos y abióticos (Hanson *et al.* 2017). La adaptación de una planta se regula a partir de la variación fisiológica y genética de sus poblaciones y esta determinada por el ambiente al cual están expuestas. Varios aspectos principalmente antropogénicos están influenciando las dinámicas climáticas causando cambios en los promedios globales de temperatura, como también alterando los patrones de precipitación, con notables incrementos en los eventos climáticos extremos y en los niveles del mar (Williams y Dumroese, 2013; Harris *et al.*, 2006). La asimétrica relación entre el incremento de los fenómenos asociados al cambio climático y la capacidad de adaptación de las poblaciones de especies vegetales, enciende las alarmas sobre los riesgos sobre la permanencia de los recursos fitogenéticos *in situ*. En ese sentido, el éxito reproductivo de las especies vegetales estará limitado por la alteración de las condiciones ambientales, particularmente en las épocas de crecimiento (Chuine, 2010).

Ante este reto, las prácticas de restauración ecológica pueden representar una oportunidad para evitar la pérdida de poblaciones vegetales en el futuro. La restauración ecológica tiene como objetivo que un ecosistema que ha sido degradado, dañado o destruido por diferentes factores, como lo puede ser el cambio climático, pueda ser recuperado. Se dice que una buena restauración ecológica debe incluir una aproximación amplia que incluya factores históricos, sociales, culturales, políticos, estéticos y morales (Higgs, 1997). Sin embargo, es el tema técnico, el cual permite restauraciones exitosas, el que se debe resolver inicialmente, y es donde el componente de la diversidad genética entre y dentro de las especies resulta un asunto clave (Harris *et al.*, 2006).

Como resultado, la aplicación de prácticas de restauración ecológica representa un desafío para proteger la biodiversidad vegetal. Las plantas logran un desarrollo y comportamiento adecuados creciendo bajo condiciones ambientales específicas en las cuales ha evolucionado. Una de las prácticas de restauración de poblaciones más comunes es la introducción de germoplasma de la misma especie con el fin de reforzar la población amenazada. Si bien la introducción de individuos de mayor adaptación en poblaciones afectadas, representaría introducir genética foránea que pudiera desplazar la genética local o cruzarse con ella, esta técnica puede llegar a representar la única oportunidad para poblaciones en serio declive. Estos reforzamientos serán exitosos en la medida que la compatibilidad genética y adaptativa entre los individuos a reintroducir y la población receptora sea mayor. Por ello la transferencia de germoplasma requiere un estricto control de la compatibilidad genética y ambiental entre poblaciones fuente y destino. Esto determinó el desarrollo de aproximaciones para realizar dichas transferencias que garantizaran al menos la compatibilidad de condiciones ambientales entre población fuente y población destino (perfiles adaptativos) y los potenciales efectos del cambio climático (Potter y Hargrove, 2012).

17.2. Zonas de transferencia de semillas

Las Zonas de Transferencia de Semillas (STZ) son un enfoque metodológico que se ha desarrollado para identificar recursos genéticos de plantas silvestres con adaptaciones apropiadas que se utilizarán en la restauración de ecosistemas afectados por cambios ambientales (Havens *et al.* 2015). En principio, las STZ se basaban sólo en la compatibilidad genética entre poblaciones fuente y receptoras para garantizar un reforzamiento o restauración efectiva.

Dado que a menudo se carece de información genética sobre las especies de plantas silvestres objetivo antes de las actividades de recolección o restauración, nuevas alternativas de STZ empezaron a surgir. Así se proponen unas STZ provisionales, que en términos prácticos, son un área geográfica que comprende dos sitios, dentro de los cuales el germoplasma de una población puede ser trasladado a otra, donde se busca una alta similitud ambiental entre fuente y destino, minimizando así el riesgo de mala adaptación (Kramer y Havens 2009; Havens *et al.* 2015). El término STZ provisionales fue propuesto por Bower *et al.* (2014), quienes argumentan la necesidad de dar soluciones provisionales a la necesidad de identificar fuentes de germoplasma para la gran mayoría de casos en los cuales no se cuenta con información genética para garantizar la compatibilidad fuente-receptor, particularmente información sobre variación genética de tipo adaptativo. Adicionalmente a la falta de información genética, las STZ provisionales también se justificaron en la medida de que la información de adaptación abiótica puede reflejar indirectamente la variación genética (Peeters *et al.*, 1990).

Las aproximaciones iniciales para generar zonas de semillas tuvieron el fin de evitar o prevenir fallas en la siembra de plantas -particularmente especies forestales- datan de principios del siglo XX (Bates 1930; Fowells 1949). Las STZ provisionales se basan en supuestos sobre características ecogeográficas que son importantes en el rango de distribución de especies y son ampliamente utilizadas por quienes ejecutan programas de restauración (Omernik y Griffith 2014; Doherty *et al.*, 2017; Gibson y Nelson 2017; Germino *et al.*, 2019; Cevallos *et al.*, 2020). Algunos estudios sobre diferenciación genética y fenotípica justifican el uso de STZ provisionales y sus bases para la transferencia de semillas (Johnson *et al.*, 2004; Doherty *et al.*, 2017; Durka *et al.*, 2017).

Como antecedentes de STZ provisionales, hay evidencia en forma de publicaciones sobre lineamientos para la transferencia de germoplasma basados en el uso de mapas de ecorregiones como parámetro para delineación de zonas de similitud ambiental (Erickson *et al.* 2004; Johnson *et al.* 2010; Miller *et al.* 2011). Sin embargo, la categorización ambiental que ofrece una ecorregión puede ser demasiado amplia para detectar ciertos factores ambientales limitantes y patrones adaptativos (Parra-Quijano *et al.* 2012a). Por otro lado, la sensibilidad a los factores ambientales limitantes es diferente para cada especie, por tanto, es necesario un enfoque específico por especie, como en el estudio desarrollado por Marinoni *et al.* (2021) para la conservación de recursos fitogenéticos de especies a partir del uso de mapas de Caracterización Ecogeográfica de Tierras (ELC). Los mapas ELC se pueden crear para una especie o un grupo de especies relacionadas filogenéticamente, utilizando variables ecogeográficas estrechamente relacionadas con la distribución de la especie objetivo, donde se delimitan diferentes escenarios adaptativos dentro de un territorio determinado (García *et al.* 2017; Parra Quijano *et al.* 2012a). Estas características hacen de los mapas ELC un sistema de STZ provisional muy apropiado como lo indica Thomas *et al.* (2018).

En todo caso, un STZ de cualquier tipo debe ser validado positivamente como método para asegurar la adaptación local del germoplasma transferido, a través de técnicas como los jardines comunes (McKay *et al.*, 2005; Miller *et al.*, 2011).

17.3. Aproximación espacio-temporal en STZ.

Los primeros estudios basados en las STZ se limitaron a situaciones de tiempo presente, considerando sólo la transferencia en términos de dos localidades geográficas (Withrow-Robinson y Johnson 2006; Bower *et al.* 2014). En años

recientes, el enfoque de las STZ provisionales ha adquirido una dimensión temporal además de la dimensión espacial previamente mencionada. Esta nueva dimensión podría ser clave en relación a las condiciones de cambio climático rápido, las necesidades de material con adaptación apropiada y restauraciones ecológicas en el futuro. El enfoque espacio-temporal consiste en producir STZ provisionales en condiciones de presente y luego proyectarlas a futuro, considerando la similitud ambiental no sólo entre dos sitios en el presente sino también entre un sitio en el presente y otro en el futuro. Por lo tanto, se establecen los límites espaciales para la transferencia de semillas y se definen para diferentes períodos de tiempo (Potter y Hargrove 2012; Havens *et al.* 2015; Richardson y Chaney 2018; Shryock *et al.* 2018).

Esa dimensión adicional fue introducida por Potter and Hargrove (2012), estableciendo la correspondencia entre condiciones de presente y futuro bajo condiciones de cambio climático para 30,000 ecorregiones globales. Sin embargo, en este estudio los autores llevaron a cabo una aproximación generalista, asumiendo que un solo mapa de ecorregiones global puede ser válido para cualquier especie vegetal. Para eso, consideraron que un mapa único de ecorregiones es capaz de delinear correctamente escenarios adaptativos para un amplio rango de especies. Sin embargo, la adaptación local puede variar significativamente de especie en especie dado que los procesos por selección natural y flujo de genes no operan de manera uniforme. Esto implica que los STZ provisionales deben ser desarrollados a nivel de especie, involucrando en el delineamiento de cada zona únicamente las variables ambientales de mayor importancia respecto a la adaptación de dicha especie (Johnson *et al.* 2004; Kramer *et al.*, 2015).

17.4. Cómo funciona Tzones

La herramienta Tzones de CAPFITOGEN3 permite obtener STZ provisionales y espacio-temporales basados en mapas ELC creados con variables bioclimáticas de presente, los cuales se proyectan a futuro a través del algoritmo Random Forest y variables bioclimáticas proyectadas a potenciales futuros escenarios de cambio climático definidos por GCM (modelos globales de circulación), SSP (rutas socio-económicas compartidas) y periodos de tiempo (años).

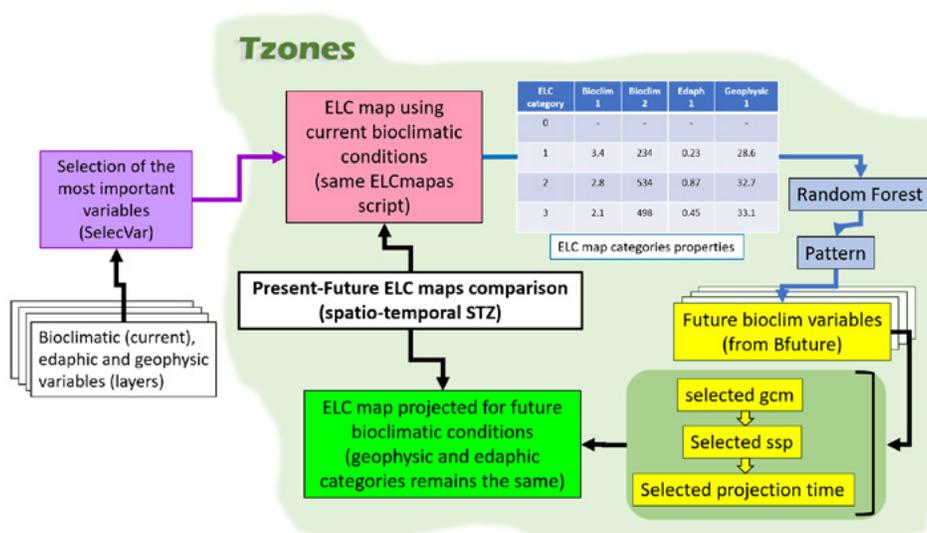


Figura 58. Procedimiento por el cual Tzones crea mapas ELC para condiciones de presente y luego los proyecta a futuro.

Adicional a la proyección del mapa ELC, Tzones determina que poblaciones podrían afectarse en el futuro debido al cambio climático, dado que la categoría de mapa ELC donde se encuentran en el presente sufre cambios bioclimáticos suficientes como para que cambie la categoría del mapa ELC proyectado. Finalmente, Tzones identifica las poblaciones que en el presente podrían servir como fuente de germoplasma para restauraciones de las poblaciones afectadas en el futuro, basado en la similitud ambiental abiótica que se asume hay entre dos sitios que perteneces a la misma categoría de un mapa ELC (ver Fig. 59 y 60).

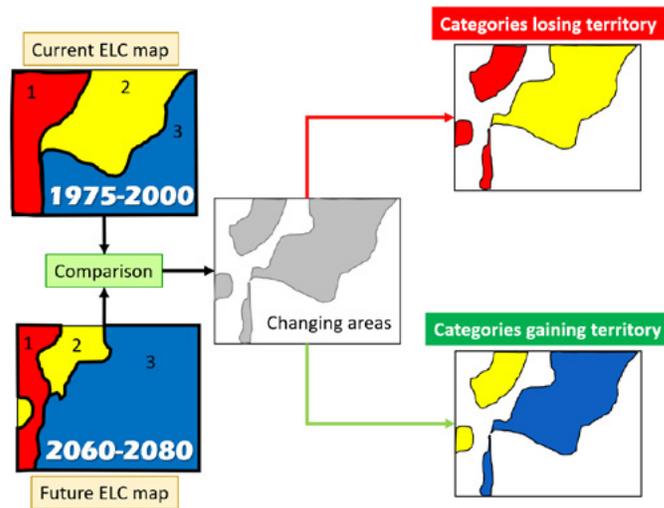


Figura 59. Ejemplo de la proyección de un mapa ELC a futuro (periodo 2070) y su comparación con el mapa de presente, la cual permite la identificación de zonas cambiantes, donde unas categorías pueden ganar territorio y otras perderlo.

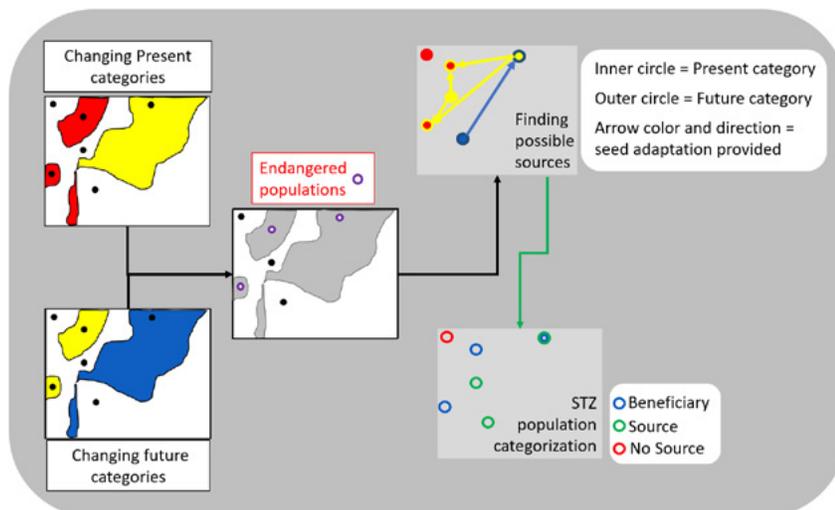


Figura 60. Resultados de la herramienta Tzones. A partir de la ocurrencia de poblaciones en áreas cambiantes en la comparación de mapas ELC presente-futuro, se identifican: a) poblaciones afectadas en el futuro que potencialmente serían beneficiarias de germoplasma, b) poblaciones fuente de germoplasma en el presente para beneficiar a las primeras en el futuro y c) poblaciones que no estarían afectadas y tampoco serían fuente de germoplasma.

17.5. Utilización de la herramienta Tzones

Una vez instaladas las herramientas CAPFITOGEN3 modo local o accedido a CAPFITOGEN3 modo on server y seleccionada la herramienta Tzones, se deben especificar una serie de parámetros por parte del usuario.

17.5.1 Parámetros iniciales definidos por el usuario

17.5.1.1 Parámetro: ruta (sólo para modo local)

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / (slash) en lugar de \ (backslash) en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

17.5.1.2 Parámetro: elcready

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si la herramienta Tzones ya ha sido ejecutada anteriormente para el mismo espacio geográfico, los mismos puntos de ocurrencia/recolección y las mismas variables ecogeográficas. Esa ejecución previa genera un archivo llamado Productos.RData en la carpeta de “resultados”. Si el usuario desea usar en el futuro el mismo mapa ELC que generó anteriormente, pero proyectándolo a otros escenarios de futuro, no es necesario volver a generar el mapa de presente, simplemente debe marcar esta opción como TRUE o ✓. Por tanto, siempre que el usuario vaya a utilizar por primera vez Tzones para un espacio, un conjunto de variables y unas presencias/sitios de recolección, el parámetro elcready debe ser FALSE o □.

17.5.1.3 Parámetro: pais

Explicación: Seleccione el país/región para el cual quiere construir el mapa ELC. En la lista desplegable dispone de más de 160 países, algunas coberturas a nivel subcontinental, a nivel global y coberturas personalizadas (provenientes de la herramienta rLayer).

17.5.1.4 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la generación del mapa. Note que 1x1 km ofrece mayor resolución, pero exige mucha más capacidad de cómputo y tarda mucho más que 5x5 km, especialmente en países de gran extensión. Mapas de alta resolución (i.e. 1x1 km) compuestos de más de 100,000 celdas (países de más de 100,000 km²) pueden generar problemas de procesamiento para algunos métodos de agrupamiento y determinación de número óptimo de grupos como “cluster”.

17.5.1.5 Parámetro: bioclimv

Explicación: Liste (modo local) o seleccione (modo on server) las variables bioclimáticas que desea analizar. Para

hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables bioclimáticas), cópielos y péguelos separados por un signo de punto y coma (;). También encontrará una línea bloqueada la selección de las 19 variables bioclim, listas para usar, simplemente eliminando el símbolo # inicial. En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro bioclimv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

17.5.1.6 Parámetro: edaphv

Explicación: Liste (modo local) o seleccione (modo on server) las variables edáficas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables edáficas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro edaphv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

17.5.1.7 Parámetro: geophysv

Explicación: Liste (modo local) o seleccione (modo on server) las variables geofísicas que desea analizar. Para hacer selección múltiple de variables en modo local, busque los nombres de las variables en el listado completo que aparece en el script de parámetros, en una línea bloqueada con # (anunciada una línea antes así: “#Lista completa de variables geofísicas), cópielos y péguelos separados por un signo de punto y coma (;). En modo on server, simplemente seleccione cada variable de interés haciendo click en su nombre. Una vez se selecciona, el nombre de la variable aparecerá en la caja frente al parámetro geophysv. Para conocer los códigos, nombres y breves descripciones de las variables, consultar el archivo “Variables names - Nombres de variables.xlsx” descargable desde: <http://www.capfitogen.net/es/wp-content/uploads/Variables-names-Nombres-de-variables.xlsx> y también disponible en la instalación de CAPFITOGEN3 modo local.

17.5.1.8 Parámetro: latitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la latitud (Y) como una variable geofísica más a ser analizada.

17.5.1.9 Parámetro: longitud

Explicación: Al indicar TRUE (modo local) o ✓ (modo on server) se incluirá la longitud (X) como una variable geofísica más a ser analizada.

17.5.1.10 Parámetro: ssp (anterior rcp)

Explicación: Seleccione la vía de concentración representativa (RCP en inglés) de la información proyectada a futuro que tiene disponible para el país/región (parámetro país) en la carpeta “rdatamapsf”. RCP es en otras palabras, el escenario de concentración de gases de invernadero para sus datos bioclimáticos de futuro. Hay cuatro RCP disponibles de acuerdo con el IPCC: RCP2.6 (aquí señalado como 26), RCP4.5 (45), RCP6 (60) y RCP8.5 (85). Estos valores tienen que ver con los valores radiativos (+2.6, +4.5, +6.0, and +8.5 W/m², respectivamente) para el año 2100 en relación a los valores de la era pre-industrial. Acerca de SSP, hay cuatro disponibles: 126 (ssp1, 2.6), 245 (ssp2, 4.5), 370 (ssp3, 7.0) y 585 (ssp5, 8.5).

17.5.1.11 Parámetro: gcm

Explicación: Sólo aplica si temp=future. Seleccione el modelo de clima global (GCM en inglés, también conocidos como modelos generales de circulación) de la información proyectada a futuro que tiene disponible para el país/región (parámetro país) en la carpeta “rdatamapsf”. Estos modelos simulan el clima dependiendo en parte de la concentración de gases de invernadero (rcp/ssp). Si desea saber más de los GCM, puede visitar <http://goo.gl/4XhU6g> o <http://goo.gl/2VUuRP>.

17.5.1.12 Parámetro: proy

Explicación: Seleccione el periodo de tiempo para el cual desea proyectar los mapas ELC de presente. La opción 50 representa el valor/año 2050 (promedio entre 2014 y 2060) y la opción 70 representa el valor/año 2070 (promedio entre 2061 y 2080). El usuario debe estar seguro que cuenta con la información bioclimática proyectada a futuro para los parámetros ssp, gcm y proy. Esta información ya adecuada al marco de trabajo (señalado en el parámetro “país”), debe estar en la carpeta rdatamapsf (en la versión local). Considere revisar la función y proceso asociado a la herramienta Bfuture para poder realizar una configuración correcta de estos tres parámetros (ssp, gcm y proy).

17.5.1.13 Parámetro: maxg

Explicación: Indique el número máximo de agrupamientos por componente (bioclimático, geofísico y edáfico) que desea permitir (a mayor número, mapas con más categorías). Se recomienda un valor menor o igual a cinco, de lo contrario se pueden llegar a formar mapas ELC de más de 125 categorías.

17.5.1.14 Parámetro: metodo

Explicación: Seleccione el método para generar los agrupamientos con determinación objetiva del número óptimo de grupos. Encontrará disponibles los seis métodos descritos en este capítulo por los términos: “kmeansbic”, “medoides”, “elbow”, “calinski”, “ssi” y “bic”

17.5.1.15 Parámetro: iterat

Explicación: Sólo aplica si para el parámetro “metodo” se ha seleccionado la opción “calinski” o “ssi”. En este campo indique el número de permutaciones usadas para generar agrupamientos K-means sobre las cuales se calcula los criterios “calinski” o “ssi”.

17.5.1.16 Parámetro: iteratf

Explicación: En este parámetro se indica el número de iteraciones o permutaciones a ser usado por el algoritmo Random Forest con el que se proyecta el mapa ELC presente con condiciones/variables bioclimáticas de futuro bajo un SSP, GCM y periodo (años) futuro.

17.5.1.17 Parámetro: transfer

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si desea realizar un análisis de zonas de transferencia de germoplasma. Este procedimiento asume que poblaciones ocurriendo en categorías cambiantes entre mapas ELC presente y futuro, se encontrarían en riesgo de mal adaptación, y que podrían requerir reforzamientos a través de introducción de germoplasma de poblaciones que en el presente se ocurren en las categorías ELC de futuro de las poblaciones afectadas. Se requiere disponer de una tabla con datos de presencia (su nombre se debe indicar en el parámetro “pasaporte”), que contiene poblaciones que podrían estar en riesgo de desaparecer por cambio climático y poblaciones que podrían servir de fuente de germoplasma en el presente para evitar su desaparición.

17.5.1.18 Parámetro: pasaporte

Explicación: Para el modo local, escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama ‘tabla’ debe escribir: ‘tabla.txt’. Recuerde que este archivo debe ser guardado previamente en la carpeta ‘Pasaporte’ que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Sus datos dentro de la carpeta ‘Pasaporte’ pueden estar dentro de subcarpetas, las cuales deben aparecer reflejadas en este campo. Por ejemplo, si su tabla se llama ‘tabla.txt’ y se encuentra en una carpeta llamada ‘Stuberosum’ que se encuentra dentro de ‘Pasaporte’, entonces en este campo debe aparecer Stuberosum/tabla.txt, usando siempre / (slash) antes que \ (backslash) en la descripción de la ruta. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados. Para el modo on server sólo debe subir el archivo .txt o seleccionarlo de lo que se hayan subido previamente y esté en la zona de archivos de usuario y resultados.

17.5.1.19 Parámetro: tecogaps

Explicación: Seleccione esta opción (TRUE en modo local o ✓ en modo on server) si su tabla de pasaporte tiene la columna “GAPTYPE” y ella contiene valores válidos (ha sido procesada por la herramienta Representa) que sirvan para filtrar aquellas presencias correspondientes a faltantes ecogeográficos prioritarios. Utilizar los faltantes detectados

en un uso previo de representa, Tzones señalaría como prioritarias para recolección las poblaciones que son fuente y a la vez faltantes ecogeográficos y espaciales.

17.5.1.20 Parámetro: **ttresh**

Explicación: Sólo es importante si tecogaps es TRUE en modo local o ✓ en modo on server. Indique para el parámetro ttresh un dígito que establecerá el valor de umbral para GAPTYPE (de la escala de 1 a 15 que genera la herramienta Representa) para determinar cuáles presencias se consideran faltantes ecogeográficos prioritarios. Este valor suele establecerse en 4 (ver capítulo para la herramienta Representa).

17.5.1.21 Parámetro: **resultados (sólo para modo local)**

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc. En esta carpeta se guardarán los mapas ELC de presente.

17.5.1.22 Parámetro: **resultados2 (sólo para modo local)**

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis a futuro. Nota: use / (slash) antes que \ (backslash). Por ejemplo C:/Resultados2, D:/MisHerramientas/Resultados2. La ruta aquí indicada debe ser diferente a la ruta del parámetro “resultados”.

17.6. Resultados de Tzones

Después de definir todos los parámetros y rutas (para modo local) que requiere Tzones, al hacer clic en el botón “Run” (modo local en RStudio) o “Iniciar” (modo on server), se iniciará el proceso de análisis de la herramienta.

Luego de un tiempo que puede variar en función del tipo de análisis solicitado, Tzones guardará los resultados en las rutas y carpetas que se hayan determinado en “resultados” y “resultados2” (en el modo local), o en la zona de archivos de usuario y resultados en modo on server.

17.6.1 Carpeta indicada en el parámetro “resultados”

Esta carpeta contiene los mismos productos de la ejecución de la herramienta ELCmapas, con la adición de un archivo, Producto.RData, en el cual se guardan los elementos necesarios de este mapa ELC (creado con variables bioclimáticas de presente) para futuras variaciones en su proyección a futuro.

17.6.2 Carpeta indicada en el parámetro “resultados2”

En esta carpeta se guardarán todos los resultados referentes a la proyección del mapa ELC bajo ciertos escenarios de futuro y cambio climático, además del análisis de los cambios entre el mapa ELC presente y futuro, los cambios de categoría para los sitios de ocurrencia/recolección, y por tanto, el sistema de fuentes de semillas – poblaciones receptoras o beneficiarias (zonas de transferencia de semillas espacio-temporales).

17.6.2.1 Tabla *TransferZoneAnalysis.txt/xls*: En esta tabla se verifica para cada presencia/recolección si la categoría ELC ha cambiado o se mantiene igual, identificando las poblaciones amenazadas por dicho cambio y las posibles fuentes de germoplasma para restauraciones a futuro. La columna “PopEndangered” en particular identifica las poblaciones para las cuales cambiará la categoría ELC (indicando el número de la categoría sobre la cual crece en el presente esta población) y la columna “Sources” indica con valor 1, aquellas poblaciones que podrían ser recolectadas en el presente para restaurar poblaciones amenazadas en el futuro.

17.6.2.2 Tabla *stats_change_table.txt/xls*: Esta tabla muestra para cada categoría del mapa ELC, el porcentaje de cambio (en términos de las celdas que cambian sobre el total de celdas de cada categoría) respecto al mapa ELC proyectado a futuro (columna “Freq”).

17.6.2.3 Tabla *stats_change_table.txt/xls*: Esta tabla muestra las características del mapa ELC proyectado a futuro en términos de cada variable ecogeográfica usada para generar el mapa ELC de presente.

17.6.2.4 Mapa *mapa_sspXXX_elcF_pais.tif/grd*: Corresponde al mapa ELC de futuro (proyectado).

17.6.2.5 Mapa *mapa_sspXXX_elcF_DIVA_pais.grd*: Es una copia del mapa del numeral anterior. Esta copia se genera porque al abrir estos mapas en DIVA-GIS, este software cambia algunas características del mismo y puede ser de interés mantener una copia original.

17.6.2.6 Mapa *mapa_sspXXX_bioclimaticoF_pais.tif/grd*: Este mapa corresponde a la proyección del mapa del componente bioclimático de presente, usada para combinarse con el mapa geofísico y el edáfico y así generar el mapa ELC.

17.6.2.7 Mapa *map_change_pres_pais.tif/grd* y *map_change_fut_pais.tif/grd*: En estos mapas se muestran en exclusiva las categorías cambiantes, resultado de la comparación entre los mapas ELC de presente y futuro. Ambos mapas contienen las mismas áreas cambiantes, sólo con la diferencia que el mapa “map_change_pres_pais” muestra las áreas con los valores que tienen en presente y “map_change_fut_pais” los valores que tienen en el mapa de futuro.

17.7. Referencias

Bates, C. G. 1930. The frost hardiness of geographic strains of Norway pine. *Journal of Forestry*, 28(3), 327-333.

Bower, A. D., Clair, J. B. S., Erickson, V. 2014. Generalized provisional seed zones for native plants. *Ecological Applications*, 24(5), 913-919.

Cevallos, D., Bede-Fazekas, Á., Tanács, E., Szitár, K., Halassy, M., Kövendi-Jakó, A., Török, K. 2020. Seed transfer zones based on environmental variables better reflect variability in vegetation than administrative units: evidence from Hungary. *Restoration Ecology*, 28(4), 911-918.

Chuine, I. 2010. Why does phenology drive species distribution? *Philosophical transactions of the royal society B*. 365, 3149-3160

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. J. 2007. Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.

Dinerstein, E. D., Olson, A., Joshi, C., Vynne, N., Burgess, E., Wikramanayake, N., Hahn, S., Palminteri, P., Hedao, R., Noss, M., Hansen, H., Locke, E., Ellis, B., Jones, C., Barber, V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baille, J., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., Souza, N., Pintea, L., Brito, J., Llewellyn, O., Miller, A., Patzelt, A., Ghazanfar, S., Timberlake, J., Klozer, H., Shenan-Farpón, Y., Kindt, R., Barnekow, J., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K., Saleem, M. 2017. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534-545.

Doherty, K. D., Butterfield, B. J., Wood, T. E. 2017. Matching seed to site by climate similarity: techniques to prioritize plant materials development and use in restoration. *Ecological Applications*, 27(3), 1010-1023.

Durka, W., Michalski, S. G., Berendzen, K. W., Bossdorf, O., Bucharova, A., Hermann, J. M., Holzel, N., Kollmann, J. 2017. Genetic differentiation within multiple common grassland plants supports seed transfer zones for ecological restoration. *Journal of Applied Ecology*, 54(1), 116-126.

Erickson, V. J., Mandel, N. L., Sorensen, F. C. 2004. Landscape patterns of phenotypic variation and population structuring in a selfing grass, *Elymus glaucus* (blue wildrye). *Canadian Journal of Botany*, 82(12), 1776-1789.

Feeley, K. J., Silman, M. R. 2009. Extinction risks of Amazonian plant species. *Proceedings of the National Academy of Sciences*, 106(30), 12382-12387.

Fowells, H. A. 1949. Cork oak planting tests in California. *Journal of Forestry*, 47(5), 357-365.

García, R. M., Parra-Quijano, M., Iriondo, J. M. 2017. A multispecies collecting strategy for crop wild relatives based on complementary areas with a high density of ecogeographical gaps. *Crop Science*, 57(3), 1059-1069.

Germino, M. J., Moser, A. M., Sands, A. R. 2019. Adaptive variation, including local adaptation, requires decades to become evident in common gardens. *Ecological Applications*, 29(2), e01842.

Gibson, A., Nelson, C. R. 2017. Comparing provisional seed transfer zone strategies for a commonly seeded grass, *Pseudoroegneria spicata*. *Natural Areas Journal*, 37(2), 188-199.

Hamann, A., Gylander, T., Chen, P. Y. 2011. Developing seed zones and transfer guidelines with multivariate regression trees. *Tree Genetics & Genomes*, 7(2), 399-408.

Hanson, J. O., Rhodes, J. R., Riginos, C., Fuller, R. A. 2017. Environmental and geographic variables are effective surrogates for genetic variation in conservation planning. *Proceedings of the National Academy of Sciences*, 114(48), 12755-12760.

Harris, J. A., Hobbs, R. J., Higgs, E., Aronson, J. 2006. Ecological Restoration and Global Climate Change. *Restoration Ecology*, 14 (2), 170–176.

Havens, K., Vitt, P., Still, S., Kramer, A. T., Fant, J. B., Schatz, K. 2015. Seed sourcing for restoration in an era of climate change. *Natural Areas Journal*, 35(1), 122-133.

Higgs, E. S. 1997. What is Good Ecological Restoration?. *Conservation biology*, 11(2), 338-348.

IPCC. 2013. IPCC Fifth Assessment Report (AR5). IPCC s. 10-12

Johnson, G. R., Sorensen, F. C., St Clair, J. B., Cronn, R. C. 2004. Pacific northwest Forest tree seed zones a template for native plants?. *Native Plants Journal*, 5(2), 131-140.

Johnson, R. C., Erickson, V. J., Mandel, N. L., St Clair, J. B., Vance-Borland, K. W. 2010. Mapping genetic variation and seed zones for *Bromus carinatus* in the Blue Mountains of eastern Oregon, USA. *Botany*, 88(8), 725-736.

Jombart, T., Collins, C. 2015. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.

Kramer, A. T., Havens, K. 2009. Plant conservation genetics in a changing world. *Trends in plant science*, 14(11), 599-607.

Kramer, A. T., Larkin, D. J., Fant, J. B. 2015. Assessing potential seed transfer zones for five forb species from the Great Basin Floristic Region, USA. *Natural Areas Journal*, 35(1), 174-188.

McKay, J. K., Christian, C. E., Harrison, S., Rice, K. J. 2005. “How local is local?”—a review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, 13(3), 432-440.

Marinoni, L., Parra Quijano, M., Zabala, J.M., Pensiero, J.F., Iriondo, J.M. 2021. Spatio-temporal seed transfer zones as an efficient restoration strategy in response to climate change. *Ecosphere*, in press.

Miller, S. A., Bartow, A., Gisler, M., Ward, K., Young, A. S., Kaye, T. N. 2011. Can an ecoregion serve as a seed transfer zone? Evidence from a common garden study with five native species. *Restoration Ecology*, 19(201), 268-276.

Omernik, J. M., Griffith, G. E. 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environmental management*, 54(6), 1249-1266.

Parra Quijano, M., Iriondo, J. M. Torres, E. 2012A. Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genetic Resources Crop Evolution*, 59, 205-217

Parra Quijano, M., Iriondo, J. M., Torres, E. 2012B. Applications of ecogeography and geographic information systems in conservation and utilization of plant genetic resources. *Spanish Journal of Agricultural Research*, 10(2), 419-429.

Parra Quijano, M., Iriondo, J.M., Frese, L., Torres, E. 2012C. Spatial and ecogeographic approaches for selecting genetic reserves in Europe. In: N. Maxted, M.E. Dulloo, B.V. Ford-Lloyd, L. Frese, J. Iriondo y M.A.A. Pinheiro de Carvalho (ed.) *Agrobiodiversity Conservation: securing the diversity of crop wild relatives and landraces*. CABI, Wallingford, UK.

Parra Quijano, M. 2016. Tools CAPFITOGEN Program to Strengthen Capabilities in National Plant Genetic Resources Programs in Latin America.

Peeters, J. P., Wilkes, H. G., Galwey, N. W. 1990. The use of ecogeographical data in the exploitation of variation from gene banks. *Theoretical and applied genetics*, 80(1), 110-112.

Potter, K. M., Hargrove, W. W. 2012. Determining suitable locations for seed transfer under climate change: a global quantitative method. *New Forests*, 43, 581-599.

Richardson, B. A., Chaney, L. 2018. Climate-based seed transfer of a widespread shrub: population shifts, restoration strategies, and the trailing edge. *Ecological Applications*, 28(8), 2165-2174.

Shryock, D., Defalco, L. A., Esque, T. C. 2018. Spatial decision-support tools to guide restoration and seed-sourcing in the Desert Southwest. *Ecosphere*, 9 (10), 1-19

Thomas, E., Alcazar, C., Moscoso L. G., Vásquez A., Osorio L. F., Salgado-Negrete, B., Gonzalez, M., Parra-Quijano, M., Bozzano, M., Loo, J., Jalonen, R., Ramírez, W. 2017. The importance of species selection and seed sourcing in forest restoration for enhancing adaptive capacity to climate change: Colombian tropical dry forest as a model. *The Lima declaration on biodiversity and climate change: contributions from science to policy for sustainable development*, (89), 122-132

Williams, M. I., Dumroese, R. K. 2013. Preparing for Climate Change: Forestry and Assisted Migration. *Journal of Forestry*, 111 (4), 287-297

Withrow-Robinson, B. A., Johnson, R. 2006. Selecting native plant materials for restoration projects: ensuring local adaptation and maintaining genetic diversity Oregon State University. URL: <https://ir.library.oregonstate.edu/downloads/g732d9349>





18 | Errores Frecuentes

En la siguiente lista se encuentran muchos de los avisos de error, que en modo local o en modo on server se pueden encontrar, en la ventana de RStudio en color rojo para el modo local (exactamente en la ventana de RStudio que muestra la Fig.2-C), o como un archivo de texto en la carpeta donde se almacenan los resultados en la zona de archivos de usuario y resultados en el modo on server.

- **Aviso de error:** An error occurred: Error en readChar(con, 5L, useBytes = TRUE) : no se puede abrir la conexión-
Calls: source -> withVisible -> eval -> eval -> load -> readChar

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Se puede presentar en cualquier herramienta

Solución(es): Este error, el más común de todos suele corresponder a introducciones o indicaciones erradas de parámetros. A veces alguna parte del enunciado cambia (de acuerdo a la herramienta o en qué parte del script se produce), pero siempre tiene en común la parte que dice: no se puede abrir la conexiónCalls Por ejemplo, en ELCmaps, si a un país como Cuba se indica a la herramienta una resolución de celda de 10x10 km, puede aparecer este error (la herramienta va a buscar esta resolución en CAPFITOGEN3/rdatamaps/cuba/, y no la va a encontrar, porque no existe). También cuando se indica una ruta equivocada donde se encuentra las herramientas o las tablas de pasaporte, etc. Para evitar este problema, revise parámetro por parámetro, confirmando que los valores son los correctos.

- **Aviso de error:** An error occurred: Error en library(package name) : there is no package called 'package name'
Calls: source -> withVisible -> eval -> eval -> library

Modo en el que se puede producir: Local

Herramienta en la que se produce: Cualquiera, pues proviene de la instalación
Solución(es): Este error indica que algún paquete de R que requiere la herramienta no fue correctamente instalado y por esa razón R no lo encuentra. Asegúrese que la estructura de carpetas y archivos de las herramientas no se encuentra localizada en la raíz del directorio (por ejemplo en K:/). Si es así, cree una carpeta (usualmente con el nombre de CAPFITOGEN3) en la raíz del directorio, corte toda la estructura de carpetas y archivos y péguela en la carpeta que acaba de crear. A continuación, reinstale las herramientas. Si esta opción no funciona, pruebe a instalar el paquete manualmente. Para ello:

- a) En el código del error aparece el nombre del paquete no instalado, exactamente donde en el ejemplo dice 'package name'. Con este nombre de paquete, abra la carpeta "packages" dentro de la estructura de carpetas y archivos de CAPFITOGEN3 y allí encontrará una serie de archivos .zip con diversos nombres. Uno de ellos corresponderá con el nombre del paquete y estará acompañado de unos números relativos a la versión. Copie el nombre del archivo (completo, incluyendo la extensión .zip).
- b) Abra RStudio y teclee la siguiente instrucción en la "R console": `Install.packages("X:/CAPFITOGEN3/packages/nombreadchivopaquete.zip")` donde X corresponde a la letra de la unidad donde se encuentran las herramientas CAPFITOGEN3 (cambie esta letra con la que sea acorde a su caso) y en la parte "nombreadchivopaquete.zip" pegue el nombre del archivo que copió en el paso 1. A continuación teclee "enter".
- c) El programa empezará a instalar el paquete y cuando acabe aparecerá un aviso similar a este: `package 'sp' successfully unpacked and MD5 sums checked`
- d) Pruebe que el paquete quedó correctamente instalado, para ello teclee a continuación: `library("package name")` "package name" es el nombre del paquete y como aparece en el aviso de error (sin códigos de

versiones ni extensión .zip). Posteriormente pulse “enter”. Aparecerá un aviso similar a este: Mensajes de aviso perdidos package ‘cluster’ was built under R version 2.15.3

- e) Pruebe de nuevo la herramienta. Es posible que aparezca el mismo error pero para un paquete diferente. Si es así, repita la operación hasta que cesen los avisos de error. Este tipo de errores no son usuales desde que se mejoró el sistema de instalación, pero ocasionalmente se presentan, especialmente en la versión de Windows 8.

- **Aviso de error:** An error occurred: Error: no se puede ubicar un vector de tamaño X.X Gb

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Puede ocurrir en varias

Solución(es): Es un problema relacionado con el tamaño de las matrices que tiene que manejar R. Suele solucionarse rebajando resoluciones de los mapas. Si se presenta en ELC mapas, se recomienda cambiar el método de determinación de número óptimo de grupos o incrementar el tamaño de celda en el parámetro resol1. También es posible que este error aparezca en GEOQUAL por un error en el contenido de las tablas de pasaporte, específicamente cuando ocurren duplicados para el campo ACCENUMB. Este campo identifica inequívocamente cada entrada y un solo duplicado puede generar este aviso de error. La solución está en revisar en la tabla que no ocurran duplicados y si ocurren, asignar números o códigos únicos para cada entrada.

- **Aviso de error:** An error occurred: Error en sample.int(m, k) : primer argumento inválidoCalls: source ... withVisible -> eval -> eval -> kmeans -> sample.int

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: ELC mapas

Solución(es): Significa que alguna variable para esa región o país es constante y al estandarizarse produce una tabla de 0 filas que luego genera el error en Kmeans (método elbow). Se soluciona deseleccionando esa variable problemática. Para identificarla, es importante saber que suele corresponder a variables de suelo especialmente en países pequeños. Por ejemplo, la variable “profundidad” suele originar este problema. También en países secos, usar variables de mínimas precipitaciones suele producir este problema

- **Aviso de error:** An error occurred: Error en clara(sdata, k, ...) : x is not a numeric dataframe or matrix.Calls: source -> withVisible -> eval -> eval -> pamk -> clara

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: ELC mapas

Solución(es): Significa que alguna variable para esa zona es constante y al estandarizarse produce una tabla de 0 filas que luego genera el error en medoides. Igual solución que error No. 2.

- **Aviso de error:** An error occurred: Error en kmeans(edaph[, -1], centers = i) : more cluster centers than distinct data points.Calls: source -> withVisible -> eval -> eval -> kmeans

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: ELC mapas

Solución(es): Significa que se ha indicado un número máximo de grupos menor al número objetivo que el método elbow determina como óptimo. Repetir operación con un número menor de grupos.

- **Aviso de error:** An error occurred: Error: objeto 'ecogeot' no encontrado
Modo en el que se puede producir: Local y on server
Herramienta en la que se produce: ECOGEO
Solución(es): Se debe seleccionar la opción geophyssv si se seleccionan variables geofísicas.
- **Aviso de error:** An error occurred: Error en validObject(Object) : invalid class "SpatialPoints" object: bbox should never contain infinite valuesCalls: source ... SpatialPoints -> new -> initialize -> initialize -> validObject
Modo en el que se puede producir: Local y on server
Herramienta en la que se produce: Representa
Solución(es): Revise el archivo de texto "process_info.txt" de la carpeta "Error" de la estructura de carpetas y archivos de las herramientas CAPFITOGEN. En la línea inferior de ese archivo de texto encontrará que puede aparecer un aviso diciendo "ATENCIÓN!!, error al eliminarse todos los registros FE al considerar datos de otros bancos como no faltantes". Esto indica que la herramienta se ha quedado sin datos de fuentes externas pues todos los aportados provienen de "bancos de germoplasma" y al indicarse que éstos no deben tomarse como faltantes, se produce un error, al no quedar datos por analizar. Elimine la opción de utilizar fuentes externas o permita que Representa tome datos de otros bancos como faltantes.
- **Aviso de error:** An error occurred: Error en dist(x[ss[[i]],], method = metric, ...) : no se permiten vectores de longitud negativaCalls: source ... withVisible -> eval -> eval -> pamk -> distcritmulti -> dist
Modo en el que se puede producir: Local y on server
Herramienta en la que se produce: ELC mapas
Solución(es): Aparece este error cuando el país o región es muy grande, la resolución es alta (un tamaño de celda menor) y se pide a la herramienta realizar la determinación del número óptimo de agrupamientos "medoides". La primera solución aplicable es ejecutar de nuevo el análisis pero usando el método "elbow". Si a pesar de esto, se genera otro error, la recomendación es usar una resolución menor (tamaño de celda mayor).
- **Aviso de error:** An error occurred: Error en merge.data.frame(as.data.frame(x), as.data.frame(y), ...) : no se permiten vectores de longitud negativaCalls: source ... merge -> merge.default -> merge -> merge.data.frame
Modo en el que se puede producir: Local y on server
Herramienta en la que se produce: ELC mapas
Solución(es): El error persiste pues las matrices que se generan son tan grandes que el método de determinación óptima de número de agrupamientos "elbow" tampoco es capaz de manejar ese tamaño de matrices. La solución ya pasa por usar una resolución menor (un tamaño de celda mayor).
- **Aviso de error:** An error occurred: Error en .checkNumericCoerce2double(obj) : cannot retrieve coordinates from non-numeric elementsCalls: source ... coordinates -> .local -> do.call -> .checkNumericCoerce2double
Modo en el que se puede producir: Local y on server
Herramienta en la que se produce: GEOQUAL
Solución(es): Error en la codificación de las coordenadas o en la elaboración de la tabla de pasaporte. Para el primer caso, corrija manualmente en Excel y vuelva a salvar en texto delimitado por tabulaciones. Para el segundo caso, el orden de las variables no es el indicado y por eso las columnas correspondientes a las coord-

nadas se encuentran desplazadas. Ceñir estrictamente el orden de las variables al formato indicado y no añada columnas ni cambie su orden.

- **Aviso de error:** An error occurred: Error en `apply(x, 2, fun2) : dim(X) must have a positive length`Calls: source ... extract -> .xyValues -> .xyvBuf -> lapply -> FUN -> apply

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Se puede presentar en herramientas con función de extracción radial

Solución(es): Este error puede darse cuando el usuario solicita una extracción radial utilizando un radio demasiado pequeño (parámetro `tamp`) respecto al tamaño de celda o resolución de las variables ecogeográficas (parámetro `resol1`). Por ejemplo, solicitar una extracción radial de 1 km usando resoluciones de celda de 10x10 km aprox (5 arc-min). Una situación así produce valores nulos de extracción que luego generan este error. Pruebe con radios de tamaño mayor, especialmente mayores al tamaño de lado de cada celda y/o intente con una resolución mayor (por ejemplo, para una extracción radial de 1 km pasar de “Celdas 5x5 km aprox (2.5 arc-min)” a “Celdas 1x1 km aprox (30 arc-seg)” podría solucionar el problema). Si aun así el problema no se soluciona, sería recomendable usar extracciones puntuales.

- **Aviso de error:** An error occurred: Error en ``colnames<-`>(*tmp*, value = "ACCENUMB") : el atributo 'names' [1] debe tener la misma longitud que el vector [0]`Calls: source -> withVisible -> eval -> eval -> colnames<-

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Se puede presentar en herramientas en donde el usuario ingrese datos de pasaporte.

Solución(es): Este aviso de error puede producirse cuando el usuario indica una tabla de pasaporte en el parámetro `pasaporte` que no tiene el número de columnas esperado, debido a se haya eliminado accidentalmente alguna columna, o que la herramienta espera columnas adicionales no incluidas (como es el caso de `ColNucleo` que espera la columna adicional `“AVAILAB”`) o que se indica en el parámetro `“geoqual”` que la tabla contiene las cuatro columnas adicionales producto del análisis de la herramienta `GEOQUAL` y en realidad no las tiene. Cerciórese del contenido de la tabla pasaporte que ingresa y acorde a ello, usar el parámetro `“geoqual”`.

- **Aviso de error:** An error occurred: Error en `if (any(puntosorig$DECLATITUDE >= 90 | puntosorig$DECLATITUDE <= : valor ausente donde TRUE/FALSE es necesario`Calls: source -> withVisible -> eval -> eval ó An error occurred: Error en `if (any(puntosorig$DECLONGITUDE >= 180 | puntosorig$DECLONGITUDE <= : valor ausente donde TRUE/FALSE es necesario`Calls: source -> withVisible -> eval -> eval

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Se puede presentar en herramientas en donde el usuario ingrese datos de pasaporte

Solución(es): Existe un error en las coordenadas de al menos una entrada, y esto puede deberse a errores mismos de la codificación de las coordenadas o a que incluso no haya valor alguno (incluyendo NA) en los campos de las coordenadas. Para solucionar el problema en el primer caso, verifique la codificación de las coordenadas en sexagesimal, que concuerden con el formato FAO/Biodiversity 2012 y para decimal, que los valores para `DECLATITUDE` estén entre -90 y 90 y para `DECLONGITUDE` entre -180 y 180. Si el error se debe al segundo caso (campos vacíos), esto puede deberse a la aparición de entradas “fantasma”, las cuales se forman al crearse

la tabla de pasaporte en Excel y tener algunas filas de más, que lamentablemente no pueden ser fácilmente identificadas porque aparecen en blanco pero que al exportarse la tabla en formato de texto, en ese momento aparecen. El sistema entiende que son entradas, porque ocupan una fila, pero no tienen datos en absoluto, incluyendo coordenadas, por lo que se genera este error.

- **Aviso de error:** An error occurred: Error en CRS(as.character(projection(crs))) : projection not namedCalls: sourcerasterFromRasterFile -> raster -> raster -> .local -> CRS

Modo en el que se puede producir: Local y on server

Herramienta en la que se produce: Se puede presentar en alguna de las herramientas que utiliza mapas ELC como Representa, ColNucleo o FIGS_R.

Solución(es): El mapa ELC producido por la herramienta ELCmapas que fue copiado y pegado en la ruta CAPFITOGEN2/ELCmapas fue editado y durante ese proceso se alteró el sistema de coordenadas, por lo cual el paquete Raster de R no lo puede leer. Por favor, utilice un mapa ELC sin editar, cópielo y péguelo en la ruta CAPFITOGEN2/ELCmapas, y con ello el problema debe desaparecer. Para tal fin, la herramienta ELCmapas produce dos copias del mapa, una para poder ser editada en DIVA-GIS y la otra se debe preservar sin edición para ser usada posteriormente.





19 | Anexos

19.1. Nombres de variables disponibles en CAPFITOGEN3, tal y como se usan en modo local y como se encuentran en modo on server al seleccionar idioma español

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
1	alt	elevación	Elevación	Geofísica
2	aspect	Orientacion	Orientación	Geofísica
3	bio_1	Temp prom anual	Temp prom anual	Bioclimática
4	bio_10	Temp prom cuarto mas cálido	Temp prom cuarto mas calido	Bioclimática
5	bio_11	Temp prom cuarto mas frio	Temp prom cuarto mas frio	Bioclimática
6	bio_12	Prec anual	Prec anual	Bioclimática
7	bio_13	Prec mes mas humedo	Prec mes mas humedo	Bioclimática
8	bio_14	Prec mes mas seco	Prec mes mas seco	Bioclimática
9	bio_15	Estacionalidad prec	Estacionalidad prec	Bioclimática
10	bio_16	Prec cuarto mas humedo	Prec cuarto mas humedo	Bioclimática
11	bio_17	Prec cuarto mas seco	Prec cuarto mas seco	Bioclimática
12	bio_18	Prec cuarto mas calido	Prec cuarto mas calido	Bioclimática
13	bio_19	Prec cuarto mas frio	Prec cuarto mas frio	Bioclimática
14	bio_2	Rango prom temp diurnas	Rango prom temp diurnas	Bioclimática
15	bio_3	Isotermalidad	Isotermalidad	Bioclimática
16	bio_4	Estacionalidad temp	Estacionalidad temp	Bioclimática
17	bio_5	Max temp mes mas calido	Max temp mes mas calido	Bioclimática
18	bio_6	Min temp mes mas frio	Min temp mes mas frio	Bioclimática
19	bio_7	Rango temp anual	Rango temp anual	Bioclimática
20	bio_8	Temp prom cuarto humedo	Temp prom cuarto humedo	Bioclimática
21	bio_9	Temp prom cuarto seco	Temp prom cuarto seco	Bioclimática
22	eastness	Esticidad	Esticidad	Geofísica
23	northness	Norticidad	Norticidad	Geofísica
24	prec_1	Prec prom 1	Prec prom 1	Bioclimática
25	prec_10	Prec prom 10	Prec prom 10	Bioclimática
26	prec_11	Prec prom 11	Prec prom 11	Bioclimática
27	prec_12	Prec prom 12	Prec prom 12	Bioclimática

Continúa

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
28	prec_2	Prec prom 2	Prec prom 2	Bioclimática
29	prec_3	Prec prom 3	Prec prom 3	Bioclimática
30	prec_4	Prec prom 4	Prec prom 4	Bioclimática
31	prec_5	Prec prom 5	Prec prom 5	Bioclimática
32	prec_6	Prec prom 6	Prec prom 6	Bioclimática
33	prec_7	Prec prom 7	Prec prom 7	Bioclimática
34	prec_8	Prec prom 8	Prec prom 8	Bioclimática
35	prec_9	Prec prom 9	Prec prom 9	Bioclimática
36	ref_depth	Profundidad	Profundidad	Edáfica
37	s_bs	Sat bases subsuelo	Sat bases subsuelo	Edáfica
38	s_caco3	CaCO3 subsuelo	CaCO3 subsuelo	Edáfica
39	s_caso4	Yesos subsuelo	Yesos subsuelo	Edáfica
40	s_cec_clay	CIC arcilla subsuelo	CIC arcilla subsuelo	Edáfica
41	s_cec_soil	CIC subsuelo gral	CIC subsuelo gral	Edáfica
42	s_clay	Arcilla en subsuelo	Arcilla en subsuelo	Edáfica
43	s_ece	Salinidad subsuelo	Salinidad subsuelo	Edáfica
44	s_esp	Sodicidad subsuelo	Sodicidad subsuelo	Edáfica
45	s_gravel	Grava en subsuelo	Grava en subsuelo	Edáfica
46	s_oc	Carbon org subsuelo	Carbon org subsuelo	Edáfica
47	s_ph_h2o	pH subsuelo	pH subsuelo	Edáfica
48	s_ref_bulk	Densidad subsuelo	Densidad subsuelo	Edáfica
49	s_sand	Arena en subsuelo	Arena en subsuelo	Edáfica
50	s_silt	Limo en subsuelo	Limo en subsuelo	Edáfica
51	s_teb	Bases int subsuelo	Bases int subsuelo	Edáfica
52	slope	Pendiente grados	Pendiente grados	Geofísica
53	t_bs	Sat bases suelo	Sat bases suelo	Edáfica
54	t_caco3	CaCO3 suelo	CaCO3 suelo	Edáfica
55	t_caso4	Yesos suelo	Yesos suelo	Edáfica
56	t_cec_clay	CIC arcilla suelo	CIC arcilla suelo	Edáfica
57	t_cec_soil	CIC suelo gral	CIC suelo gral	Edáfica
58	t_clay	Arcilla en suelo	Arcilla en suelo	Edáfica

Continúa

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
59	t_ece	Salinidad suelo	Salinidad suelo	Edáfica
60	t_esp	Sodicidad suelo	Sodicidad suelo	Edáfica
61	t_gravel	Grava en suelo	Grava en suelo	Edáfica
62	t_oc	Carbon org suelo	Carbon org suelo	Edáfica
63	t_ph_h2o	pH suelo	pH suelo	Edáfica
64	t_ref_bulk	Densidad suelo	Densidad suelo	Edáfica
65	t_sand	Arena en suelo	Arena en suelo	Edáfica
66	t_silt	Limo en suelo	Limo en suelo	Edáfica
67	t_teb	Bases int suelo	Bases int suelo	Edáfica
68	tmax_1	Temp max 1	Temp max 1	Bioclimática
69	tmax_10	Temp max 10	Temp max 10	Bioclimática
70	tmax_11	Temp max 11	Temp max 11	Bioclimática
71	tmax_12	Temp max 12	Temp max 12	Bioclimática
72	tmax_2	Temp max 2	Temp max 2	Bioclimática
73	tmax_3	Temp max 3	Temp max 3	Bioclimática
74	tmax_4	Temp max 4	Temp max 4	Bioclimática
75	tmax_5	Temp max 5	Temp max 5	Bioclimática
76	tmax_6	Temp max 6	Temp max 6	Bioclimática
77	tmax_7	Temp max 7	Temp max 7	Bioclimática
78	tmax_8	Temp max 8	Temp max 8	Bioclimática
79	tmax_9	Temp max 9	Temp max 9	Bioclimática
80	tmean_1	Temp prom 1	Temp prom 1	Bioclimática
81	tmean_10	Temp prom 10	Temp prom 10	Bioclimática
82	tmean_11	Temp prom 11	Temp prom 11	Bioclimática
83	tmean_12	Temp prom 12	Temp prom 12	Bioclimática
84	tmean_2	Temp prom 2	Temp prom 2	Bioclimática
85	tmean_3	Temp prom 3	Temp prom 3	Bioclimática
86	tmean_4	Temp prom 4	Temp prom 4	Bioclimática
87	tmean_5	Temp prom 5	Temp prom 5	Bioclimática
88	tmean_6	Temp prom 6	Temp prom 6	Bioclimática
89	tmean_7	Temp prom 7	Temp prom 7	Bioclimática

Continúa

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
90	tmean_8	Temp prom 8	Temp prom 8	Bioclimática
91	tmean_9	Temp prom 9	Temp prom 9	Bioclimática
92	tmin_1	Temp min 1	Temp min 1	Bioclimática
93	tmin_10	Temp min 10	Temp min 10	Bioclimática
94	tmin_11	Temp min 11	Temp min 11	Bioclimática
95	tmin_12	Temp min 12	Temp min 12	Bioclimática
96	tmin_2	Temp min 2	Temp min 2	Bioclimática
97	tmin_3	Temp min 3	Temp min 3	Bioclimática
98	tmin_4	Temp min 4	Temp min 4	Bioclimática
99	tmin_5	Temp min 5	Temp min 5	Bioclimática
100	tmin_6	Temp min 6	Temp min 6	Bioclimática
101	tmin_7	Temp min 7	Temp min 7	Bioclimática
102	tmin_8	Temp min 8	Temp min 8	Bioclimática
103	tmin_9	Temp min 9	Temp min 9	Bioclimática
104	POINT_X	Longitud	Longitud	Geofísica
105	POINT_Y	Latitud	Latitud	Geofísica
106	vapr_1	Presion de vapor 1	Presion de vapor 1	Bioclimática
107	vapr_2	Presion de vapor 2	Presion de vapor 2	Bioclimática
108	vapr_3	Presion de vapor 3	Presion de vapor 3	Bioclimática
109	vapr_4	Presion de vapor 4	Presion de vapor 4	Bioclimática
110	vapr_5	Presion de vapor 5	Presion de vapor 5	Bioclimática
111	vapr_6	Presion de vapor 6	Presion de vapor 6	Bioclimática
112	vapr_7	Presion de vapor 7	Presion de vapor 7	Bioclimática
113	vapr_8	Presion de vapor 8	Presion de vapor 8	Bioclimática
114	vapr_9	Presion de vapor 9	Presion de vapor 9	Bioclimática
115	vapr_10	Presion de vapor 10	Presion de vapor 10	Bioclimática
116	vapr_11	Presion de vapor 11	Presion de vapor 11	Bioclimática
117	vapr_12	Presion de vapor 12	Presion de vapor 12	Bioclimática
118	vapr_annual	Presion de vapor anual	Presion de vapor anual	Bioclimática
119	wind_1	Velocidad viento 1	Velocidad viento 1	Geofísica
120	wind_2	Velocidad viento 2	Velocidad viento 2	Geofísica

Continúa

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
121	wind_3	Velocidad viento 3	Velocidad viento 3	Geofísica
122	wind_4	Velocidad viento 4	Velocidad viento 4	Geofísica
123	wind_5	Velocidad viento 5	Velocidad viento 5	Geofísica
124	wind_6	Velocidad viento 6	Velocidad viento 6	Geofísica
125	wind_7	Velocidad viento 7	Velocidad viento 7	Geofísica
126	wind_8	Velocidad viento 8	Velocidad viento 8	Geofísica
127	wind_9	Velocidad viento 9	Velocidad viento 9	Geofísica
128	wind_10	Velocidad viento 10	Velocidad viento 10	Geofísica
129	wind_11	Velocidad viento 11	Velocidad viento 11	Geofísica
130	wind_12	Velocidad viento 12	Velocidad viento 12	Geofísica
131	wind_annual	Velocidad viento anual	Velocidad viento anual	Geofísica
132	srad_1	Radiacion solar 1	Radiacion solar 1	Geofísica
133	srad_2	Radiacion solar 2	Radiacion solar 2	Geofísica
134	srad_3	Radiacion solar 3	Radiacion solar 3	Geofísica
135	srad_4	Radiacion solar 4	Radiacion solar 4	Geofísica
136	srad_5	Radiacion solar 5	Radiacion solar 5	Geofísica
137	srad_6	Radiacion solar 6	Radiacion solar 6	Geofísica
138	srad_7	Radiacion solar 7	Radiacion solar 7	Geofísica
139	srad_8	Radiacion solar 8	Radiacion solar 8	Geofísica
140	srad_9	Radiacion solar 9	Radiacion solar 9	Geofísica
141	srad_10	Radiacion solar 10	Radiacion solar 10	Geofísica
142	srad_11	Radiacion solar 11	Radiacion solar 11	Geofísica
143	srad_12	Radiacion solar 12	Radiacion solar 12	Geofísica
144	srad_annual	Radiacion solar anual	Radiacion solar anual	Geofísica
145	t_awc1	Avail soil water cap h1 top	Disp agua cap h1 top	Edáfica
146	s_awc1	Avail soil water cap h1 sub	Disp agua cap h1 sub	Edáfica
147	t_awc2	Avail soil water cap h2 top	Disp agua cap h2 top	Edáfica
148	s_awc2	Avail soil water cap h2 sub	Disp agua cap h2 sub	Edáfica
149	t_awc3	Avail soil water cap h3 top	Disp agua cap h3 top	Edáfica
150	s_awc3	Avail soil water cap h3 sub	Disp agua cap h3 sub	Edáfica
151	t_awcts	Sat water cont top	Cont agua sat top	Edáfica

Continúa

ID	CODE	VARIABLE MODO LOCAL	VARIABLE MODO ON SERVER	COMPONENTE
152	s_awcts	Sat water cont sub	Cont agua sat sub	Edáfica
153	depth_rock	Depth to bedrock	Profund a rocas	Edáfica
154	r_horizon	R horizon	Horizonte R	Edáfica
155	t_bulk_dens	Bulk density top	Densidad aparente top	Edáfica
156	s_bulk_dens	Bulk density sub	Densidad aparente sub	Edáfica
157	t_cecsol	Cation exchange cap top	Cap interc cation top	Edáfica
158	s_cecsol	Cation exchange cap sub	Cap interc cation sub	Edáfica
159	t_clay_cont	Clay content top	Contenido arcilla top	Edáfica
160	s_clay_cont	Clay content sub	Contenido arcilla sub	Edáfica
161	t_coarse_frag	Coarse fragments top	Fragment grueso top	Edáfica
162	s_coarse_frag	Coarse fragments sub	Fragment grueso sub	Edáfica
163	t_oc_dens	Organic carbon dens top	Dens carbon org top	Edáfica
164	s_oc_dens	Organic carbon dens sub	Dens carbon org sub	Edáfica
165	t_oc_stock	Organic carbon stock top	Stock carbon org top	Edáfica
166	s_oc_stock	Organic carbon stock sub	Stock carbon org sub	Edáfica
167	t_oc_cont	Organic carbon content top	Cont carbon org top	Edáfica
168	s_oc_cont	Organic carbon content sub	Cont carbon org sub	Edáfica
169	t_ph_hox	Soil pH H2O top	pH suelo en H2O top	Edáfica
170	s_ph_hox	Soil pH H2O sub	pH suelo en H2O sub	Edáfica
171	t_ph_kcl	Soil pH KCl top	pH suelo en KCl top	Edáfica
172	s_ph_kcl	Soil pH KCl sub	pH suelo en KCl sub	Edáfica
173	sodicity	Sodic soil grade	Grado sodicidad suelo	Edáfica
174	t_silt_cont	Silt content top	Contenido limo top	Edáfica
175	s_silt_cont	Silt content sub	Contenido limo sub	Edáfica
176	t_sand_cont	Sand content top	Contenido arena top	Edáfica
177	s_sand_cont	Sand content sub	Contenido arena sub	Edáfica
178	t_soilwater_cap	Avail soil water cap top	Disp agua punto marchitez top	Edáfica
179	s_soilwater_cap	Avail soil water cap sub	Disp agua punto marchitez sub	Edáfica

19.2. Descripción de las variables disponibles en CAPFITOGEN3 (ID es un campo común con ID de la tabla 19.1)

ID	DESCRIPCIÓN	FUENTE
1	Elevación. Metros sobre el nivel del mar	http://worldclim.org
2	Orientación (en grados) de la superficie del terreno. 0 y 359 corresponden al norte	NA
3	Temperatura media anual	http://worldclim.org
4	Temperatura media del cuarto más cálido (los tres meses más cálidos)	http://worldclim.org
5	Temperatura media del cuarto más frío (los tres meses más fríos)	http://worldclim.org
6	Precipitación anual	http://worldclim.org
7	Precipitación del mes más húmedo	http://worldclim.org
8	Precipitación del mes más seco	http://worldclim.org
9	Estacionalidad en la Precipitación (coeficiente de variación)	http://worldclim.org
10	Precipitación del cuarto más húmedo (los tres meses más lluviosos)	http://worldclim.org
11	Precipitación del cuarto más seco (los tres meses más secos)	http://worldclim.org
12	Precipitación del cuarto más cálido (los tres meses más cálidos)	http://worldclim.org
13	Precipitación del cuarto más frío (los tres meses más fríos)	http://worldclim.org
14	Rango medio de temperaturas diurnas (medias mensuales (máxima temperatura-mínima temperatura))	http://worldclim.org
15	Isotermalidad (BIOCLIM2/BIOCLIM7) (x100)	http://worldclim.org
16	Estacionalidad en la temperatura (desviación estandarx100)	http://worldclim.org
17	Máxima temperatura del mes más cálido	http://worldclim.org
18	Mínima temperatura del mes más frío	http://worldclim.org
19	Rango de temperatura anual (BIOCLIM5 - BIOCLIM6)	http://worldclim.org
20	Temperatura media del cuarto más húmedo (los tres meses más lluviosos)	http://worldclim.org
21	Temperatura media del cuarto más seco (los tres meses más secos)	http://worldclim.org
22	Esticidad. Toma valores cercanos a 1 si la Orientación tiende al este, -1 si tiende al oeste y 0 si la Orientación tiende a norte o sur.	NA
23	Norticidad. Toma valores cercanos a 1 si la Orientación tiende al norte, -1 si tiende al sur y 0 si la Orientación tiende a este u oeste.	NA
24	Precipitación media de enero	http://worldclim.org
25	Precipitación media de octubre	http://worldclim.org

Continúa

ID	DESCRIPCIÓN	FUENTE
26	Precipitación media de noviembre	http://worldclim.org
27	Precipitación media de diciembre	http://worldclim.org
28	Precipitación media de febrero	http://worldclim.org
29	Precipitación media de marzo	http://worldclim.org
30	Precipitación media de abril	http://worldclim.org
31	Precipitación media de mayo	http://worldclim.org
32	Precipitación media de junio	http://worldclim.org
33	Precipitación media de julio	http://worldclim.org
34	Precipitación media de agosto	http://worldclim.org
35	Precipitación media de septiembre	http://worldclim.org
36	Profundidad referencia de la unidad de suelo. La profundidad referencia de todas las unidades están ajustadas a 100 cm (con algunas excepciones)	HWSD (http://t.ly/dbry)
37	Saturación de bases en subsuelo.	HWSD (http://t.ly/dbry)
38	Carbonato de calcio en subsuelo. Este campo ofrece el contenido de cal en subsuelo.	HWSD (http://t.ly/dbry)
39	Yesos en subsuelo. Este campo ofrece el contenido de sulfato de calcio en subsuelo.	HWSD (http://t.ly/dbry)
40	CEC de arcilla en subsuelo. CEC arcilla es la capacidad de intercambio catiónico en la fracción de arcilla	HWSD (http://t.ly/dbry)
41	CEC del suelo (general) en subsuelo. Este campo ofrece la capacidad de intercambio catiónico en subsuelo.	HWSD (http://t.ly/dbry)
42	Contenido de arcilla en subsuelo	HWSD (http://t.ly/dbry)
43	Salinidad en subsuelo. Este campo ofrece la conductividad eléctrica en subsuelo.	HWSD (http://t.ly/dbry)
44	Sodicidad en subsuelo. Este campo ofrece el porcentaje de sodio intercambiable en subsuelo.	HWSD (http://t.ly/dbry)
45	Contenido de grava en subsuelo	HWSD (http://t.ly/dbry)
46	Contenido de carbón orgánico en subsuelo	HWSD (http://t.ly/dbry)
47	pH en subsuelo. pH, medido en solución suelo-agua, es una medida de acidez o alcalinidad del suelo.	HWSD (http://t.ly/dbry)
48	Densidad aparente de referencia en subsuelo. Es la masa de muchas partículas del material dividido por el volumen que ocupan.	HWSD (http://t.ly/dbry)
49	Contenido de arena en subsuelo	HWSD (http://t.ly/dbry)
50	Contenido de limo en subsuelo	HWSD (http://t.ly/dbry)

Continúa

ID	DESCRIPCIÓN	FUENTE
51	Bases intercambiables totales en subsuelo	HWSD (http://t.ly/dbry)
52	Pendiente (en grados) de la superficie del terreno	NA
53	Saturación de bases en suelo superficial.	HWSD (http://t.ly/dbry)
54	Carbónato de calcio en suelo superficial. Este campo ofrece el contenido de cal en suelo superficial.	HWSD (http://t.ly/dbry)
55	Yesos en suelo superficial. Este campo ofrece el contenido de sulfato de calcio en suelo superficial.	HWSD (http://t.ly/dbry)
56	CEC de arcilla en suelo superficial. CEC arcilla es la capacidad de intercambio cationico en la fraccion de arcilla	HWSD (http://t.ly/dbry)
57	CEC del suelo (general) en suelo superficial. Este campo ofrece la capacidad de intercambio cationico en suelo superficial.	HWSD (http://t.ly/dbry)
58	Contenido de arcilla en suelo superficial	HWSD (http://t.ly/dbry)
59	Salinidad en suelo superficial. Este campo ofrece la conductividad electrica en suelo superficial.	HWSD (http://t.ly/dbry)
60	Sodicidad en suelo superficial. Este campo ofrece el porcentaje de sodio intercambiable en suelo superficial.	HWSD (http://t.ly/dbry)
61	Contenido de grava en suelo superficial	HWSD (http://t.ly/dbry)
62	Contenido de carbón organico en suelo superficial	HWSD (http://t.ly/dbry)
63	pH en suelo superficial. pH, medido en solucion suelo-agua, es una medida de acidez o alcalinidad del suelo.	HWSD (http://t.ly/dbry)
64	Densidad aparente de referencia en suelo superficial. Es la masa de muchas particulas del material dividido por el volumen que ocupan.	HWSD (http://t.ly/dbry)
65	Contenido de arena en suelo superficial	HWSD (http://t.ly/dbry)
66	Contenido de limo en suelo superficial	HWSD (http://t.ly/dbry)
67	Bases intercambiables totales en suelo superficial	HWSD (http://t.ly/dbry)
68	Temperatura máxima de enero	http://worldclim.org
69	Temperatura máxima de octubre	http://worldclim.org
70	Temperatura máxima de noviembre	http://worldclim.org
71	Temperatura máxima de diciembre	http://worldclim.org
72	Temperatura máxima de febrero	http://worldclim.org
73	Temperatura máxima de marzo	http://worldclim.org
74	Temperatura máxima de abril	http://worldclim.org
75	Temperatura máxima de mayo	http://worldclim.org
76	Temperatura máxima de junio	http://worldclim.org

Continúa

ID	DESCRIPCIÓN	FUENTE
77	Temperatura máxima de julio	http://worldclim.org
78	Temperatura máxima de agosto	http://worldclim.org
79	Temperatura máxima de septiembre	http://worldclim.org
80	Temperatura media de enero	http://worldclim.org
81	Temperatura media de octubre	http://worldclim.org
82	Temperatura media de noviembre	http://worldclim.org
83	Temperatura media de diciembre	http://worldclim.org
84	Temperatura media de febrero	http://worldclim.org
85	Temperatura media de marzo	http://worldclim.org
86	Temperatura media de abril	http://worldclim.org
87	Temperatura media de mayo	http://worldclim.org
88	Temperatura media de junio	http://worldclim.org
89	Temperatura media de julio	http://worldclim.org
90	Temperatura media de agosto	http://worldclim.org
91	Temperatura media de septiembre	http://worldclim.org
92	Temperatura mínima de enero	http://worldclim.org
93	Temperatura mínima de octubre	http://worldclim.org
94	Temperatura mínima de noviembre	http://worldclim.org
95	Temperatura mínima de diciembre	http://worldclim.org
96	Temperatura mínima de febrero	http://worldclim.org
97	Temperatura mínima de marzo	http://worldclim.org
98	Temperatura mínima de abril	http://worldclim.org
99	Temperatura mínima de mayo	http://worldclim.org
100	Temperatura mínima de junio	http://worldclim.org
101	Temperatura mínima de julio	http://worldclim.org
102	Temperatura mínima de agosto	http://worldclim.org
103	Temperatura mínima de septiembre	http://worldclim.org
104	Longitud para el centroide de la celda	NA
105	Latitud para centroide de la celda	NA
106	Presión de vapor de enero	http://worldclim.org
107	Presión de vapor de febrero	http://worldclim.org

Continúa

ID	DESCRIPCIÓN	FUENTE
108	Presión de vapor de marzo	http://worldclim.org
109	Presión de vapor de abril	http://worldclim.org
110	Presión de vapor de mayo	http://worldclim.org
111	Presión de vapor de junio	http://worldclim.org
112	Presión de vapor de julio	http://worldclim.org
113	Presión de vapor de agosto	http://worldclim.org
114	Presión de vapor de septiembre	http://worldclim.org
115	Presión de vapor de octubre	http://worldclim.org
116	Presión de vapor de noviembre	http://worldclim.org
117	Presión de vapor de diciembre	http://worldclim.org
118	Presión de vapor promedio anual	http://worldclim.org
119	Velocidad viento enero	http://worldclim.org
120	Velocidad viento febrero	http://worldclim.org
121	Velocidad viento marzo	http://worldclim.org
122	Velocidad viento abril	http://worldclim.org
123	Velocidad viento mayo	http://worldclim.org
124	Velocidad viento junio	http://worldclim.org
125	Velocidad viento julio	http://worldclim.org
126	Velocidad viento agosto	http://worldclim.org
127	Velocidad viento septiembre	http://worldclim.org
128	Velocidad viento octubre	http://worldclim.org
129	Velocidad viento noviembre	http://worldclim.org
130	Velocidad viento diciembre	http://worldclim.org
131	Velocidad viento anual	http://worldclim.org
132	Radiación solar enero	http://worldclim.org
133	Radiación solar febrero	http://worldclim.org
134	Radiación solar marzo	http://worldclim.org
135	Radiación solar abril	http://worldclim.org
136	Radiación solar mayo	http://worldclim.org
137	Radiación solar junio	http://worldclim.org
138	Radiación solar julio	http://worldclim.org

Continúa

ID	DESCRIPCIÓN	FUENTE
139	Radiación solar agosto	http://worldclim.org
140	Radiación solar septiembre	http://worldclim.org
141	Radiación solar octubre	http://worldclim.org
142	Radiación solar noviembre	http://worldclim.org
143	Radiación solar diciembre	http://worldclim.org
144	Radiación solar anual	http://worldclim.org
145	Capacidad agua disponible suelo (fracción volumétrica) para h1 - top	https://soilgrids.org
146	Capacidad agua disponible suelo (fracción volumétrica) para h1 - sub	https://soilgrids.org
147	Capacidad agua disponible suelo (fracción volumétrica) para h2 - top	https://soilgrids.org
148	Capacidad agua disponible suelo (fracción volumétrica) para h2 - sub	https://soilgrids.org
149	Capacidad agua disponible suelo (fracción volumétrica) para h3 - top	https://soilgrids.org
150	Capacidad agua disponible suelo (fracción volumétrica) para h3 - sub	https://soilgrids.org
151	Contenido agua saturada (fracción volumétrica) para tS - topsoil	https://soilgrids.org
152	Contenido agua saturada (fracción volumétrica) para tS - subsoil	https://soilgrids.org
153	Profundidad a lecho rocoso (horizonte R) hasta 200 cm	https://soilgrids.org
154	Probabilidad ocurrencia horizonte R	https://soilgrids.org
155	Densidad aparente (tierra fina) en kg / metro cúbico - topsoil	https://soilgrids.org
156	Densidad aparente (tierra fina) en kg / metro cúbico - subsoil	https://soilgrids.org
157	Capacidad intercambio catiónico suelo en cmolc/kg - topsoil	https://soilgrids.org
158	Capacidad intercambio catiónico suelo en cmolc/kg - subsoil	https://soilgrids.org
159	Contenido arcilla (0-2 micro metros) fracción de masa en % - topsoil	https://soilgrids.org
160	Contenido arcilla (0-2 micro metros) fracción de masa en % - subsoil	https://soilgrids.org
161	Fragmentos gruesos volumétricos en % topsoil	https://soilgrids.org
162	Fragmentos gruesos volumétricos en % subsoil	https://soilgrids.org
163	Densidad de carbono orgánico en suelo kg por m cúbico topsoil	https://soilgrids.org
164	Densidad de carbono orgánico en suelo kg por m cúbico subsoil	https://soilgrids.org
165	Stock de carbono orgánico en tons por ha topsoil	https://soilgrids.org
166	Stock de carbono orgánico en tons por ha subsoil	https://soilgrids.org
167	Contenido carbón orgánico suelo (fracción tierra fina) en g por kg topsoil	https://soilgrids.org
168	Contenido carbón orgánico suelo (fracción tierra fina) en g por kg subsoil	https://soilgrids.org
169	pH suelo en H2O topsoil	https://soilgrids.org

Continúa

ID	DESCRIPCIÓN	FUENTE
170	pH suelo en H2O subsoil	https://soilgrids.org
171	pH suelo en KCl topsoil	https://soilgrids.org
172	pH suelo en KCl subsoil	https://soilgrids.org
173	Grado sodicidad del suelo	https://soilgrids.org
174	Contenido limo (2-50 micro metros) fracción de masa en % topsoil	https://soilgrids.org
175	Contenido limo (2-50 micro metros) fracción de masa en % subsoil	https://soilgrids.org
176	Contenido arena (50-2000 micro metros) fracción de masa en % topsoil	https://soilgrids.org
177	Contenido arena (50-2000 micro metros) fracción de masa en % subsoil	https://soilgrids.org
178	Capacidad agua disponible suelo (fracción volumétrica) hasta punto marchitez topsoil	https://soilgrids.org
179	Capacidad agua disponible suelo (fracción volumétrica) hasta punto marchitez subsoil	https://soilgrids.org

19.3. Explicación de columnas extra de la tabla “tabla_de_analisisGEOQUAL.txt/xls”

Variable	Explicación
globlandc	Valor extraído de GLC 2000 (Global Land Cover 2000).
DISTOLAND	Anillo de distancia a la tierra dentro del cual las coordenadas se encuentran. (0 = tierra, 1 = 1 km, 10 = 10 km, etc.).
SUITQUAL	Parámetro SUITQUAL (valores de 0 a 20).
ID_0	Valores extraídos de GADM identifying the area of the country.
ISO	Valores extraídos de GADM compared with ORIGCTY.
NAME_0	Valores extraídos de GADM for the country's full name.
ID_1	Valores extraídos de GADM identifying the area at the NAME_1 level.
NAME_1	Valores extraídos de GADM compared with ADM1.
VARNAME_1	Valores extraídos de GADM for alternative names to NAME_1.
ENGTYPE_1	Valores extraídos de GADM defining the type of administration represented by NAME_1.
ID_2	Valores extraídos de GADM identifying the area at the NAME_2 level.
NAME_2	Valores extraídos de GADM compared with ADM2.
VARNAME_2	Valores extraídos de GADM for alternative names to NAME_2.
ENGTYPE_2	Valores extraídos de GADM defining the type of administration represented by NAME_2.

Continúa

Variable	Explicación
ID_3	Valores extraídos de GADM identifying the area at the NAME_3 level.
NAME_3	Valores extraídos de GADM compared with ADM3.
VARNAME_3	Valores extraídos de GADM for alternative names to NAME_3.
ENGTYPE_3	Valores extraídos de GADM defining the type of administration represented by NAME_3.
ID_4	Valores extraídos de GADM identifying the area at the NAME_4 level.
NAME_4	Valores extraídos de GADM compared with ADM4.
VARNAME_4	Value extracted from GADM for alternative names to NAME_4.
ENGTYPE4	Value extracted from GADM defining the type of administration represented by NAME_4.
NIVELMAX	Depending on the country, this is the lowest administrative level included in GADM.
LOCALQUAL	LOCALQUAL parameter (values 0 to 20).
COORQUAL	COORQUAL parameter (values 0 to 20).
intertemp	COORQUAL intertemp sub-parameter
errors	COORQUAL errors sub-parameter
precis	COORQUAL precis sub-parameter
georble	COORQUAL georble sub-parameter
TOTALQUAL	TOTALQUAL parameter (values from 0 to 40 or 0 to 60, depending on whether LOCALQUAL is included or not).
TOTALQUAL100	TOTALQUAL100 parameter (values from 0 to 100).



<http://www.capfitogen.net/es/>
