UNIVERSIDAD
NACIONAL
DE COLOMBIA

Caracterización del genoma de *Haemoproteus* (*Haemoproteus*) *columbae*: como herramienta para el estudio evolutivo del orden Haemosporida

Characterization of *Haemoproteus* (*Haemoproteus*) *columbae* genome: as a tool for evolutionary study of Haemosporida order

Axl S. Cepeda

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Biología
Maestría en Ciencias - Biología
Bogotá D.C., Colombia
2019

Characterization of *Haemoproteus* (*Haemoproteus*) *columbae* genome: as a tool for evolutionary study of Haemosporida order

Axl S. Cepeda

Tesis presentada como requisito parcial para optar al título de:
Magister en Ciencias - Biología

Directora:
Ph.D. Nubia Estela Matta Camacho
Departamento de Biología, Universidad Nacional de Colombia.

Línea de Investigación:
Relación Parasito Hospedero: Genómica
Grupo de Investigación:
Caracterización Genética e Inmunología

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Biología
Bogotá D.C., Colombia
2019

Lema

"Un gran pensador inglés dijo que *la verdadera Universidad hoy en día son los libros*, y esta verdad, a pesar del desarrollo que modernamente han tenido las instituciones docentes, es en la actualidad más cierta que nunca. Nada aprende mejor el hombre que lo que aprende por sí mismo, lo que le exige un esfuerzo personal de búsqueda y de asimilación; si los maestros sirven de guías y orientadores, las fuentes perennes del conocimiento está en los libros".

Próposito del Libro; Fausto - Goethe

"... Creéis que todo tiene un límite, y así estáis todos, **limitados** ..."

Cuidado - Eskorbuto

Acknowledgments

- A la vida, que me ha demostrado que los fracasos son inexistentes bajo la luz del esfuerzo.
- A mi mamá, quien sacrificó su vida por hacer de mi una persona virtuosa.
- A la profesora Nubia Matta, quien nunca ha dejado de creer en mi talento, pasión y amor por lo que yo hago.
- A Melisa, que mientras estuvo en mi camino, no desfalleció en su intento de hacerme alcanzar mis sueños.
- A Ingrid Lotta por sus consejos, su paciencia y su apoyo incondicional.
- A los profesores María Andreina Pacheco, Ananias Escalante, Juan Fernando Alzate y Gediminas Valkiunas por sus enseñanzas y valiosos aportes.
- A mis compañeros y amigos del Laboratorio: Fredy Colorado, Rafael Gutiérrez, Germán Gutiérrez, Paola González, Juan Valencia y Carolina Vargas, por compartir un espacio de formación personal y profesional.
- A la nueva generación GERPH: que mis pasos recorridos sean un ejemplo para que ustedes crean en sus talentos, pasiones, capacidades, etc. - Recuerden que no hay límites.
- A mis amigos Aimer Gutiérrez, Ricardo Barrera, Mauro Díaz, Daniel Arias y Jonathan Duque por tantos años de amistad inquebrantable.

Abstract

Research on malaria has focused during a long time on the parasites that infect humans. However, it is also true that most information about the biology of this parasites comes from experimental models. For this reason, this thesis focuses on a Haemosporida parasite closely related to the *Plasmodium* genus, which is the *Haemoproteus* parasites.

Methods include standardizing an experimental animal model for the *Haemoproteus* transmission. The approach involved the natural host Rock Pigeon (*Columba livia*), the louse flies (*Pseudolynchia canariensis*) which are the vectors and the parasite *Haemoproteus columbae*. The first hole in the road to overcoming was to increase the number of parasites present in the blood sample (parasitemia); in this way, it was possible to reduce the gap between the proportion of host DNA and parasite DNA. Besides, there was necessary to standardize the conditions to reared louse flies in the lab, and the methodologies that allow following the infection both in the vector and in the vertebrate host these results are shown in **Chapter 6**.

On the other hand, total genomic DNA was sequenced on Illumina HiSeqX 150-bp technology, resulting in a total of 628′859,636 pair-end reads. It allows obtaining the complete apicoplast and mitochondrial genomes, with enough coverage to be considered reference genomes. These results are shown in **Chapters 7** and **8**.

Regarding the Apigenome, phylogenetic, phylogenomic and evolutionary analyses were carried out, highlighting an evolutionary dynamic related to GC content bias within Haemosporida order, which had an impact on the Substitution Saturation of coding sequences as well as in the Codon Usage. Finally, a draft nuclear genome was assembled, and 3976 genes were annotated. Likewise, evidence of LTR-transposon sequences was found, which play a critical role in the evolutionary dynamics of genomes (**Chapter 8**).

In conclusion, this thesis present to the scientific community an experimental animal model that undoubtedly will allow new approximations to characterize the life cycle of vector-borne parasites. The apicoplast genome of this parasite allowed us to study the evolutionary dynamics of this organelle inside and outside the Haemosporida order. Finally, we generate the first draft genome of parasite belonging to *Haemoproteus* genus, subgenus *Haemoproteus*, which it could be useful for genetic, immunological, evolutionary, and ecological studies, among others. Altogether, this thesis generates valuable information that open possibilities to explore new approaches to characterize in-depth the biology, evolution and phylogenetic relationships of apicomplexan parasites.

Outline

List of figures

List of tables

Objectives

General objective

To characterize the *Haemoproteus columbae* genome, in order to increase the phylogenetic resolution of Haemosporida order.

Specific objectives

1. To obtain genomic DNA of *H. columbae* from the enrichment of parasites present in blood samples from the host.

2. To characterize structurally and functionally some genes present in assembled sequences of *H. columbae* genome.

3. To carry out a phylogenomic approach of the Haemosporida order, including genomic sequences of *H. columbae*.

Introduction

The Hemosporidia are an order of parasites widely distributed in the world, capable of infecting amphibians, birds, mammals and reptiles; of which, three genera have special importance in epidemiology, public health and in wildlife conservation: *Plasmodium*, *Leucocytozoon* and *Haemoproteus* (Bennett *et al.*, 1965; Valkiūnas, 2005). However, *Plasmodium* has been the most historically studied parasite, because it contains the most pathogenic species for humans such as *P. falciparum* and *P. vivax* (WHO, 2015) and for birds *P. relictum* (van Riper *et al.*, 1986).

Although there are approximately $2,405$ molecular lineages of the mitochondrial *cytochrome b* gene deposited in the parasitic database of MalAvi (Bensch *et al.*, 2009), there is still no consensus on the phylogenetic relationships of this order. Some authors, such as Martinsen *et al.,* (2008) and Pacheco *et al.,* (2017) propose to genus *Haemoproteus* as a **polyphyletic** group. Whereas Valkiūnas *et al.,* (2008), Levin *et al.,* (2013), and Field *et al.,* (2018) propose to genus *Haemoproteus* as a **monophyletic** group. Therefore, it is necessary to increase the omic information to carry out more robust analyzes at an evolutionary, phylogenetic, genetic and epidemiological level, among others in Haemosporida order.

I think that four major problems persist within this order that prevent a more accelerated advance in the inconsistency of phylogenetic hypotheses: **1)** low phylogenetic resolution generated by small sizes of the molecular markers availables (470bp for cyt b gene; Hellgren *et al.*, 2004), paucity of experimental models (Bukauskaitė *et al.*, 2015; Bukauskaité and Valkiūnas, 2016; Cepeda *et al.*, 2019a); **2)** the proportion of parasitic DNA versus host DNA (for haemosporida that infect birds, reptiles and amphibians) can reach values of $1:100000$, due to the presence of nucleated erythrocytes, parasite haploid phase, parasitaemia, among others (Videvall, 2019); **3)** Insufficient information for parasites present in wildlife (Bahl *et al.*, 2003; Bensch *et al.*, 2009; Bensch *et al.*, 2016; Böhme *et al.*, 2018); and **4)** few analysis on evolutionary features of these nuclear, mitochondrial and apicoplast genomes (Pacheco *et al.*, 2017; Cepeda *et al.*, 2019b).

Currently, except for the known models in macaques and mice, there are few animal models available to study and characterize the host-parasite-vector interactions in avian Haemosporidian, that at the beginning demonstrated were very important in advance of the knowledge of malaria parasites (Valkiūnas, 2005). One of the models is *Serinus canaria* - *Culex quinquefasciatus* and *Culex pipiens* mosquitoes and the another one is *Agelaius phoeniceus* and *Cx. pipiens*, which are used for modelling interactions between malaria parasites (*Plasmodium* species) and avian host (LaPointe *et al.*, 2005; Kimura, 2008; Valkiūnas *et al.*, 2015). For

that reason one of the main aims of this thesis project was to standardize an experimental model for parasites of the subgenus Haemoproteus (Cepeda et al., 2019a submited).

Important to mention that the statistical analysis centre of the European Nucleotide Archive (ENA, 2017) annually reports that information at the genomic level is doubled, with 884,4 million sequences deposited in public databases for the date. Nonetheless, there are only available 22 reference nuclear genomes of the genus *Plasmodium* (PlasmoDB) and one of the genus *Haemoproteus* (subgenus *Parahaemoproteus*; Bensch *et al.*, 2016) from the 500 species that build up the order Haemosporidia; 114 mtDNA genomes from species belonging to four genera: *Leucocytozoon, Haemoproteus* (subgenera *Haemoproteus* and *Parahaemoproteus*), *Plasmodium*, and *Hepatocystis* (Pacheco *et al.*, 2017); at least 20 apicoplast genomes belonging to *Plasmodium* species (Arisue *et al.*, 2012; Arisue *et al.*, 2019) and one from *L. caulleryi* (Imura *et al.*, 2014). However, there is a general bias in genomic evolution studies of Haemosporidian order, as most researchers focus parasites that infect humans or organisms close to humans. Therefore, the availability of data and analyses of parasites that infect wildlife is limited (Arisue *et al.*, 2012; Imura *et al.*, 2014; Bensch *et al.*, 2016; Böhme *et al.*, 2018; Field *et al.*, 2018; Arisue *et al.*, 2019; Cepeda *et al.*, 2019b).

In addition, the recent publication of omic data from parasites infecting birds (Bensch *et al.*, 2016; Böhme *et al.*, 2018; Field *et al.*, 2018), could demonstrate the importance of this information to improve genetic and evolutionary approaches in Haemosporida order. Therefore, this thesis also aims to generate new genomic information about *Haemoproteus* (subgenos *Haemoproteus*) parasites.

References

Arisue, N., Hashimoto, T., Kawai, S., Honma, H., Kume, K., and Horii, T. (2019). Apicoplast phylogeny reveals the position of *Plasmodium vivax* basal to the asian primate malaria parasite clade. *Scientific reports*, 9(1):7274.

Arisue, N., Hashimoto, T., Mitsui, H., Palacpac, N. M., Kaneko, A., Kawai, S., Hasegawa, M., Tanabe, K., and Horii, T. (2012). The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Molecular biology and evolution*, 29(9):2095–2099.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., et al. (2003). Plasmodb: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic acids research*, 31(1):212–215.

Bennett, G., Garnham, P., and Fallis, A. (1965). On the status of the genera *Leucocytozoon* ziemann, 1898 and haemoproteus kruse, 1890 (Haemosporidiida: Leucocytozoidae and Haemoproteidae). *Canadian Journal of Zoology*, 43(6):927–932.

Bensch, S., Canbäck, B., DeBarry, J. D., Johansson, T., Hellgren, O., Kissinger, J. C., Palinauskas, V., Videvall, E., and Valkiūnas, G. (2016). The Genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome biology and evolution*, 8(5):1361–1373.

Bensch, S., Hellgren, O., and Pérez-Tris, J. (2009). Malavi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Molecular Ecology Resources*, 9(5):1353–1358.

Böhme, U., Otto, T. D., Cotton, J. A., Steinbiss, S., Sanders, M., Oyola, S. O., Nicot, A., Gandon, S., Patra, K. P., Herd, C., et al. (2018). Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome research*, 28(4):547–560.

Bukauskaitė, D., Žiegytė, R., Palinauskas, V., Iezhova, T. A., Dimitrov, D., Ilgūnas, M., Bernotienė, R., Markovets, M. Y., and Valkiūnas, G. (2015). Biting midges (culicoides, diptera) transmit *Haemoproteus* parasites of owls: evidence from sporogony and molecular phylogeny. *Parasites & vectors*, 8(1):303.

Bukauskaité, D., B. R. I. T. and Valkiūnas, G. (2016). Mechanisms of mortality in culicoides biting midges due to *Haemoproteus* infection. *Parasitology*, 143(13):1748–1754.

Cepeda, A. S., Lotta, I. A., Pinto Osorio, D. F., Macías Zapata, J., Valkiūnas, G., Barato, P., and Matta, N. E. (2019a). The experimental characterization of complete life cycle of *Haemoproteus columbae*, with description of natural host-parasite system to study this infection. *International Journal for Parasitology*, submitted.

Cepeda, A. S., Pacheco, M. A., Escalante, A. A., Alzate, J. F., and Matta, N. E. (2019b). *Haemoproteus columbae* apigenome: as an approach for evolutionary and phylogenetic studies of the apicoplast. *In preparation*.

Field, J. T., Weinberg, J., Bensch, S., Matta, N. E., Valkiūnas, G., and Sehgal, R. N. (2018). Delineation of the genera haemoproteus and plasmodium using rna-seq and multi-gene phylogenetics. *Journal of molecular evolution*, 86(9):646–654.

Hellgren, O., Waldenström, J., and Bensch, S. (2004). A new pcr assay for simultaneous studies of *LeucocytozoonI*, *Plasmodium*, and *Haemoproteus* from avian blood. *Journal of Parasitology*, 90(4):797–803.

Imura, T., Sato, S., Sato, Y., Sakamoto, D., Isobe, T., Murata, K., Holder, A. A., and Yukawa, M. (2014). The apicoplast genome of *Leucocytozoon caulleryi*, a pathogenic apicomplexan parasite of the chicken. *Parasitology research*, 113(3):823–828.

Kimura, M. (2008). Understanding avian plasmodium distribution: the role of vector and host.

LaPointe, D. A., Goff, M. L., and Atkinson, C. T. (2005). Comparative susceptibility of introduced forest-dwelling mosquitoes in hawai'i to avian malaria, plasmodium relictum. *Journal of Parasitology*, 91(4):843–850.

Levin, I., Zwiers, P., Deem, S., Geest, E., Higashiguchi, J., Iezhova, T., Jiménez-Uzcátegui, G., Kim, D., Morton, J., Perlut, N., et al. (2013). Multiple lineages of avian malaria parasites (plasmodium) in the galapagos islands and evidence for arrival via migratory birds. *Conservation Biology*, 27(6):1366–1377.

Martinsen, E. S., Perkins, S. L., and Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Molecular phylogenetics and evolution*, 47(1):261–273.

Pacheco, M. A., Matta, N. E., Valkiūnas, G., Parker, P. G., Mello, B., Stanley Jr, C. E., Lentino, M., Garcia-Amado, M. A., Cranfield, M., Kosakovsky Pond, S. L., et al. (2017). Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. *Molecular biology and evolution*, 35(2):383–403.

Valkiūnas, G. (2005). Avian malaria parasites and other haemosporidia CRC press. *Florida, Boca Raton*.

Valkiūnas, G., Zehtindjiev, P., Dimitrov, D., Križanauskienė, A., Iezhova, T. A., and Bensch, S. (2008). Polymerase chain reaction-based identification of *Plasmodium* (*Huffia*) *elongatum*, with remarks on species identity of haemosporidian lineages deposited in genbank. *Parasitology research*, 102(6):1185–1193.

Valkiūnas, G., Žiegytė, R., Palinauskas, V., Bernotienė, R., Bukauskaitė, D., Ilgūnas, M., Dimitrov, D., and Iezhova, T. A. (2015). Complete sporogony of *Plasmodium relictum* (lineage pgrw4) in mosquitoes *culex pipiens pipiens*, with implications on avian malaria epidemiology. *Parasitology research*, 114(8):3075–3085.

van Riper, C., van Riper, S. G., Goff, M. L., and Laird, M. (1986). The epizootiology and ecological significance of malaria in hawaiian land birds. *Ecological monographs*, 56(4):327–344.

Videvall, E. (2019). Genomic advances in avian malaria research. *Trends in parasitology*.

Theoretical framework

Haemosporida order: life cycle and genome

The Haemosporida order is a group of protozoan parasites that infect different classes of the animal kingdom and includes more than 500 morphological species organized into four families: Garniidae, Haemoproteidae, Leucocytozoidae, Plasmodiidae (Valkiūnas, 2005; Adl *et al.*, 2012).

This order presents a complex life cycle, which alternates between a vertebrate host (haploid phase) and an invertebrate vector (diploid phase). Sporozoites are the infective form that enters into vertebrate host when a vector insect takes blood meal. These forms have tropism by diverse tissues that include liver, spleen, brain, lung, among others (Valkiūnas, 2005). After some tissue cycles, the parasites come out to blood and infect mature or immature blood cells depending on species.

The final phase culminates with the formation of gametocytes named microgametocytes and macrogametocytes that circulate in blood. The fertilization occurs in the invertebrate when the vector feeds on infected blood with these parasites; the microgametocytes generate male gametes process called exflagellatio (Valkiūnas, 2005; Coral *et al.*, 2015), which fertilize the female gamete (from macrogametocytes)), forming the zygote.

The zygotes transform into mobile forms called ookinetes, which are lodged in the intestinal epithelium of the vector forming oocysts. The oocysts begin a meiotic cycle followed by a mitotic cycle, which the latter generates the sporozoites, which will travel to the glands of the vector (Valkiūnas, 2005).

Haemosporida genomes are usually small (20 to 33.6mb; Fig. **5-1**) and arranged into 14 chromosomes; the length of each chromosome is characteristic of each species. In addition, most of these genomes have a low GC content, being approximately between 19 and 30 % (GC content bias; Fig. **5-1**; Kissinger and DeBarry, 2011; Carlton *et al.*, 2013; Bensch *et al.*, 2016; Böhme *et al.*, 2018). However, species such as *P. vivax* and *P. knowlesi* have a higher GC content, around 40 % (Fig. **5-1**; Rutledge *et al.*, 2017).

Recently Boehme *et al.*, (2018) found evidence of transposable elements in *Plasmodium* genomes that infect birds; sequences that have not been reported in *Plasmodium* genomes that infect mammals (Fig. **5-1**; Carlton *et al.*, 2013).

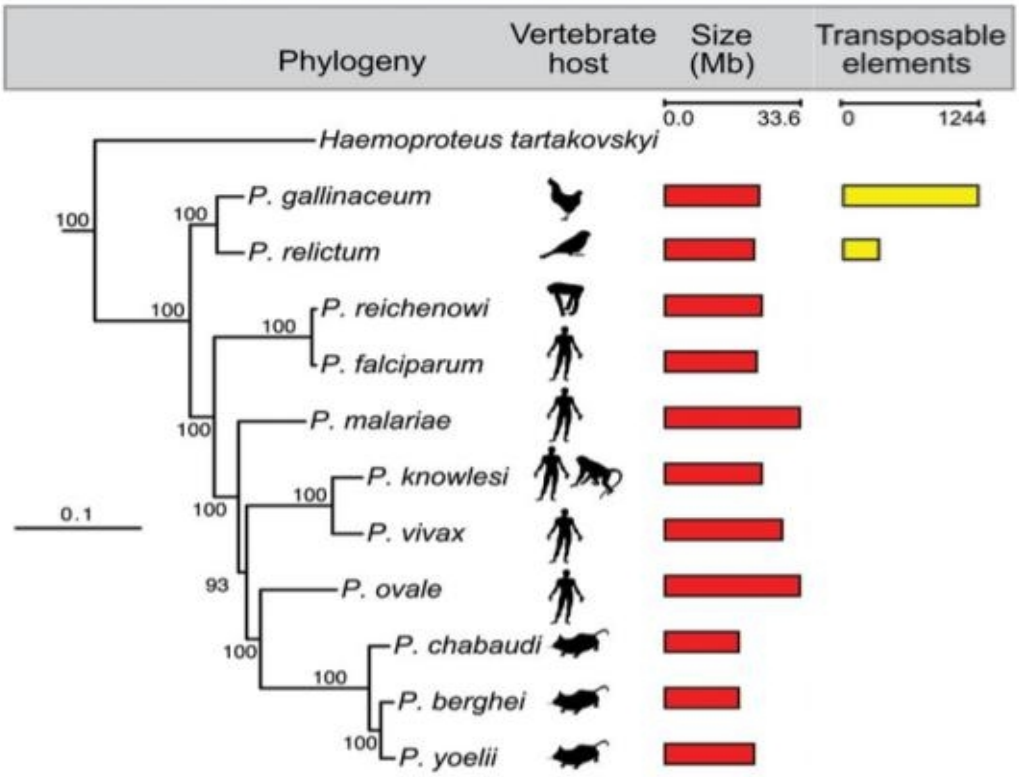


Figure 5-1.: Graphical overview of key features of all Haemosporida reference genomes. Figure taken and modified from Boehme *et al.*, (2018)

Haemoproteus genus

Haemoproteus genus belongs to family Haemoproteidae and about 150 species have been described (Valkiūnas, 2005; Iezhova *et al.*, 2011). *Haemoproteus* has wide distribution, including countries with cold climates (Oakgrove *et al.*, 2014). *Haemoproteus* spp. are able to infect wild, domestic and poultry birds (Saif *et al.*, 2003), causing serious pathologies (Olias *et al.*, 2011) that manage to affect host fitness, parental care, increase probability of predation, and even death (Ferrell *et al.*, 2007; Møller and Nielsen, 2007; Islam *et al.*, 2014). Therefore, the study of this parasitic infection has been intensified in order to determine economic losses and to generate control strategies (Islam *et al.*, 2014).

Similar to *Plasmodium* parasites, *Haemoproteus* is recognized morphologically by its production of hemozoin granules (malarial pigment) in host blood cells. Bennett *et al.,* (1965) propose to divide this genus into two sugberena, *Haemoproteus* and *Parahaemoproteus*, differentiating themselves from each other by ecological features, life history , morphology and vectors implied in their transmission. *Haemoproteus* subgenus infects mainly Columbifor-

mes and fregates belonging to Pelecaniformes and Charadriiformes (Valkiūnas, 2005; Levin *et al.*, 2011; Levin *et al.*, 2012), and is transmitted by diptera of the Hippoboscidae family. These findings on infections in endemic bird species by Haemoproteus subgenus can have devastating consequences on biodiversity and wildlife, as occurred recently with the arrival of *P. relictum* in Hawaii (van Riper *et al.*, 1986).

By contrast, *Parahaemoproteus* subgenus infects birds that do not belong to the Columbiform order, and is transmitted by diptera of family Ceratopogonidae (Valkiūnas, 2005). The division in two subgenera is soported by molecular tools (Martinsen *et al.*, 2008).

Genomic assembly

Genome is understood as "*the complete group of sequences in the genetic material of an organism. Which includes the sequence of each chromosome, plus any DNA present in organelles such as mitochondria*" (Brown, 2008). On the other hand, the approximation or draft genome refers to "*a genomic DNA sequence with less accuracy than the final sequence; some segments are missing or in the wrong order or orientation*" (MGI, 2017).

The assembly is assumed as the "*computational reconstruction of a long sequence from multiple reads of small sequences*"(Ekblom and Wolf, 2014); this groups of reads are grouped in contigs and contigs, in scaffolds. The contigs provide a multiple alignment of the reads and the consensus sequence; through the scaffolds, the order, orientation and size of the gaps are obtained between the contigs (Miller *et al.*, 2010). There are two types of assembly approach: *de novo* and comparative. The first one refers to the reconstruction of a contiguous sequences without making use of a reference genome, while the comparative assembly uses a reference genome as a guide (Ekblom and Wolf, 2014).

Additionally, the assembly is based on graph algorithms. A graph is the representation of a set of objects through nodes or vertices that are connected by edges (Miller *et al.*, 2010). The different paths that can be formed between the nodes have been considered for genome sequences assembly such as the Hamiltonian path (each node is visited only once) and the Eulerian path (each edge is visited only once; El-Metwally *et al.*, 2014). The three types of assembly algorithms are: Overlap/Layout/Consensus (OLC), Greedy graphics and Bruijin graphics.

OLC algorithm is based on finding the Hamiltonian path, originally used for the assembly of data from sequencing by Sanger, later it was optimized for large genomes and used in the assemblers of Celera, Arachne, CAP and PCAP. These assemblers perform a pre-calculation

through all the reads to select the overlapping candidates according to which it uses the *k-mer* identified as seeds of alignment, and elaboration of the overlap chart through which the consensus sequence is obtained (Miller *et al.*, 2010).

Greedy algorithm looks for sub-strings of whole set of reads with maximum score, and it calculates alignments pairs of all the fragments choosing the fragments with the greater overlapping, then joins the different fragments that previously it has evaluated. The above procedure is performed until the assembly is complete. According to El-Metwally *et al.,* (2014), the main problem with this algorithm is that it gets stuck in local maxima. Some assemblers with the focus of the Greddy algorithm are SSAKE and SHARCS.

Bruijin algorithm is based on the Eurelian path, in a scenario where the reads do not present errors, therefore, through a *k-mer*, the construction of the graph is performed and consensus sequence (Compeau *et al.*, 2011). Velvet, SPAdes, ABySS and Euler are some assemblers base on Bruijin graph algorithm (El-Metwally *et al.*, 2014).

Structural and functional genome annotation

Genomic annotation occurs at two levels: structural and functional. The first one consists of searching for biologically relevant sites, determining a coherent model for the whole assembled sequence in which each target is properly defined and each component of the object has a unique location. The second level corresponds to processing information, it is consisting of attributing specific and relevant information to the assembled sequence, for example, molecular function, biological function, metabolic function of each structurally annotated gene (Rouzé *et al.*, 1999).

Besides, structural annotation has as its main objective the search or prediction of genes, in this case only those coding for proteins, encoding structural RNAs or simply comments on the sequence. The structural annotation will recognize genes, their locations in the assembled sequences, the structure of the genes (promoter, UTR, start codon, exons, introns and stop codon). There are three methods to carry out the structural annotation of genomic sequences:

1. *ab initio*: this method uses only the properties of the sequence to predict the location of genes, based on algorithms that discriminate coding and non-coding regions, through the presence of open reading frames (ORF) to infer where the gene is located.

2. homology: this method is based on use of algorithms (such as BLAST) to deduce the location and structure of genes from the comparison of assembled genomic sequences

with databases.

3. hybrid: it makes use of the two previous methods.

Likewise, structural annotation is able to occur at the level of nucleotides, answering the questions: where are the genes located ? And at the level of proteins: what genes are present in assembled genomic sequences? (Rouzé *et al.*, 1999).

Functional annotation of genes refers to the comparison and statistical analysis of several lists of genes, which by statistical methods identifies functional annotations which the analyzed genes are significantly related (Rouzé *et al.*, 1999). Like structural annotation, the functional annotation can be carried out by 3 different methodologies:

1. Over-representation analysis (ORA): it checks statistical over-representation of a list of interest genes in a reference list. Methods such as Fisher's exact one-tailed test or hypergeometric distribution are used.

2. Gene set enrichment analysis (GSEA): it incorporates the expression values, FC values or *p values* of all the genes to test of statistical significance analysis.

3. Integrative and modular enrichment analysis (IMEA): it takes into account the dependencies among genes inferred from biological networks, ontologies graphs or combinations of different types of annotations.

References

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, 59(5):429–514.

Bennett, G., Garnham, P., and Fallis, A. (1965). On the status of the genera *Leucocytozoon* ziemann, 1898 and haemoproteus kruse, 1890 (Haemosporidiida: Leucocytozoidae and Haemoproteidae). *Canadian Journal of Zoology*, 43(6):927–932.

Bensch, S., Canbäck, B., DeBarry, J. D., Johansson, T., Hellgren, O., Kissinger, J. C., Palinauskas, V., Videvall, E., and Valkiūnas, G. (2016). The Genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome biology and evolution*, 8(5):1361–1373.

Böhme, U., Otto, T. D., Cotton, J. A., Steinbiss, S., Sanders, M., Oyola, S. O., Nicot, A., Gandon, S., Patra, K. P., Herd, C., et al. (2018). Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome research*, 28(4):547–560.

Brown, T. (2008). *Genomas/Genome*. Ed. Médica Panamericana.

Carlton, J. M., Das, A., and Escalante, A. A. (2013). Genomics, population genetics and evolutionary history of plasmodium vivax. *Adv Parasitol*, 81:203–222.

Compeau, P. E., Pevzner, P. A., and Tesler, G. (2011). How to apply de bruijn graphs to genome assembly. *Nature biotechnology*, 29(11):987–991.

Coral, A. A., Valkiūnas, G., González, A. D., and Matta, N. E. (2015). In vitro development of *Haemoproteus columbae* (haemosporida: Haemoproteidae), with perspectives for genomic studies of avian haemosporidian parasites. *Experimental parasitology*, 157:163–169.

Ekblom, R. and Wolf, J. B. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary applications*, 7(9):1026–1042.

El-Metwally, S., Ouda, O. M., and Helmy, M. (2014). *Next generation sequencing technologies and challenges in sequence assembly*, volume 7. Springer Science & Business.

Ferrell, S. T., Snowden, K., Marlar, A. B., Garner, M., and Lung, N. P. (2007). Fatal hemoprotozoal infections in multiple avian species in a zoological park. *Journal of Zoo and Wildlife Medicine*, 38(2):309–316.

Iezhova, T. A., Dodge, M., Sehgal, R. N., Smith, T. B., and Valkiūnas, G. (2011). New avian haemoproteus species (haemosporida: Haemoproteidae) from african birds, with a critique of the use of host taxonomic information in hemoproteid classification. *Journal of Parasitology*, 97(4):682–694.

Islam, M. S., Alim, M. A., Das, S., Ghosh, K. K., Pervin, S., Lipi, A., Siddiki, A. Z., Masuduzzaman, M., and Hossain, M. A. (2014). Prevalence of haemoproteus sp in domestic pigeon at chittagong and khulna district in bangladesh. *J Adv Parasitol*, 1:24–26.

Kissinger, J. C. and DeBarry, J. (2011). Genome cartography: charting the apicomplexan genome. *Trends in parasitology*, 27(8):345–354.

Levin, I. I., Valkiūnas, G., Iezhova, T. A., O'brien, S. L., and Parker, P. G. (2012). Novel *Haemoproteus* species (haemosporida: Haemoproteidae) from the swallow-tailed gull (lariidae), with remarks on the host range of hippoboscid-transmitted avian hemoproteids. *The Journal of parasitology*, pages 847–854.

Levin, I. I., Valkiūnas, G., Santiago-Alarcon, D., Cruz, L. L., Iezhova, T. A., O'Brien, S. L., Hailer, F., Dearborn, D., Schreiber, E., Fleischer, R. C., et al. (2011). Hippoboscid-transmitted haemoproteus parasites (haemosporida) infect galapagos pelecaniform birds: Evidence from molecular and morphological studies, with a description of haemoproteus iwa. *International Journal for Parasitology*, 41(10):1019–1027.

Martinsen, E. S., Perkins, S. L., and Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Molecular phylogenetics and evolution*, 47(1):261–273.

Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6):315–327.

Møller, A. P. and Nielsen, J. T. (2007). Malaria and risk of predation: a comparative study of birds. *Ecology*, 88(4):871–881.

Oakgrove, K. S., Harrigan, R. J., Loiseau, C., Guers, S., Seppi, B., and Sehgal, R. N. (2014). Distribution, diversity and drivers of blood-borne parasite co-infections in alaskan bird populations. *International journal for parasitology*, 44(10):717–727.

Olias, P., Wegelin, M., Zenker, W., Freter, S., Gruber, A. D., and Klopfleisch, R. (2011). Avian malaria deaths in parrots, europe. *Emerging Infectious Diseases*, 17(5):950.

Rouzé, P., Pavy, N., and Rombauts, S. (1999). Genome annotation: which tools do we have for it? *Current opinion in plant biology*, 2(2):90–95.

Rutledge, G. G., Böhme, U., Sanders, M., Reid, A. J., Cotton, J. A., Maiga-Ascofare, O., Djimdé, A. A., Apinjoh, T. O., Amenga-Etego, L., Manske, M., et al. (2017). Plasmodium malariae and p. ovale genomes provide insights into malaria parasite evolution. *Nature*, 542(7639):101.

Saif, Y., Barnes, H., Glisson, J., Fadly, A., McDougald, L., and Swayne, D. (2003). *Diseases of Poultry*. Wiley.

Valkiūnas, G. (2005). Avian malaria parasites and other haemosporidia CRC press. *Florida, Boca Raton*.

van Riper, C., van Riper, S. G., Goff, M. L., and Laird, M. (1986). The epizootiology and ecological significance of malaria in hawaiian land birds. *Ecological monographs*, 56(4):327–344.

The experimental characterization of complete life cycle of *Haemoproteus columbae*, with description of natural host-parasite system to study this infection

Axl S. Cepeda[a], Ingrid A. Lotta-Arévalo[a], David F. Pinto-Osorio[a], Jhon Macías-Zacipa[a,b], Gediminas Valkiūnas[c], Paola Barato[d], Nubia E. Matta[a]

a. Departamento de Biología, Grupo de Investigación Caracterización Genética e Inmunología, Sede Bogotá-Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia.

b. Programa Bacteriología y Laboratorio Clínico, Facultad de Ciencias de la Salud, Universidad Colegio Mayor de Cundinamarca. Bogotá, Colombia.

c. Nature Research Centre, Vilnius, Lithuania.

d. Corporación Patología Veterinaria, CORPAVET, Bogotá, Colombia.

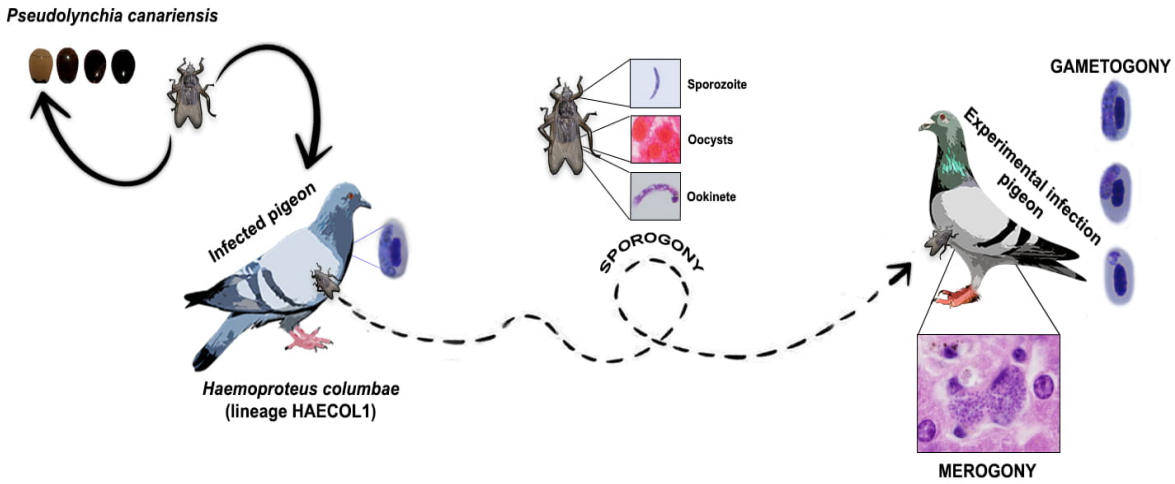


Figure 6-1.: Graphical Abstract

6.1. Abstract

Characterization of complete life cycles of haemoparasites requires maintaining suitable susceptible vertebrate hosts and vectors for long periods in captivity, in order to follow the complete parasitic cycle. Such studies require to follow the development of different parasite stages in definitive and intermediate hosts. Currently, there are few host-parasite models established in avian haemosporidian research, and they have been developed mainly for Passeriformes species and their parasites. This study aimed at developing an experimental methodology to access the complete life cycle of *Haemoproteus columbae* (cytb lineage HAECOL1), which parasitize the Rock Pigeon *Columba livia* and *Pseudolynchia canariensis* (louse fly). A colony of louse flies (Hippoboscidae), which are the

natural vectors of this parasite, was established. Thirty newly emerged insects were exposed to *H. columbae* infection and used to infect naïve Rock Pigeons. The peak of parasitaemia (acute stage) was seen between 27 to 32 dpi when up to 70 % of red blood cells were infected. The crisis occurred approximately one week after the peak, and the long-lasting chronic parasitaemia stage was followed. Exo-erythrocytic meronts were seen mainly in the lungs where extensive tissue damage was reported, but also in the kidneys and spleen. In vector, the sporogonic cycle of *H. columbae* was completed between 13 to 16 days post infection (dpi) at the average temperature ranging between 12 and 15°C. This host-parasite model is tractable to maintain in captivity. It is recommended to use in studies aiming for detailed characterization of host-parasite relationships in areas such as physiology, pathology, immunobiology, genetics as well as for evaluative treatments and to follow the infection in any stage of parasite development both in the vertebrate or invertebrate host.

6.2. Introduction

Haemoproteus species (Haemosporida, Haemoproteidae) are vector-borne parasites widely distributed in birds worldwide (Thomas *et al.*, 2008; Valkiūnas, 2005). There are approximately 170 described species in the genus There are approximately 170 described species in the genus *Haemoproteus*, which are classified into two subgenera based not only on their genetic differences but also, in the vectors where the sexual development take place (Martinsen *et al.*, 2008). The parasites that belong to the subgenus *Parahaemoproteus* are transmitted by biting midges (Culicoides), while Hippoboscidae species transmit parasites of the subgenus *Haemoproteus* (Adie *et al.*, 1915; Baker, 1957; Valkiūnas, 2005; Valkiūnas *et al.*, 2010).

The sexual process and sporogonic cycle of *H. columbae* occur in *Pseudolynchia canariensis* (louse fly; Adie *et al.*, 1915; Valkiūnas, 2005). In these louse flies, the eggs hatch into the mother's uterus, and there three stages of larvae development take place. The larvae feed on nutrient fluids secreted by paired milk glands in the louse fly uterine wall until they become pupae (Bishopp, 1929; Baker, 1967; Harwood *et al.*, 1979;). Female flies deposit pupae on the substrate in or around pigeon nests (Bishopp, 1929; Waite *et al.*, 2012a).

The sporogonic cycle of Haemoproteus species in flies takes between 6.5 and 10 days until parasites reach the salivary glands (Adie, 1915; Baker, 1966). Coral *et al.,* (2015;) demonstrated that exflagellation *in vitro* occurs as quickly as 3 min at 40°C AEA (after exposure of infected blood to air), when micro- and macrogametes appear; zygotes appear 5 min AEA, developing ookinetes 45 min AEA and mature ookinetes were observed 20 hours AEA. Gallucci (1974;), indicates that there is no way to distinguish the ookinetes of *H. columbae* developing in vitro from those differentiating *in vivo*, but different sizes of these mature structures have been reported in other studies. *In vivo* the oocysts have been seen protruding outside of the midgut wall when mature (Baker, 1957;); they are usually larger than 30µm in diameter and contain numerous germinative centres; this may partly explain why the sporogonic cycle of the Haemoproteus parasite is longer than that of the

Parahaemoproteus species, in which tiny oocysts (< 15 µm in diameter) with one germinative centre occur (Adie *et al.*, 1915; Valkiūnas, 2005). In louse flies, the salivary glands are tightly packed and wrapped by the digestive tract (Adie *et al.*, 1915). Sporozoites (infective stages for birds) are elongate bodies that appear 10 to 12 days post-infection (dpi; Adie *et al.*, 1915)). Histological changes (disruption of basal membranes and inflammation) have been reported in heavily infected salivary glands of louse flies (Klei and De Giusti, 1973).

The merogony and development of gametocytes of *H. columbae* occur in its natural host, the Rock Pigeon *Columba livia* (Valkiūnas, 2005). This bird species is of cosmopolitan distribution and is widely raised as domesticated ornamental birds. In Colombia, these pigeons are considered an invasive species (Baptiste *et al.*, 2010) and/or a pest that even may cause health problems in humans due to the high number of associated pathogens (Villalba-Sánchez and J., 2014).

Haemoproteus parasites do not multiply in peripheral blood, and for that reason, it is unlikely that they will infect a new avian host by direct inoculation of blood, as it readily occurs in *Plasmodium* species (Valkiūnas, 2005). To access the infective stage (sporozoites) of haemoproteids, the sporogonic development must be followed in the louse fly vector. Sporozoites of the parasites reach the salivary glands of the vector, which inoculates them to birds during feeding. Another non-natural alternative to achieve an infection by *Haemoproteus* parasites is the inoculation of either the sporozoites from crushed infectious flies (Ahmed and Mohammed, 1978), or, the mature tissue stages of the parasites (Atkinson, 1986) in the susceptible recipient avian host, but this mode of infection is difficult to achieve in practise.

Currently, there are few animal models available to study and characterize the host-parasite-vector interactions in avian haemosparidia. For instance, the interactions between avian malaria parasites (*Plasmodium* species) and their hosts have been studied mainly using two standardized models which involves *Culex* (*Cx.*) *quinquefasciatus* and *Cx. pipiens* mosquitoes to transmit infections to *Serinus canaria* or *Agelaius phoeniceus*. However, in the past other models involved chicken and ducks (LaPointe *et al.*, 2005; Valkiūnas *et al.*, 2015). Additionally, various wild birds and biting midges of the genus Culicoides (mainly *Culicoides impunctatus* and *C. nubeculosus*) were used to study sporogony of some species of the subgenus *Parahaemoproteus* (Bukauskaitė *et al.*, 2015; Bukauskaité and Valkiūnas, 2016).

These animal models are tractable for using in experimental research due to the availability of both avian host and vectors that can be maintained under controlled laboratory conditions. This provides opportunities to sample parasitological material during experimental exposure of hosts. Model organisms to access development of haemosporidian parasites of the subgenus *Haemoproteus* remain non-accessed in experimental research. The main aims of this study were: 1) to develop a use-able methodology to use for experimental research with *H. columbae* using its natural avian host (Rock Pigeon) and its vector (louse fly) and 2) to follow the complete life cycle of *H. columbae*

(*cytb* lineage HAECOL1) in experimentally infected insects and birds.

6.3. Materials and methods

6.3.1. Ethical considerations

The Bioethics Committee (Facultad de Ciencias of the Universidad Nacional de Colombia Act number: 04 of 2017 and 03 of 2018) approved the methodology used in this study. Fieldwork was done under permit No. 0255 granted by Autoridad Nacional de Licencias Ambientales (ANLA).

6.3.2. Parasite lineage isolation and characterization

The *cytochrome b* lineage of *H. columbae* HAECOL1 was found in a naturally infected *C. livia* (pigeon No. 20) captured in Bogotá city-Colombia. In order to maintain the same clone of the parasite lineage, three naïve pigeons were infected experimentally by louse fly bites infected with the HAECOL1 lineage (Table **6-1**; Fig. **6-2**).

6.3.3. Cloning the infection of *Haemoproteus columbae* (lineage HAECOL1)

The exposure procedures of Rock Pigeons and louse flies to HAECOL1 lineage were as follows. Experimental pigeons were purchased from a pet shop (Table **6-1**). Once in the laboratory, pigeons were screened for haemoparasites by microscopy and PCR. On the other hand, eighteen naïve adult louse flies (six per pigeon) were used to transmit the parasite to three naive pigeons (No. 68, 70 and 71). They were left in starvation from 12 to 14h in the incubator at 28°C with 30-40 % relative humidity. Then, the louse flies were allowed to feed during 24h on No. 20, the pigeon harbouring mature gametocytes of the HAECOL1 lineage (2.4 % of parasitaemia). At this parasitaemia, it was estimated that one of six individuals got the infection (Table **6-3**). Preening was avoided by using restraint collars (Elizabethan bird collars) that were left until the end of the experiments. It is important to mention that despite the use of such collars, birds were able to feed and move freely. Then, all louse flies were collected manually and placed on uninfected pigeons (naïve) where they were maintained permanently. The entire sporogonic cycle occurred in the insects during their normal process of feeding, and the insects infected Rock Pigeons by natural bites when sporogony was completed. This procedure mimics the natural mode of infection and allows to clone the same lineage of the parasite (HAECOL1) in new individual pigeons. The last procedure guarantees maintenance of the same parasite lineage (HAECOL1) in its natural host.

6.3.4. Rock-pigeon sample and maintenance

The common Rock Pigeons were maintained in an outdoor aviary of the Biology Department-Universidad Nacional de Colombia, where the average year-round temperature ranged between 12

and 15°C (Bernal *et al.*, 2007). Each bird was kept separately in a cage (50x50x50 cm) under a natural photoperiod (approximately 12 h of dark and 12 h light). Likewise, a silk net covered each cage to avoid contamination with external insects.

Eleven Rock Pigeons were purchased in a pet store and used in the present study as described in Table **6-1**. Once the individuals arrived at the laboratory, and before any experiment, pigeons were tested for possible natural infection with Haemoproteus or another blood parasite using microscopic and molecular methods every week during a month period. These pigeons were ringed and fed three times daily, and water was provided ad libitum according to the requirements of this species (Soto and Acosta, 2010). Main procedures of the experimental assays are shown in Fig.**6-2**.

Rock Pigeon Identification (number of individual)	Use in the experiment
GERPH-UN820 (20)	Original pigeon infected with HAECOL1, the source of the lineage
GERPH-UN868* (68); GERPH UN870 (70) GERPH-UN871*(71)	To follow the infection in the pigeons
GERPH-UN870 (70); GERPH-UN875 (75)	To follow the infection in the louse flies
GERPH-UN873 (73); GERPH-UN869 (69)	Negative controls
And four pigeons without number	Naïve pigeons to maintain the colony of louse flies

Table 6-1.: Identification of *C. livia* used in the present study. *GERPH-UN871 died at 64 dpi, GERPH-UN868 was euthanized 33 dpi

6.3.5. Microscopic examination and PCR protocols for detection of *Haemoproteus columbae*

Pigeons were bled from the brachial vein. About 80µL of whole blood was taken in heparinized micro haematocrit tubes (NRIS, vitrex medical A/S Ref 161315) to prepare three smears and the remainder was stored in EDTA for molecular analysis. Additionally, another 80µL of blood was collected in capillary tubes and centrifuged 5min at 5.000rpm (Scientific, Model HC-12A-Zenith Lab INC, USA) and a Haematocrit Reader Card was used for haematocrit estimation.

Three blood smears were fixed in methanol and stained using 4 % Giemsa solution (Valkiūnas, 2008). Parasitaemia was estimated by counting the number of parasites per 10000 erythrocytes. Blood films for microscopic examination were prepared daily starting from the first-day until 64 days post-infection (dpi) to follow the course of parasitaemia. Uninfected pigeons (controls) were tested once a week. *H. columbae* was identified according to Valkiūnas (2005). DNA from the blood was extracted using DNeasy Blood & Tissue kit (Qiagen, GmbH, Hiden, Germany) and tested for *H. columbae* by amplifying 480 bp of parasite mitochondrial *cytochrome b* gene (*cytb*) according to Hellgren et al., (2004). The amplifications were evaluated by running 1.5µL of the final PCR product on a 1.5 % agarose gel.

6.3.6. Histopathological analysis

The pigeon No. 68 was euthanized at 33 dpi with the aim to evaluate the impact of the infection in selected tissues and to identify sites of exo-erythrocytic merogony. Brain, heart, kidney, liver, lungs, and spleen were fixed in 10 % buffered formalin and embedded in paraffin for HE routine staining. A veterinary pathologist (P.B.) evaluated the case with an Olympus BX43 light microscope. Tissue structures compatible with meronts and lesions in organs were reported. Digital images of parasites were taken using an Olympus DP27 digital camera coupled to the microscope and processed with the cellSens™ Microscope Imaging Standard software (Olympus, Tokyo, Japan).

6.3.7. Collection and maintenance of louse flies infected with *H. columbae* (HAECOL1)

Louse flies were captured by hand from pigeons in Bogotá, Colombia (N 4°35'53" W 74°4'33"), and they were removed manually from more than 100 feral Rock Pigeons, placed in rearing silk mesh boxes (15x15x15cm) and transported to the laboratory. The collected insects were placed on a non-infected caged pigeon for maintenance and reproduction. Since female louse fly can produce one pupa every two days after they lay their first pupa (Herath, 1966), birdcages were examined once a week looking for pupae that were mainly found in the cage litter. Pupae were placed in containers and maintained in the incubator at 28°C and 30-40 % relative humidity (Memmert model INB400, Germany) until the emergence of the imagos. To increase the number of louse flies in the colony, all emerged adults were maintained on uninfected pigeons and used as the parental for further reproduction, as described above.

The louse flies were identified to species level using morphological characters and the key by Hutson (1984). The species identity was confirmed by amplification of a 658 bp barcode fragment of *Cytochrome Oxidase I* (*co1*) using the primers LCO1490/HCO2198 (Vrijenhoek, 1994) according protocol by Colorado-Garzón *et al.* (2016).

6.3.8. The course of infection in louse flies

Twenty-three newly emerged non-infected louse flies (reared in the laboratory from pupae) were distributed in 4 cohorts and infected in a period of 4 consecutive days, one cohort per day. Louse flies belonging to each cohort were marked in the wings with a distinctive color (2012a), and left in starvation for 12-14h in an incubator at 28°C with 30-40 % of relative humidity. Then, they were allowed to take blood meals on an infected pigeon N° 70 (**??**), whose parasitemia ranged between 1.3 and 2.0 % during experiments. Detailed information about the number of individuals tested in each group and the corresponding parasitemia of blood donor is provided in Table **6-3**. After 24 h, insects were gathered from the infected pigeon No. 70 and transferred to a cage with one naïve pigeon (No. 75). A maximum of 3 individuals were dissected every 2 or 3 days until 16 dpi. The louse flies were dissected in order to follow the sporogonic development of the parasite according to the

protocols suggested by Adie (1915). Dissections were carried out using a Carl Zeiss™ Stemi™ DV4 binocular stereo microscope (Oberkochen, Alemania). Salivary glands are tightly packed along with the intestine in the abdomen, thus, using a blade, an incision was made in the posterior segments, following the middle line. Then, abdomen contents were gently pulled out using an entomological needle. Furthermore, to extract the goblet-shaped organ, which may contain sporozoites, a thin cross-section of the ventral chitin plate of the thorax was performed using a sterile razor blade. Thin films of midgut contents and salivary glands were prepared, fixed in absolute methanol and stained with Giemsa as blood films. The midguts were stained with Mercurochrome 2 % for 10-15 min and then examined under the microscope. If oocysts were found, permanent preparations were performed; entire midguts were fixed in formalin, stained with Ehrlich's hematoxylin and mounted in diluted Canada balsam following the protocols suggested by Valkiūnas (2005), Kazlauskienė *et al.,* (2013) and Bukauskaitė *et al.*, (2015). Images were prepared from both haematoxylin and Mercurochrome stained preparations.

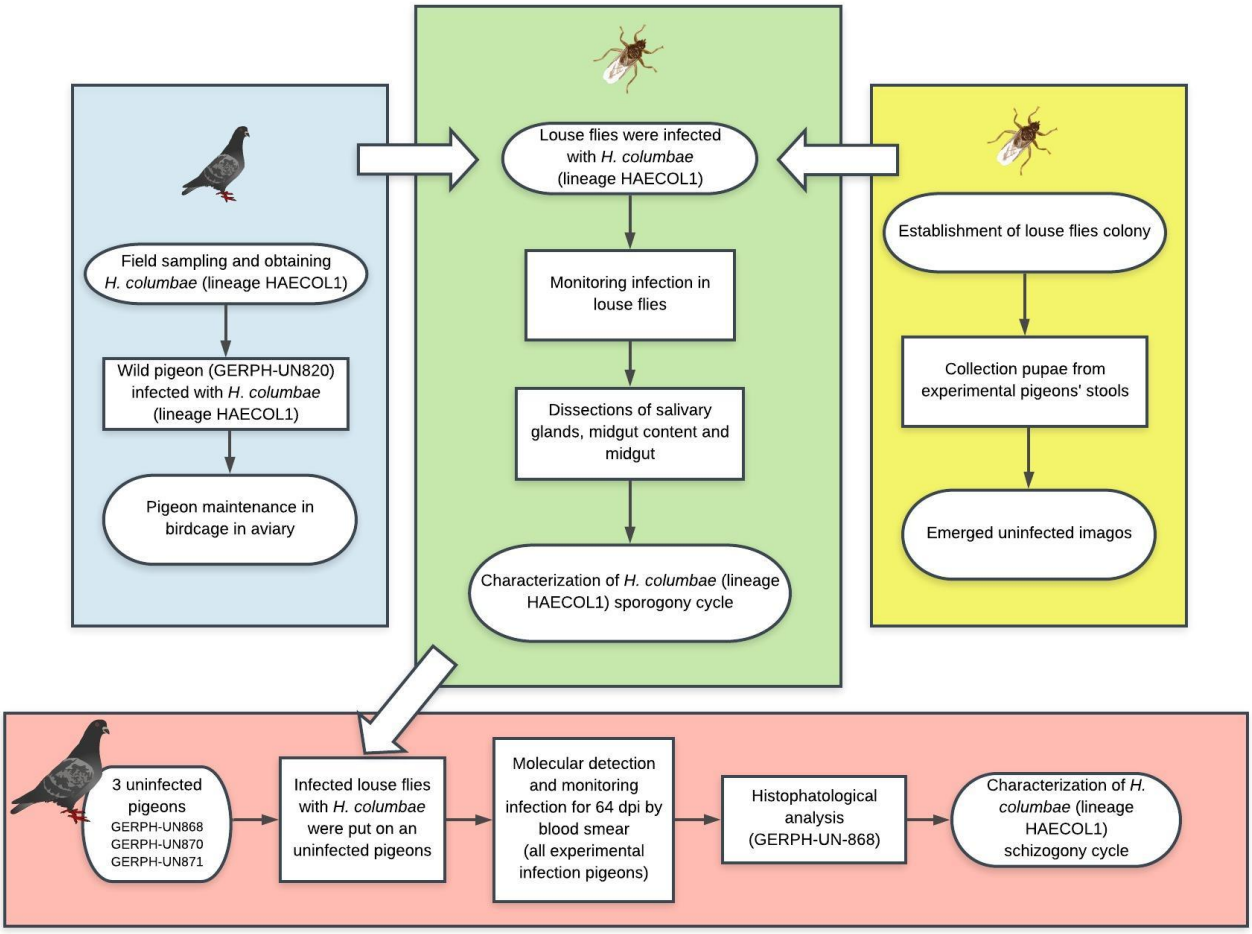


Figure 6-2.: Diagrammatic representation of the experimental approaches, which were used for the study of *H. columbae* life cycle in *C. livia* and *P. canariensis*.

6.3.9. Microscopic examination of vector preparations and parasite morphology

An Olympus BX43 light microscope was used to examine preparations, prepare illustrations, and to take measurements. All preparations were first examined at low magnification (400x) and then at high magnification (1000x). Digital images of parasites were taken using an Olympus DP27 digital camera coupled to the microscope and processed with the cellSens™ Microscope Imaging Standard software (Olympus, Tokyo, Japan). The morphometric features studied were those defined by Valkiūnas (2005). Voucher specimens of ookinetes (GERPH-UNI002:HAE; GERPH-UNI001:HAE), oocysts (GERPH-UNI006:HAE; GERPH-UNI0067:HAE) and sporozoites (GERPH-UNI022) of *H. columbae* lineage HAECOL1, as well as the blood films, and histological preparations (biological record ID: UNAL:GERPH:UN868:HAE), were deposited in the Biological collection GERPH, Biology Department, Universidad Nacional de Colombia Bogotá-Colombia.

6.4. Results

6.4.1. The establishment of a natural model system for maintenance of *H. columbae* infection

We developed a methodology to establish a natural avian host-parasite-vector model to maintain haemoproteid parasites belonging to the subgenus *Haemoproteus*. The described procedures were tested several times and allowed 1) to obtain *H. columbae* infected avian hosts, 2) to sample sufficient numbers of infected and non-infected louse flies, 3) to infect louse flies and to follow the complete sporogonic development of *H. columbae*, and 4) to infect naïve birds and to follow the entire life cycle in the natural vertebrate host. This methodology and host parasite model open opportunities for precise investigations of various aspects of the biology of haemosporidian infections and access more details about the development of *H. columbae* (HAECOL1) infection. Application of molecular and morphological tools in parallel indicated that the louse flies *P. canariensis* are cosmopolitan ectoparasite of Columbiformes birds. Also, these flies are competent vectors of the lineage HAECOL1 and are tractable to use in experimental vector research.

6.4.2. The course infection and dynamics of parasitaemia of *H. columbae* (HAECOL1) in pigeons

Prepatent period was 19-20 dpi. The acute phase, determined by the continuous production and constant accumulation of immature gametocytes with a parasitaemia lower than 15 % (Ahmed and Mohammed, 1978), started between 22 and 25 dpi in different individual birds (Fig. **6-3**). After that, crisis characterized by an accelerated elimination of parasites present in blood (Ahmed and Mohammed, 1978) occurred, and parasitaemia dropped sharply from approximately 29-30 dpi and turned to the chronic stage within the next 7-8 days (Fig. **6-3**).

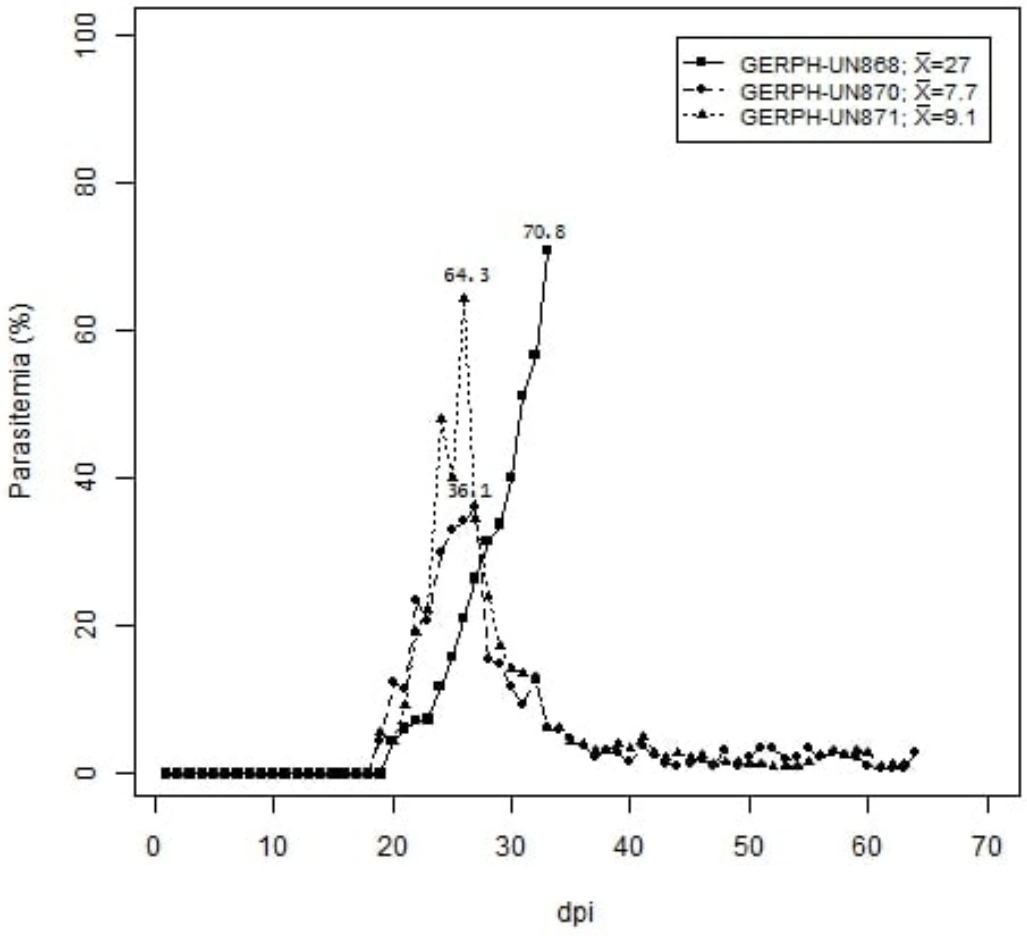


Figure 6-3.: Dynamics of *H. columbae* parasitemia in three experimentally infected rock pigeons *C. livia* (GERPH-UN868, GERPH-UN870, and GERPH-UN871).

Fig. **6-4** shows the development of gametocytes of *H. columbae* (HAECOL1) in the pigeon GERPH-UN868. The immature gametocytes (Fig. **6-4** A-D, G-H) predominated in the beginning of the acute stage. The first fully grown mature gametocytes (as determined by displacement of host cell nuclei and presence of pigment granules on the ends of parasite; Fig. **6-4** E-F, I) were seen 3-5 days after the microscopical detection of parasites in the blood. The proportion of macro- to microgametocytes was 1:3 during the course of infection. Multiple infections of the same red blood cell with several growing gametocytes were common during high parasitaemia (Fig. **6-4** G-I), but not during the chronic infection stage.

Intensity of parasitaemia varied among pigeons and dpi, e.g. the peak of parasitaemia for pigeon No. 68 was 70.8 % (33 dpi), but in pigeon No. 70 it was of 36.1 % (28 dpi) and for No. 71 was 64.3 % (27 dpi; **6-3**). The pigeon No. 71 died 64 dpi with a parasitaemia of 1.4 % and haematocrit value

36 % (Table **6-2**). The average of haematocrit value for the two negative controls was 51 %.

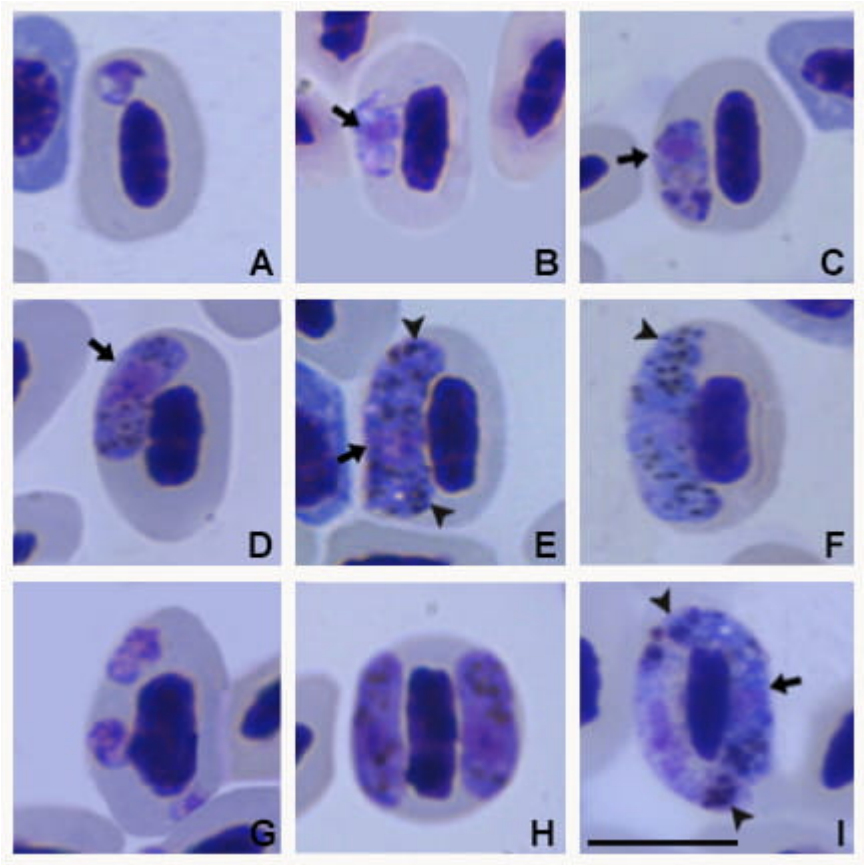


Figure 6-4.: Development of gametocytes of *H. columbae* (lineage HAECOL1) in experimentally infected the rock pigeon GERPH-UN868. Immature gametocytes (A-B, G) 20 days post infection (dpi), 21 dpi (C-D), and 22dpi (H). Mature macrogametocytes (E, F, I) and microgametocytes (H, I) 22 dpi (E), 23 dpi (F) and 24dpi (I). Arrows show parasite nuclei and arrowheads indicate hemozoin granules. Scale bar = 10µm.

	GERPH UN868					GERPH UN870					GERPH UN871				
dpi	**P**	**Ma**	**Mi**	**Im**	**Ht**	**P**	**Ma**	**Mi**	**Im**	**Ht**	**P**	**Ma**	**Mi**	**Im**	**Ht**
1	0	0	0	0	48	0	0	0	0	54	0	0	0	0	57
2	0	0	0	0	52	0	0	0	0	53	0	0	0	0	57
3	0	0	0	0	50	0	0	0	0	52	0	0	0	0	55
4	0	0	0	0	50	0	0	0	0	55	0	0	0	0	58
5	0	0	0	0	53	0	0	0	0	51	0	0	0	0	56
6	0	0	0	0	49	0	0	0	0	53	0	0	0	0	57
7	0	0	0	0	50	0	0	0	0	51	0	0	0	0	56
8	0	0	0	0	50	0	0	0	0	47	0	0	0	0	56
9	0	0	0	0	52	0	0	0	0	49	0	0	0	0	55
10	0	0	0	0	54	0	0	0	0	52	0	0	0	0	54
11	0	0	0	0	48	0	0	0	0	52	0	0	0	0	57
12	0	0	0	0	53	0	0	0	0	54	0	0	0	0	56
13	0	0	0	0	49	0	0	0	0	51	0	0	0	0	56
14	0	0	0	0	52	0	0	0	0	50	0	0	0	0	58

Table 6-2 Parasitemia and Hematrocrit values from individuals studied

	GERPH UN868					GERPH UN870					GERPH UN871				
15	0	0	0	0	52	0	0	0	0	51	0	0	0	0	54
16	0	0	0	0	48	0	0	0	0	53	0	0	0	0	54
17	0	0	0	0	48	0	0	0	0	53	0	0	0	0	53
18	0	0	0	0	49	0	0	0	0	53	0	0	0	0	55
19	0	0	0	0	51	4.4	0	0	4.4	55	5.5	0.1	0.1	5.3	56
20	4.4	0	0.4	4.0	50	12.5	0	0	12.5	48	4.4	0.2	0.1	4.1	51
21	6.1	0	0.3	5.8	51	11.4	0	0.1	11.4	49	9.4	0.3	0.5	8.6	51
22	7.3	0	2.1	5.2	50	23.5	0.8	0.0	22.7	50	19.2	0.1	0.1	19.1	54
23	7.4	0.30	1.3	5.8	50	20.9	0.0	0.2	20.7	50	22.3	0.2	0.6	21.4	54
24	11.8	0.06	2.4	9.3	49	30.0	0.5	10.4	19.0	51	47.9	0.9	1.4	45.6	55
25	15.8	1.38	3.2	11.2	51	33.2	1.1	19.2	12.9	51	39.9	0.7	1.1	38.0	54
26	21.1	1.42	4.4	15.3	53	34.4	9.4	5.1	19.9	50	64.3	1.2	0.8	62.3	55
27	26.4	2.24	4.3	19.9	50	36.1	21.8	7.8	6.5	52	34.6	0.5	0.6	33.6	53
28	31.5	3.33	0.2	28.0	52	15.4	0.3	10.6	4.5	49	24.0	0.3	0.3	23.4	48
29	33.8	2.14	4.8	26.8	51	14.8	0.4	13.6	0.8	46	17.4	0.4	4.0	13.0	48
30	40.2	2.40	5.9	31.8	51	11.9	0.1	10.1	1.6	44	14.3	0.3	1.7	12.3	48
31	51.2	4.26	9.5	37.4	52	9.3	0.3	8.6	0.3	45	13.6	2.2	1.3	10.1	44
32	56.8	5.49	4.2	47.1	53	13.2	5.1	5.7	2.3	47	12.5	4.9	5.2	2.4	44
33	70.8	6.57	11.7	52.6	50	6.3	0.1	5.3	0.9	48	6.1	0.0	4.3	1.8	42
34	-	-	-	-	-	6.1	0.1	5.1	0.9	49	6.3	0.0	4.2	2.1	41
35						4.8	0.5	3.6	0.7	49	4.3	1.0	3.2	0.0	42
36						3.7	0.4	2.1	1.2	44	3.9	1.6	2.0	0.3	41
37						2.4	0.4	1.8	0.3	44	3.0	0.3	2.4	0.3	39
38						3.1	0.2	2.3	0.5	42	3.2	0.1	2.4	0.7	35
39						2.9	0.1	2.1	0.7	40	3.9	0.5	2.8	0.6	36
40						1.8	0.2	1.5	0.2	44	3.5	0.1	2.2	1.3	33
41						3.8	1.4	2.1	0.3	41	5.1	1.6	2.2	1.3	34
42						2.6	0.9	1.0	0.7	42	2.9	0.5	1.2	1.2	33
43						1.3	0.2	0.9	0.2	41	2.3	0.2	0.7	1.4	30
44						1.0	0.1	0.8	0.2	44	2.8	0.1	1.2	1.5	30
45						1.4	0.2	0.9	0.2	44	2.1	0.3	0.6	1.2	29
46						2.0	0.1	0.9	1.0	51	2.4	0.2	1.3	0.9	28
47						1.1	0.1	0.8	0.3	53	1.3	0.1	0.7	0.5	28
48						3.2	0.1	0.8	2.3	54	1.7	0.1	0.9	0.7	28
49						1.1	0.0	0.5	0.6	54	1.5	0.0	0.8	0.6	28
50						2.4	0.1	2.0	0.4	52	1.4	0.2	1.0	0.2	29
51						3.6	0.1	2.9	0.6	54	1.3	0.1	0.5	0.7	28
52						3.5	0.1	2.8	0.6	53	1.1	0.1	0.2	0.8	31
53						2.1	0.2	1.0	0.9	54	1.1	0.1	0.1	0.9	35
54						2.4	0.6	1.4	0.4	51	1.1	0.8	0.1	0.2	38
55						3.6	0.6	2.8	0.1	51	1.5	0.1	0.5	0.9	36
56						2.2	0.7	1.2	0.4	52	2.4	0.3	1.8	0.3	34
57						2.9	1.4	1.2	0.3	52	3.2	1.6	0.6	1.0	38
58						2.7	1.0	1.4	0.3	52	2.6	1.2	0.3	1.1	36
59						2.2	0.2	1.8	0.3	53	3.2	1.4	0.5	1.4	38
60						1.1	0.2	0.8	0.1	54	2.9	0.1	2.1	0.7	38
61						0.7	0.4	0.4	0.0	50	1.0	0.3	0.6	0.1	36
62						0.8	0.5	0.2	0.0	52	1.2	0.5	0.6	0.0	35
63						0.9	0.1	0.8	0.0	54	1.4	0.6	0.7	0.2	36
64						2.8	0.3	2.4	0.1	54	-	-	-	-	-

Table 6-2.: Monitoring total parasitaemia and gametocytaemia values (%) of the infected individuals. ppp: days after infection; *: day on individual was sacrificed; **: day on individual died. P: Parasitaemia; Ma: Macrogametocyte; Mi: Microgametocyte; Im: Immature; Ht: Hematocrit

6.4.3. The establishment of colony of *Pseudolynchia canariensis* and the sporogonic development of *H. columbae*

This study shows that the louse flies can be maintained under semi-natural conditions, and that several generations of these insects were successfully reared. A decline in the number of louse flies placed on the pigeons was also reported; the main reason of insect mortality was the killing of insects by grooming pigeons. Therefore, in order to maintain a sufficient number of louse flies in the colony and to increase the genetic variability of the flies, new wild-caught insects were introduced to the colony every month. Pupae were mainly found in the cage litter, and we were able to differentiate a scale of at least four colours grades of the pupae varying from beige to black, and it was associated with the level of their maturation (data no shown). The development of pupae until the emergence of imago took between 10 and 15 days under the laboratory conditions, as described above.

Serial dissections of 23 louse flies allowed to find ookinetes from 24 h post exposure (Fig. **6-5** A, B), while the few oocysts observed appear in midgut 4 dpi, and mature oocysts can be seen 13 dpi (Fig. 4 **6-5** C, D). Sporozoites were reported from 13 to 16 dpi (Fig. **6-5** E-F). Overall prevalence was 39.1 % (9/23 flies infected). Both, low prevalences and intensities of infection were observed particularly in oocysts (no more than of 6 structures reported per infected insect, Fig. **6-6** A), and sporozoites (2 to 9 parasites recorded in the preparations) (Table **6-3**). All parasitic stages in the vector were measured and compared with values from previous studies (Table **6-4**).

			Number of positives		
Cohort Number	**Gametocytemia (%)**	**-2*n infected (n tested)**	**Ookinete**	**Oocyst**	**Sporozoites**
1	1.27-1.33	6 (9)	2	3	1
2	1.68	0 (2)	0	0	0
3	1.8-1.87	0 (6)	0	0	0
4	2.2- 1.97	3 (6)	2	0	1
	Total	**9 (23)**	**4**	**3**	**2**

Table 6-3.: Number of louse flies found infected after exposure to an infected pigeon (*C. livia*). Gametocytemia intervals of the bird is provided

	Measuremets µm					
Features	*H.columbae/ P. canariensis* This study	H. columbae/ O. avicularia (Baker, 1957)	H. columbae/ P. canariensis (Adie *et al.*, 1915)	*H. palumbis/ O. avicularia* (Baker, 1966)	*P. relictum/ C. p. pipiens* (Kazlauskienė *et al.*, 2013)	*H tartakovskyi/ C. impunctatuts* (Valkiūnas *et al.*, 2002)
Ookinete	n=29	-	-	-	n = 10	n=32
Lenght	9,06-19,86 (15,73±2,33)	20	-	15-16	9.8–19.1 (15.9 ± 2.6)	17.7-30.1 (22.8 ± 2.6)
Width	1,36-2,45(1,904±0,28)	2,25	-	2-3	1.4–2.8 (1.9 ± 0.5)	2.5-3.5 (3.0 t 0.3)
Oocysts	n=30	n=3	-	-	n = 21	n=31
Diameter	12,33-43,76 (26,64 ±9,19)	36-46 (41,67±5,13)	36,5	32-75	23.7–64.3 (40.0 ± 11.8)	2.4-4.4 (3.4 ± 0.5)
Sporozoites	n=8	n=3	-	-	n = 21	n=32
Lenght	6,74-9,45 (8,69 ± 0,87)	7,5-8,5 (7,8± 0,58)	7-10	6-11	11.9–16.8 (13.9 ± 1.7)	8.6-15.2 (11.5 ± 1.6)
Width	0,95-1,26 (1,06 ± 0,12)	0,5	-	0.8-1.0	0.8–1.3 (1.0 ± 0.1)	0.9-1.8 (1.2 ± 0.2)

Table 6-4.: Comparative measurements of ookinete, oocysts and sporozoites stages, reported in haemosporidian parasites

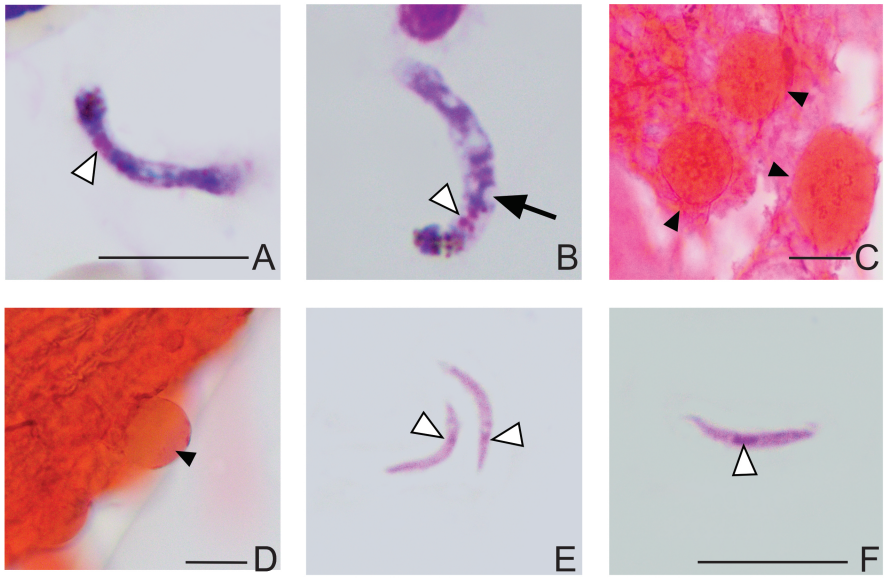


Figure 6-5.: Sporogonic stages of *H. columbae* in experimentally infected louse flies. Methanol-fixed and Giemsa-stained preparations of ookinetes (A, B) and sporozoites (E, F). White arrowheads - parasite nuclei, black arrow - vacuole. Scale bar (A, E, F); scale bar = 10µm. Formalin-fixed and Erlich's hematoxylin stained (C) and mercurochrome stained fresh preparations of midguts (D) showing developing oocysts. Black arrowheads indicate oocysts. Scale bar (C, D); Scale bar = 20µm. Note gathering of pigment granules at the distal end of ookinetes (A, B), and heterogenous structure of developing oocysts (C).

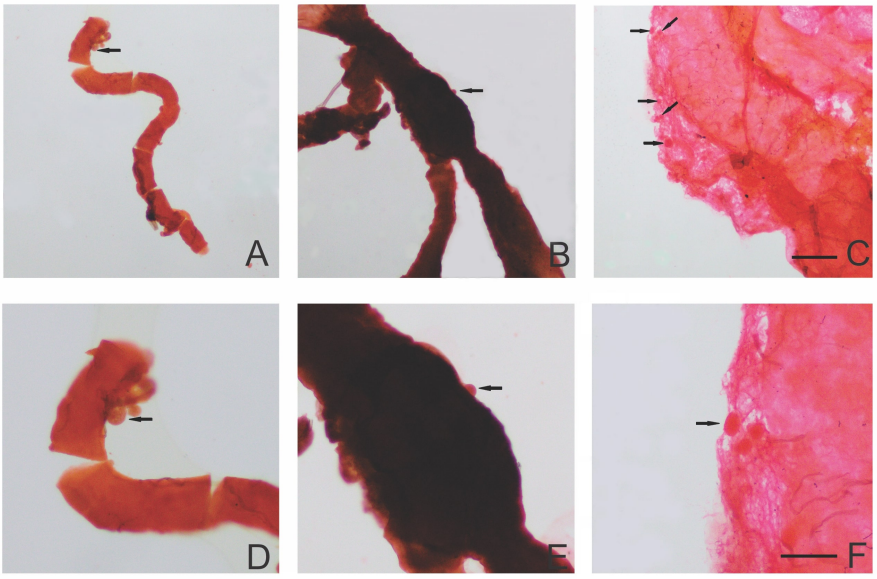


Figure 6-6.: Sporogonic stages of *H. columbae* in experimentally infected louse flies. Methanol-fixed and Mercurochrome stained fresh preparations of midguts showing developing oocysts. Black arrowheads indicate oocysts. Scale bar = 200µm (A), =100µm (B, D) and 40=µm (C, E and F).

6.4.4. Histopathological findings

Examination of histological organ sections of the euthanized infected pigeon reveals that lungs presented the most critical damage generated by *H. columbae* infection, with moderate multifocal perivascular and mild lymphoplasmacytic pneumonia with abundant and multifocal exo-erythrocytic meronts located in capillary vessels (Fig.**6-7** A). In the liver, moderate and multifocal lymphoplasmacytic hepatitis was observed, accompanied with moderate intracytoplasmic presence of meronts of *H. columbae* in mononuclear cells located in sinusoids (Fig.**6-7** B and C). Also, multifocal and moderate hemosiderosis were associated with the presence of meronts (Fig.**6-7** D). Meronts were characterized by irregular, often lobular-like shape and variable size (Table **6-5**).

	Meronts in lungs n=46		Meronts in kidneys n=3	Meronts in spleen n=3
	Sagittal section	Transversal section		
	n=36	n=10		
Lenght	15.1-44.7 (27.5±9.1)	9.1-15.4 (13.3 ± 1.8)	10.6-23.6 (15.04 ± 7.42)	13.5-27.5 (20.3 ± 6.9)
Width	4.6-17.14 (11.8±3.6)	6.4 -9.5 (7.7 ± 1.05)	8.7 -11.9 (10.6 ± 1.7)	4.4 -10.7 (7.4 ± 3.2)
Area	59.6-216.4 (137.1±52.5)	47.97-91.5 (65.05 ± 17.8)	59.9-136.4 (107.6 ± 41.5)	83.8-303.6 (164.6 ± 120.1)
N° of merozoites	34-115 (63.82 ± 24.40)	24 -65 (40.1 ± 11.9)	31-74 (46.7 ± 23.7)	27 -96 (57 ± 35.4)

Table 6-5.: Measurements in µm of Meronts found in lungs, kidneys and spleen from pigeon GERPH-UN868. Length, width, area and number of merozoites information are given.

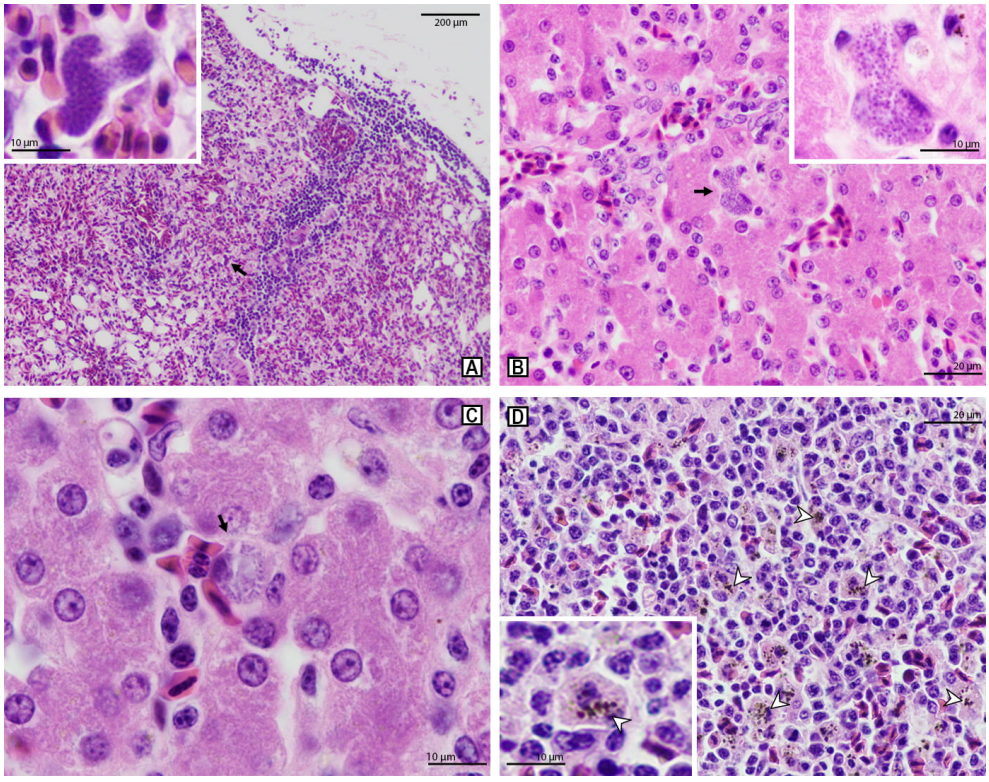


Figure 6-7.: Exo-erythrocytic meronts of *H. columbae* in lungs (A), liver (B, C) and spleen (D) of experimentally infected rock pigeon (GERPH UN868) 33 days post infection. Black arrows – meronts, white simple arrowheads - hemosideriosis. Note branching shape of meronts in lungs (A)

6.5. Discussion

The key result of this study is the establishment of a useful avian host-parasite-vector model, which provides opportunities to access all stages of the life cycle in the haemosporidian parasite *H. columbae* under controlled laboratory conditions. Currently, this model system is available for investigation of biology of avian haemosporidia parasites of the subgenus *Haemoproteus*.

Several species of parasites belonging to this subgenus have been described parasitizing Columbiform and marine birds (Valkiūnas, 2005; Valkiūnas *et al.*, 2010; Levin *et al.*, 2012), and the system described here will allow detailed studies of the life cycle of parasites of this subgenus. Being important due to it has been reported that haemoproteids cause serious diseases in pigeons and doves (Earle *et al.*, 1993).

Few *cytb* lineages of *H. columbae* have been reported around the world, and the genetic distances between them can be as big as 1 % (e.g., hCOQUI05 and hCOLIV03; Chagas *et al.*, 2016). Because different parasite lineages of same parasite species could produce different immune responses, or pathology in their hosts and vectors (Kazlauskienė *et al.*, 2013; Zélé *et al.*, 2014; Dimitrov et. al., 2015), in the present research we mimic the natural system of infection and transmission of *H. columbae* and generate a population of pigeons infected with the same lineage of parasite HAECOL1(clones). The lineage HAECOL1 is of wide distribution; there are reports of this parasite in Africa (Waldenström *et al.*, 2002), Colombia (González *et al.*, 2015; Coral *et al.*, 2015), Italy (Scaglione *et al.*, 2015) and Brazil (Chagas *et al.*, 2016). This provides opportunities to develop comparative research using the same lineage in different sites and laboratories in the future.

It is important to note that a louse fly colony, which is free of parasites was established. That provides opportunities to maintain the *H. columbae* infection in their natural hosts, the Rock pigeon and *P. canariensis*. This study provided the first information about the development of the lineage HAECOL1 both in the insect vector and in the natural avian host *C. livia*.

The host parasite model "*Haemoproteus columbae-Columba livia*" has been used in studies of avian haemosporidians (Aragão, 1908; Valkiūnas, 2005). Some observations on the parasite life cycle, development and the course of infection have been published in France, Egypt, India and other countries (Sergent and Sergent, 1906; Aragão, 1908, Adie *et al.*, 1915; Rendtorff *et al.*, 1949; Hickman, 1952; Mohammed *et al.*, 1958; Ahmed and Mohammed, 1978). The prevalence of this infection is high, and an active transmission occurs year-round in tropical countries, with a predominance of chronic infections in pigeons (Adriano and Cordeiro, 2001; Villalba-Sánchez and J., 2014; Coral *et al.*, 2015; González *et al.*, 2015). This study provides the first experimental data on behaviour of the HAECOL1 lineage both in the vector and vertebrate hosts.

The course of *H. columbae* infection reported in this study, particularly the prepatent period, was

similar to that reported by Rendtorff *et al.* (1949) and Waite *et al.,* (2014) despite the infection method used (louse fly bite or injection of infected macerated flies). Nevertheless, it is shorter in comparison to the prepatent period reported by other authors (Sergent and Sergent, 1906; Adie *et al.*, 1915; Hickman, 1952; Mohammed *et al.*, 1958; Ahmed and Mohammed, 1978; Table **6-6**). Indeed, the most prolonged period (15 days) was obtained by Ahmed and Mohammed, (1978), who also used various modes of infection (louse fly bites, intramuscular, intravenous, and intraperitoneal inoculation of sporozoites). Porter *et al.,* (1952), refer that these differences in time of the prepatent period (Table **6-6**) may be due to how the infections were carried out. However, more in-depth studies should be designed to solve this issue.

Authors	Location	Prepatent phase (dpi range)	Acuate phase (dpi range)
Sergent and Sergent, 1906	No data	28	No data
Adie, 1924	No data	28	No data
Rendtorff *et al.,* 1949	No data	17-33	No data
Coatney and Hickman 1952	No data	26-38	No data
Mohammed, 1958	Cairo, Egypt	25-34	No data
Ahmed and Mohammed, 1978	Cairo, Egypt	22-37	31-57
Waite *et al.,* 2014	USA	∼18-19	∼25-38
This study	Bogot, Colombia	19-20	26-32

Table 6-6.: Data on prepatent and acute stages during *H. columbae* infection.

The acute stage in this study was slightly shorter than reported by Ahmed and Mohammed (1978) and similar to Waite *et al.,* (2014; Table **6-6**). Such differences can be due to numerous factors, such as the climatic conditions of the aviary, the geographical origin of the sample, the immune response and physical condition of the hosts, sex, age, and the parasite lineage (Valkiūnas, 2005; Gayathri and Hegde, 2006; Donovan *et al.*, 2008; Olias *et al.*, 2011; Waite *et al.*, 2012b; Ghosh *et al.*, 2014; Valkiūnas and Iezhova, 2017).

It has been shown that infection by haemosporidians, e.g. *P. gallinaceum* (de Macchi *et al.*, 2013) or *Haemoproteus* species (Cannell *et al.*, 2013) has a direct effect on haematocrit values causing anaemia or sometimes death (Donovan *et al.*, 2008; Cannell *et al.*, 2013). Previous reports on haematocrit values in healthy pigeons vary between 41 to 50 % (Gayathri and Hegde, 2006; Glomski and Pica, 2016). Our results, however, show two opposite cases; the first pigeon GERPH-UN868 at the time of its sacrifice had a parasitaemia of 70.8 % and its haematocrit was 50 %. The second pigeon GERPH-UN871 showed very low values of haematocrit in several days of the infection with parasitaemia < 2 %. This probably indicates differences in immune response of the hosts. It is possible that such grade of infection in the pigeon GERPH-UN868 does not cause anaemia, and probably, if this infection can cause disease or death in nature, it would be associated with the rupture of exo-erythrocytic meronts, which can induce an uncontrolled immune response (Lee *et al.*, 2016). It is essential to keep in mind that the individual host immune condition also could affect the course

and the final result of the infection (parasitaemia burden).

Recently, significant advances have been made in the characterization of the sporogonic development of haemosporidia parasites infecting birds. For the particular case of *Haemoproteus* (*Haemoproteus*) species that are transmitted by louse flies, sporogony takes from 6.5 to 7 days (*H. palumbis*; Baker, 1966) or even ten days (*H. columbae*; Adie *et al.*, 1915). Nevertheless, it has been reported that this process for the parasites transmitted by louse flies may take longer periods of time, since oocysts are large so need more time to mature (Valkiūnas, 2005). Despite this, our results contrast with those reported by Adie (1915), where the sporogony of the same parasite species takes about ten days to be completed in "*Lynchia maura*" (*P. canariensis*) (Table **6-4**). Probably the aviary and environmental conditions differences in different studies might be an explanation. However, Baker (1966) dismisses the environmental temperature variation as a probable cause of the variation in the duration of the extrinsic period of the parasite, since being ectoparasites, louse flies remain in close contact with the skin of the vertebrate host, thereby, under a more constant temperature.

Adie (1915) reported that the parasite sporogony is completed within nine to ten days when the insects were maintained on an infected bird. In our study, the insects did not remain on the infected host (where reinfection of the insect is possible), but instead the louse flies were transferred to an uninfected bird after 24 h of contact with the infected pigeons. The complete sporogony in our experiment took between 13 and16 dpi. Some environmental features or stress due to manipulation of the insects or host could also cause these changes.

Vector competence may fluctuate according to the strain of the vector (Collins *et al.*, 1986), the parasite lineage, or the presence of endosymbionts (Weiss and Aksoy, 2011; Zélé *et al.*, 2014). Our current knowledge about the impact of such variables in this system is limited, and further studies on these matters need to be developed.

Previous studies have demonstrated a deleterious effect of the high gametocytaemia (over 1 %) of *Parahaemoproteus* and *Plasmodium* parasites on its natural vectors (Valkiūnas *et al.*, 2015; Bukauskaité and Valkiūnas, 2016). In this study, such gametocytaemia seems not to be harmful in louse flies, probably because biting midges and mosquitoes are tiny insects in comparison to louse flies.

This study showed numerous exo-erythrocytic meronts in infected birds, with particularly high intensity in the lungs. Moderate focal granulomatous inflammation in the spleen, necrosis with central classification necrosis, and moderate generalized hemosiderosis (overload of iron deposition associated with hemolysis due to malaria infection, (MacDonald, 1963) were observed. Recent reports have confirmed the virulence of infection of haemoproteids that for several decades was considered benign. That is probably because it is difficult to follow infections in the wild, and sick avian hosts are easy prey in wildlife, so are difficult to sample during field-work. A few studies discuss in detail the pathological mechanism underlying the infection or mortality caused by these parasites.

Earlé *et al.* (1993), analyzed histopathological sections of sick pigeons due to possible *H. columbae* infection and found numerous meronts and multi-lobular megalomeronts in several organs, particularly in the striated muscles. It was concluded that extensive fiber necrosis produced by rupture of megalomeronts could be the cause of mortality. Ortiz-Catedral et al. (2019) found that *H. minutus* cause lethal infections in Australasian and South American parrots, even in the absence of parasites in peripheral blood. The infections were associated with multifocal extended haemorrhages caused by disrupted meronts in the heart and gizzard muscle. More research is needed for better understanding virulence of avian haemoproteids.

In our study, there was evidence of extensive damage, mainly in pigeon's lungs (Fig. **6-7** A). However, in contrast to Earlé *et al.* (1993) report, megalomeronts were not observed in this study (meronts in Fig. **6-7** A-C). The meronts observed in our study were more similar to those reported in the lungs, by Aragão (1908) in Brazil, which have an irregular shape. Probably differences in the host species or even the parasite lineage might explain such changes (Atkinson *et al.*, 1995; Valkiūnas, 2005).

6.6. Conclusions

This study has developed and used an experimental model organism to access haemosporidian parasites belonging to the subgenus *Haemoproteus* at all stages of their life cycle in avian hosts and insect vectors. This experimental methodology and host-parasite model provide opportunities to collect biological parasite material to answer various research issues. Due to easy access to different parasite life cycle strategies (gametocytes, gametes, ookinetes, oocysts, sporozoites), this methodology enables the testing of hypotheses or solving questions related to various morphological, physiological, immunological or genetic issues in studies aiming better understanding haemosporidian infections. Moreover, the model allows precise monitoring of the course of infection under different and well-controlled biotic and abiotic variables. Data on the life cycle of widely distributed *H. columbae* (lineage HAECOL1) were obtained, and this facilitates extrapolation of results from different geographical regions. Both sporogony and merogony of *H. columbae* can be readily followed in laboratory conditions, providing opportunities to obtain valuable information about this infection at any time using samples either from the vector or the vertebrate host.

Acknowledgements

The authors thank the former and new students of the Host-Parasite Relationship Research Group: Avian Haemoparasites Model for field assistance to captured wild louse flies, caring for pigeons and for always being willing to help in all the experiments described in this manuscript. To the anonymous reviewer which significantly improve with his/her comments this manuscript.

References

Adie, H. et al. (1915). The sporogony of *Haemoproteus columbae*. *Indian Journal of Medical Research*, 2(3).

Adriano, E. A. and Cordeiro, N. S. (2001). Prevalence and intensity of *Haemoproteus columbae* in three species of wild doves from brazil. *Memorias do Instituto Oswaldo Cruz*, 96(2):175–178.

Ahmed, F. E. and Mohammed, A.-H. H. (1978). *Haemoproteus columbae*: course of infection, relapse and immunity to reinfection in the pigeon. *Zeitschrift für Parasitenkunde*, 57(3):229–236.

Aragão, H. d. B. (1908). Sobre o cyclo evolutivo ea transmissão do *Haemoproteus columbae*. *Revista Médica de São Paulo*, 11(20).

Atkinson, C., Woods, K., Dusek, R. J., Sileo, L., and Iko, W. (1995). Wildlife disease and conservation in hawaii: pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected iiwi (vestiaria coccinea). *Parasitology*, 111(S1):S59–S69.

Atkinson, C. T. (1986). Host specificity and morphometric variation of *Haemoproteus meleagridis* levine, 1961 (protozoa: Haemosporina) in gallinaceous birds. *Canadian Journal of Zoology*, 64(11):2634–2638.

Baker, J. (1957). A new vector of *Haemoproteus columbae* in england. *The Journal of Protozoology*, 4(3):204–208.

Baker, J. (1966). *Haemoproteus palumbis* sp. nov.(sporozoa, haemosporina) of the english woodpigeon *Columba palumbus*. *The Journal of protozoology*, 13(3):515–519.

Baker, J. (1967). A review of the role played by the hippoboscidae (diptera) as vectors of endoparasites. *The Journal of parasitology*, pages 412–418.

Baptiste, E., Piedad, M., Castaño, N., Cárdenas López, D., Gutiérrez, F. d. P., Gil, D., Lasso, C. A., et al. (2010). Análisis de riesgo y propuesta de categorización de especies introducidas para Colombia.

Bernal, G., Rosero, M., Cadena, M., Montealegre, J., and Sanabria, F. (2007). Estudio de la caracterización climática de bogotá y cuenca alta del río tunjuelo. *Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM—Fondo de Prevención y Atención de Emergencias FOPAE.*

Bishopp, F. (1929). The pigeon fly—an important pest of pigeons in the united states. *Journal of Economic Entomology*, 22(6):974–980.

Bukauskaitė, D., Žiegytė, R., Palinauskas, V., Iezhova, T. A., Dimitrov, D., Ilgūnas, M., Bernotienė, R., Markovets, M. Y., and Valkiūnas, G. (2015). Biting midges (culicoides, diptera) transmit *Haemoproteus* parasites of owls: evidence from sporogony and molecular phylogeny. *Parasites & vectors*, 8(1):303.

Bukauskaité, D., B. R. I. T. and Valkiūnas, G. (2016). Mechanisms of mortality in culicoides biting midges due to *Haemoproteus* infection. *Parasitology*, 143(13):1748–1754.

Cannell, B., Krasnec, K., Campbell, K., Jones, H., Miller, R., and Stephens, N. (2013). The pathology and pathogenicity of a novel *Haemoproteus* spp. infection in wild little penguins (eudyptula minor). *Veterinary parasitology*, 197(1-2):74–84.

Chagas, C. R. F., de Oliveira Guimarães, L., Monteiro, E. F., Valkiūnas, G., Katayama, M. V., Santos, S. V., Guida, F. J. V., Simões, R. F., and Kirchgatter, K. (2016). Hemosporidian parasites of free-living birds in the são paulo zoo, brazil. *Parasitology research*, 115(4):1443–1452.

Collins, F. H., Sakai, R. K., Vernick, K. D., Paskewitz, S., Seeley, D. C., Miller, L. H., Collins, W. E., Campbell, C. C., and Gwadz, R. W. (1986). Genetic selection of a *Plasmodium*-refractory strain of the malaria vector *Anopheles gambiae*. *Science*, 234(4776):607–610.

Colorado-Garzón, F. A., Adler, P. H., García, L. F., Muñoz de Hoyos, P., Bueno, M. L., and Matta, N. E. (2016). Estimating diversity of black flies in the *Simulium ignescens* and *Simulium tunja* complexes in Colombia: Chromosomal rearrangements as the core of integrative taxonomy. *Journal of Heredity*, 108(1):12–24.

Coral, A. A., Valkiūnas, G., González, A. D., and Matta, N. E. (2015). In vitro development of *Haemoproteus columbae* (haemosporida: Haemoproteidae), with perspectives for genomic studies of avian haemosporidian parasites. *Experimental parasitology*, 157:163–169.

de Macchi, B. M., Miranda, F. J. B., de Souza, F. S., de Carvalho, E. C. Q., Albernaz, A. P., do Nascimento, J. L. M., and DaMatta, R. A. (2013). Chickens treated with a nitric oxide inhibitor became more resistant to *Plasmodium gallinaceum* infection due to reduced anemia, thrombocytopenia and inflammation. *Veterinary research*, 44(1):8.

Donovan, T. A., Schrenzel, M., Tucker, T. A., Pessier, A. P., and Stalis, I. H. (2008). Hepatic hemorrhage, hemocoelom, and sudden death due to *Haemoproteus* infection in passerine birds: eleven cases. *Journal of Veterinary Diagnostic Investigation*, 20(3):304–313.

Earle, R., Bastianello, S. S., Bennett, G., and Krecek, R. (1993). Histopathology and morphology of the tissue stages of *haemoproteus columbae* causing mortality in columbiformes. *Avian Pathology*, 22(1):67–80.

Galluci, B. B. (1974). Fine structure of *haemoproteus columbae* kruse during differentiation of the ookinete. *The Journal of protozoology*, 21(2):264–275.

Gayathri, K. and Hegde, S. (2006). Alteration in haematocrit values and plasma protein fractions during the breeding cycle of female pigeons, *Columba livia*. *Animal reproduction science*, 91(1-2):133–141.

Ghosh, S., Waite, J. L., Clayton, D. H., and Adler, F. R. (2014). Can antibodies against flies alter malaria transmission in birds by changing vector behavior? *Journal of theoretical biology*, 358:93–101.

Glomski, C. A. and Pica, A. (2016). *The avian erythrocyte: its phylogenetic odyssey*. CRC Press.

González, A. D., Lotta, I. A., García, L. F., Moncada, L. I., and Matta, N. E. (2015). Avian haemosporidians from neotropical highlands: evidence from morphological and molecular data. *Parasitology international*, 64(4):48–59.

Harwood, R. F., James, M. T., et al. (1979). *Entomology in human and animal health.* Number 7th edition. Macmillan Publishing Co. Inc. New York; Baillière Tindall, 35 Red Lion.

Herath, P. R. (1966). *Colonizing Pseudolynchia canariensis on hosts other than the pigeon: Columba livia*. PhD thesis, Wayne State University, Department of Biology.

Hickman, B. (1952). The course of sporozoite-induced *Haemoproteus columbae* infection in the pigeon. *J. Parasit*, 38:12.

Hutson, A. M. et al. (1984). Keds, flat-flies and bat-flies. diptera, hippoboscidae and nycteribiidae. *Keds, flat-flies and bat-flies. Diptera, Hippoboscidae and Nycteribiidae.*, 10(7).

Kazlauskienė, R., Bernotienė, R., Palinauskas, V., Iezhova, T. A., and Valkiūnas, G. (2013). *Plasmodium relictum* (lineages psgs1 and pgrw11): complete synchronous sporogony in mosquitoes *Culex pipiens pipiens*. *Experimental parasitology*, 133(4):454–461.

Klei, T. R. and De Giusti, D. L. (1973). Ultrastructural changes in salivary glands of *Pseudolynchia canariensis* (dipteria: Hippoboscidae) infected with sporozoites of *haemoproteus columbae*. *Journal of Invertebrate Pathology*, 22(3):321–328.

LaPointe, D. A., Goff, M. L., and Atkinson, C. T. (2005). Comparative susceptibility of introduced forest-dwelling mosquitoes in hawai'i to avian malaria, plasmodium relictum. *Journal of Parasitology*, 91(4):843–850.

Lee, H. R., Koo, B.-S., Jeon, E.-O., Han, M.-S., Min, K.-C., Lee, S. B., Bae, Y., and Mo, I.-P. (2016). Pathology and molecular characterization of recent *leucocytozooni caulleryi* cases in layer flocks. *Journal of biomedical research*, 30(6):517.

Levin, I. I., Valkiūnas, G., Iezhova, T. A., O'brien, S. L., and Parker, P. G. (2012). Novel *Haemoproteus* species (haemosporida: Haemoproteidae) from the swallow-tailed gull (lariidae), with remarks on the host range of hippoboscid-transmitted avian hemoproteids. *The Journal of parasitology*, pages 847–854.

MacDonald, R. A. (1963). Hemochromatosis and hemosiderosis. *Kanzo*, 5(1):34–36.

Martinsen, E. S., Perkins, S. L., and Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Molecular phylogenetics and evolution*, 47(1):261–273.

Mohammed, A. H. et al. (1958). Systematic and experimental studies on protozoal blood parasites of egyptian birds. vols. i. & ii. *Systematic and Experimental Studies on Protozoal Blood Parasites of Egyptian Birds. Vols. I. & II.*

Olias, P., Wegelin, M., Zenker, W., Freter, S., Gruber, A. D., and Klopfleisch, R. (2011). Avian malaria deaths in parrots, europe. *Emerging Infectious Diseases*, 17(5):950.

Ortiz-Catedral, L., Brunton, D., Stidworthy, M. F., Elsheikha, H. M., Pennycott, T., Schulze, C., Braun, M., Wink, M., Gerlach, H., Pendl, H., et al. (2019). *Haemoproteus minutus* is highly virulent for australasian and south american parrots. *Parasites & vectors*, 12(1):40.

Porter, R. J., Laird, R. L., and Dusseau, E. M. (1952). Studies on malarial sporozoites. i. effect of various environmental conditions. *Experimental Parasitology*, 1(3):229–244.

Rendtorff, R., Jones, W., and Coatney, G. (1949). Studies on the life-cycle of *Haemoproteus columbae*. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 43(1):7.

Scaglione, F. E., Pregel, P., Cannizzo, F. T., Pérez-Rodríguez, A. D., Ferroglio, E., and Bollo, E. (2015). Prevalence of new and known species of haemoparasites in feral pigeons in northwest italy. *Malaria journal*, 14(1):99.

Sergent, E. and Sergent, E. (1906). Sur le second hôte de l'*Haemoproteus* (halteridium) du pigeon. *Comptes Rendus Hebdomadaires des Séances et Mémoires de la Société de Biologie et de ses Filiales*, 58(2).

Soto, C. and Acosta, I. (2010). Prevención y enfermedades de la paloma doméstica. *Revista electrónica de Veterinaria*, 11(11):5–79.

Thomas, N. J., Hunter, D. B., and Atkinson, C. T. (2008). *Infectious diseases of wild birds*. John Wiley & Sons.

Valkiūnas, G. (2005). Avian malaria parasites and other haemosporidia CRC press. *Florida, Boca Raton*.

Valkiūnas, G. and Iezhova, T. A. (2017). Exo-erythrocytic development of avian malaria and related haemosporidian parasites. *Malaria journal*, 16(1):101.

Valkiūnas, G., Liutkevičius, G., and Iezhova, T. A. (2002). Complete development of three species of *Haemoproteus* (haemosporida, haemoproteidae) in the biting midge culicoides impunctatus (diptera, ceratopogonidae). *Journal of Parasitology*, 88(5):864–869.

Valkiūnas, G., Santiago-Alarcon, D., Levin, I. I., Iezhova, T. A., and Parker, P. G. (2010). A new *Haemoproteus* species (haemosporida: Haemoproteidae) from the endemic galapagos dove zenaida galapagoensis, with remarks on the parasite distribution, vectors, and molecular diagnostics. *Journal of Parasitology*, 96(4):783–793.

Valkiūnas, G., Žiegytė, R., Palinauskas, V., Bernotienė, R., Bukauskaitė, D., Ilgūnas, M., Dimitrov, D., and Iezhova, T. A. (2015). Complete sporogony of *Plasmodium relictum* (lineage pgrw4) in mosquitoes *culex pipiens pipiens*, with implications on avian malaria epidemiology. *Parasitology research*, 114(8):3075–3085.

Villalba-Sánchez, C., O.-L. A. and J., O. (2014). *Columba livia* domestica gmelin, 1789: plaga o símbolo. *Revista Colombiana de Ciencia Animal-RECIA*, pages 363–368.

Vrijenhoek, R. (1994). Dna primers for amplification of mitochondrial *cytochrome c oxidase subunit I* from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, 3(5):294–9.

Waite, J. L., Henry, A. R., Adler, F. R., and Clayton, D. H. (2012a). Sex-specific effects of an avian malaria parasite on an insect vector: support for the resource limitation hypothesis. *Ecology*, 93(11):2448–2455.

Waite, J. L., Henry, A. R., and Clayton, D. H. (2012b). How effective is preening against mobile ectoparasites? an experimental test with pigeons and hippoboscid flies. *International journal for parasitology*, 42(5):463–467.

Waite, J. L., Henry, A. R., Owen, J. P., and Clayton, D. H. (2014). An experimental test of the effects of behavioral and immunological defenses against vectors: do they interact to protect birds from blood parasites? *Parasites & vectors*, 7(1):104.

Waldenström, J., Bensch, S., Kiboi, S., Hasselquist, D., and Ottosson, U. (2002). Cross-species infection of blood parasites between resident and migratory songbirds in africa. *Molecular Ecology*, 11(8):1545–1554.

Weiss, B. and Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends in parasitology*, 27(11):514–522.

Zélé, F., Vézilier, J., L'Ambert, G., Nicot, A., Gandon, S., Rivero, A., and Duron, O. (2014). Dynamics of prevalence and diversity of avian malaria infections in wild *Culex pipiens* mosquitoes: the effects of wolbachia, filarial nematodes and insecticide resistance. *Parasites & vectors*, 7(1):437.

Haemoproteus columbae ApiGenome: as an approach for evolutionary and phylogenetic studies of the Apicoplasts

Axl S. Cepeda[a], M. Andreína Pacheco[b], Ananias A. Escalante[b], Juan F. Alzate[c], Nubia E. Matta[a]

a. Departamento de Biología, Grupo de Investigación Caracterización Genética e Inmunología, Sede Bogotá-Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia.

b. Department of Biology, Institute for Genomics and Evolutionary Medicine (igem), Temple University, Philadelphia, PA

c. 3Centro Nacional de Secuenciación Genómica – CNSG, SIU, Grupo de Parasitología, Facultad de Medicina, Universidad de Antioquia, Medellín, Antioquia, Colombia.

7.1. Abstract

Haemoproteus (*Haemoproteus*) *columbae* is a haemosporidian parasite highly prevalent in Columbiformes and close related to the *Plasmodium* genus. We report the complete sequence and annotation of the apicoplast genome of the *H. columbae* (lineage HAECOL1). The genome consists of a 29.8 kb circular molecule encoding CDS, ORFs, tRNAs and rRNAs. Genome analysis and annotation revealed a conserved structure among the Haemosporidian parasites. The values of relative synonymous codon usage (RSCU) and the effective number of codon (ENc) were calculated and showed a bias codon usage. Based on the information obtained from *H. columbae* ApiGenome, it was possible to identify genes useful for phylogenetical analyses based on the index of substitution saturation and the relative estimation of evolutionary rates, such as *clpC*. Besides, we designed primers for the amplification of the *clpC* gene as a taxonomic and phylogenetic marker of the order Haemosporida. This additional information is particularly useful for the development of therapeutic targets as well as evolutionary studies, comparative genomics, among others, since the mitochondrial genome and the RNA-Seq data of this parasite are already available. Nevertheless, the obtained phylogenetic hypotheses demonstrate the importance of increasing the number of ApiGenomes.

7.2. Introduction

Haemosporidian belong to the phylum Apicomplexa, which is characterized by the presence of a non-photosynthetic plastid, apparently originated from a secondary endosymbiosis of an ancestral algal known as Apicloplast (McFadden, 2011). There is evidence of ancestry of a red algal chloroplast genome, which by means of degeneration/reduction processes has given rise to an ApiGenoma (Janouškovec *et al.*, 2010). In addition, its genetic information is inherited maternally, and it is in a single copy (Okamoto *et al.*, 2009). This organelle is essential for the parasite survival since it participates in metabolic pathways such as the provision of fatty acids, isoprenoids, heme biosynthesis and possibly iron-sulfur cluster to the parasitic cell stage (Yeh and DeRisi, 2011; Sigala and Goldberg, 2014; Shears *et al.*, 2015; Mehlhorn, 2016). Thus, the deeply knowledge of this pseudo-organelle is desirable as some of its sequences could be plausible for new therapeutical targets for the design of antimalarial drugs (Soldati, 1999; Ralph *et al.*, 2001; Srimath P. *et al.*, 2017).

The circular ApiGenome possesses protein genes, rRNAs, tRNAs, and other genes, and its size varies depending on the species from 30 to 35 kb (Arisue *et al.*, 2012; Arisue and Hashimoto, 2015). Likewise, some apicoplast genes could be useful in phylogenetic analyses considering the current limitations of evolutionary hypotheses in haemosporidian, which present low phylogenetic resolution derived from the limited number of molecular markers and the sizes of the sequences used in those analyses (Martinsen *et al.*, 2008; Borner *et al.*, 2016). Therefore, it is desirable the use of new molecular markers from nuclei (Hellgren *et al.*, 2004; Martinsen *et al.*, 2008) or ApiGenome (Valkiūnas *et al.*, 2019).

Currently, there are at least 20 apicoplast genomes published belonging to *Plasmodium* species (Aurrecoechea *et al.*, 2008; Arisue *et al.*, 2012; Arisue *et al.*, 2019) and one from *Leucocytozoon caulleryi* (Imura *et al.*, 2014). In this study, we sequenced and characterized the first complete apicoplast genome from a parasite belonging to the *Haemoproteus* genus (subgenus *Haemoproteus*) and highlighted the great importance of this genus due to its closeness to *Plasmodium* genus (Valkiūnas, 2005; Borner *et al.*, 2016; Pacheco *et al.*, 2017). In addition, in order to understand more about the evolution of this organellar genome, we evaluated phylogenetic hypotheses based on analyses of its Index of substitution saturation, codon usage and estimation of its relative evolutionary rates. Based on the aforementioned analyzes, *clpC* gene is a good candidate for phylogenetic approaches within Haemosporida order; therefore, we proposed new primers for a nested PCR protocol, which will allow amplifying this gene.

7.3. Material and Methods

7.3.1. Ethical considerations

The methodology used in this study was approved by the Bioethics Committee from the "*Facultad de Ciencias*" of the "*Universidad Nacional de Colombia*" (Act number: 04 of 2017). Fieldwork was done under permit No. 0255 granted by "*Autoridad Nacional de Licencias Ambientales* (ANLA)".

7.3.2. Sample collection, gDNA Extraction and Sequencing

Almost 120µL of whole blood was taken in heparinized microhaematocrit tubes (NRIS, vitrex medical A/S Ref 161315) from the *Columba livia* (GERPH-UN868) infected with *H. columbae* (lineage HAECOL1) parasitemia 76.8 %. DNA from the blood was extracted using DNeasy Blood & Tissue kit (Qiagen, GmbH, Hiden, Germany), resulting in a DNA concentration of 924.113 ng/µL as determined by a NanoDrop spectrophotometer (Thermoscientific nanolite).

The NGS sequencing library was prepared from 1.2µg gDNA by random fragmentation of the DNA sample, followed by 5' and 3' adapter ligation according to the TruSeq DNA PCR free Sample Preparation Kit (insert size average 350bp; Illumina). Sequencing and library preparation were conducted at Macrogen, Korea, in an Illumina Hiseq X. One full lane was used for this experiment

generating PE reads of 150 bases.

7.3.3. Raw Data, Preprocessing, Assembly and Annotation of *H. columbae* ApiGenome

The Illumina HiSeq X generated raw images and base calling through an integrated primary analysis software called RTA2 (Real Time Analysis 2). The BCL (base calls) binary was converted into FASTQ using illumina package bcl2fastq2-v2.20.0. A total of 628,859,636 reads were produced, and the total read bases were 95 Gb (compressed file). The GC content (%) was 41.84 %, and Q30 was 86.7 %. Reads quality was determined by the FastQC software (Bioinformatics, 2017). Low quality read ends and adapter sequences were removed using Trimmomatic software (Bolger *et al.*, 2014). The trimmed reads were mapped to the *C. livia* genome (accession number **GCA_001887795.1**) using Burrows-Wheeler Aligner long-read alignment (BWA-mem; Li and Durbin, 2010). Reads mapped to the *C. livia* genome were removed, and only unmapped reads were used in the following *de novo* assembly processes. FLASh software (Fast Length Adjustment of Short reads; Magoč and Salzberg, 2011) was used to extend reads when it is possible.

De novo assembly was done using St. Petersburg genome assembler (SPAdes; Bankevich *et al.*, 2012) with default parameters with kmers of 33, 55, 77 and 99. Extended reads (with FLASH) as well as cleaned PE reads were used for this process. ApiGenome was entirely obtained in one single scaffold, and annotation was carried out manually by MEGA7 (Kumar *et al.*, 2016) using *L. caulleryi* ApiGenome (accession number **AP013071**) as a reference. *Haemoproteus columbae* ApiGenome map was obtained using CGView ServerBETA (Stothard and Wishart, 2004).

7.3.4. Data Retrieval, Alignment construction, Synteny, Index of substitution saturation and Codon usage

The complete or nearly complete ApiGenome sequences were obtained from GenBank (Table **7-1**). The sequence alignment was achieved by using Clustal Omega (Madeira *et al.*, 2019) and Muscle as implemented in SeaView v4.3.5 (Gouy *et al.*, 2009) with manual editing. This alignment was constructed with all mentioned above ApiGenome sequences (29.798bp excluding gaps) and genomic comparisons were performed using Mauve (Darling *et al.*, 2010).

NCBI code	Haemosporidian Parasites	Strain	ApiGenome features		Vertebrate Hosts
			Length (bp)	GC content	
	H. columbae	HAECOL1	29798	12,3	*Columba livia*
AP013071	*L. caulleryi*	Niigata	34779	14,9	*Gallus domesticus*
AB649424	*P. gallinaceum*	A8	28981	12,8	*G. domesticus*
$NC_0$31964	*P. relictum*	SGS1	29365	13,1	

NCBI code	Haemosporidian Parasites	Strain	ApiGenome features		Vertebrate Hosts
			Length (bp)	GC content	
AB649421	*P. berghei*	ANKA	29264	14	*Grammomys sp.*
AB649423	*P. chabaudi*	AS	29198	13,7	*Thamnomys sp.*
AB649420	*P. coatney*	CDC	29055	13,1	*Macaca fascicularis*
AP018101	*P. cynomolgi*	Ceylonensis	34378	14,2	*M. nemestrina*
LT841394	*P. cynomolgi*	M	34521	14,2	*M. nemestrina*
AP018102	*P. cynomolgi*	Berok	34515	14,2	*M. nemestrina*
$NC_0$36769	*P. falciparum*	3D7	34250	14,2	*H. sapiens*
AP018105	*P. fieldi*	ABI	34377	14,3	*M. nemestrina, M. fascicularis*
AP018104	*P. fragile*	Hackeri	34375	14,2	*M. radiata, M. mulatta, Prebytis spp.*
CM003884	*P. gaboni*	SY75	29387	13	*Pan troglodytes*
AP018109	*P. gonderi*	ATCC 30045	34039	14,1	*Cercocebus atys*
AP018107	*P. hylobati*	WAK	34341	14,2	*Hylobati moloch*
AP018108	*P. inui*	Celeves	34401	14	*M. radiata, M. mulatta, Prebytis spp.*
AP018103	*P. knowlesi*	ATCC 30158	34332	14,5	*M. nemestrina, M. fascicularis, M. nigra*
AB649418	*P. malariae*	Kisii67	28968	12,5	*Homo sapiens*
AB649417	*P. ovale*	Nigeria II	29075	13,1	*H. sapiens*
AP018106	*P. simiovale*	ATCC 30104	34354	14,2	*M. sinica*
AB649419	*P. vivax*	Salvador I	29093	13,2	*H. sapiens*
AB649422	*P. yoelii*	17NXL	29227	13,9	*Thamnomys sp.*

Table 7-1.: Complete list of haemosporidian species and supporting information about the sequences included in this investigation.

Given that substitution saturation declines phylogenetic signal contained in the sequences, the entropy-based index of substitution saturation (Xia *et al.*, 2003) was estimated using Dambe v6.4.81 (Xia, 2017) for each gene and nonprotein coding regions. The estimation of the phylogenetic signal using this method was performed separately for the first, second, and third codon positions of genes, and the joined sites (1st+2nd and all sites).

For codon usage analyses all concatenated CDS and CDS greater than 500bp were taken into account. Also, ApiGenome sequences from other Apicomplexan species were included (Table dfhdfghfghfghgfhf) in order to evaluate how codon usage has changed in organisms that contain the same genes in ancestral or similar organelles. GC content, Effective Number Codon (**ENc**), and

Relative Synonymous Codon Usage (**RSCU**) were measured using Dambe v6.4.81 (Xia, 2017). The ENC measures the degree of codon bias, which correlates negatively with codon usage bias; values ranging from 20 (use of a single codon per amino acid) to 61 (use of all codons). Therefore, when ENC = 20, it means a complete bias towards a synonymous codon, while ENC = 61 indicates a neutral codon usage (Wright, 1990). Besides, the RSCU shows the deviation of synonymous codon usage from their even usage. Thus, RSCU is a measure that represents the ratio between the observed frequency of a synonymous codon and the expected frequency of that codon when all codons are used in a similar mode for a specific amino acid; the values of RSCU > 1 indicate a codon usage preference, RSCU < 1 indicates a less frequent usage of that codon, and RSCU $= 1$ indicates that there is no bias in the codon usage (Goldman and Yang, 1994).

7.3.5. Phylogenetic Analyses and Estimation of Relative evolutionary rates

Phylogenetic relationships were inferred based on three different approaches: complete ApiGenome, CDS without substitution saturation (Table sdfggafdg) and *clpC* gene. Bayesian methods implemented in MrBayes v3.2.7a with the default parameters (Ronquist and Huelsenbeck, 2003) and the Maximum Likelihood (ML) analysis in W-IQ-Tree (Trifinopoulos *et al.*, 2016) were used to estimate the trees. For both phylogenetic methods, a general time reversible model with gamma-distributed substitution rates and a proportion of invariant sites (GTR+Γ+I) was used. This model was the one with the lowest Bayesian Information Criterion (BIC) scores for all alignments as estimated by MEGAv7 (Kumar *et al.*, 2016). Bayesian support for all nodes was inferred by sampling every 500 generations from two independent chains lasting $4x10^6$ Markov Chain Monte Carlo (MCMC) steps, and for ML analyses, 1,000 bootstrap replicates were evaluated for statistical confidence.

Estimation of relative evolutionary rates (non-calibrated) was calculated in a non-Bayesian context with RelTime method. Here, calculations were accomplished on the command line version of MEGAv7 (Kumar *et al.*, 2016) and the substitution model was the same as the one used for Bayesian analyses.

7.3.6. *clpC* Primers design

Due to the results obtained in the analyses of the present study, *clpC* gene primers were designed using conserved regions, based on the previously constructed alignment, following the conditions recommended by Jennings (2017): **(i)** optimum length varies between 18 and 30 bp; **(ii)** optimum melting temperatures are in the range of 52-58ºC; **(iii)** optimal GC content; **(iv)** primers have one or two G and / or C bases within the last five bases from the 3 'end of the primers; **(v)** secondary structures of the primer, such as the forks and the cross dimer, were avoided; **(vi)** in the case of degenerate primers, only a maximum of four positions were allowed in the oligonucleotide containing a mixture of base pairs. All these requirements were verified using the Oligo Calc online tool: Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html).

7.4. Results

7.4.1. Features of *Haemoproteus columbae*) ApiGenome

Haemoproteus columbae apicoplast genome is a circular molecule of 29.8 Kbp and 87.7 % A+T rich. Besides, it has a coding density of over 98 % encoding the elongation factor Tu (tufA), clpC chaperone, FeS cluster assembly protein SufB, seven ORFs, four subunits of the RNA polymerase, 17 ribosomal proteins, 25 tRNAs, and one copy of SSU and LSU rRNAs (Fig. 1). **ATG** is the unique start codon for all 31 coding sequences (CDSs) and the stop codons of the CDSs are **TAA** (24 of the 31 CDSs), **TGA** (6 of the 31 CDSs) and **TAG** (1 of the 31 CDSs).

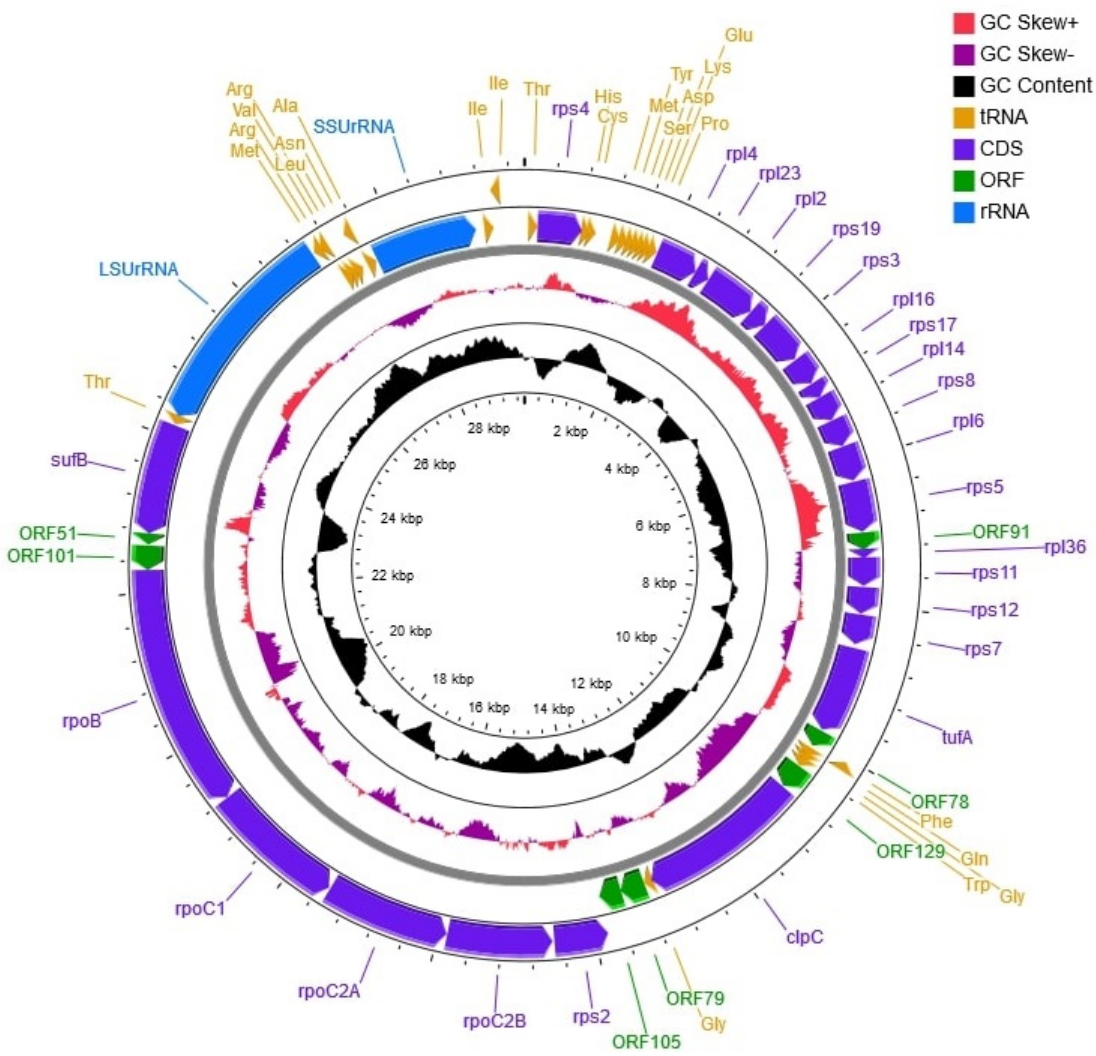


Figure 7-1.: Graphical representation of the H. columbae ApiGenome. The map was designed using CGView Server^{BETA}. From outside to center: genes (3'-5'), genes (5'-3'), GC skew, % G+C, and base coordinates

Likewise, the ApiGenome was compared with other haemosporidian species (genera *Leucocytozoon* and *Plasmodium*) and two species of the Piroplasmida order (*Theileria parva* and *Babesia bovis*). Haemosporidian parasites have highly structural conservation of the genome in the number and organization of genes. However, *H. columbae* possesses a reduction of the genome corresponding to a single copy of the rRNAs (as occurs in other parasites of the genus *Plasmodium* i.e. *P. chabaudi*) compared to other Haemosporidia that present double copy of these genes as *P. falciparum* or *L. caulleryi* (Fig. **7-2**). Regarding to the piroplasmids parasites, there is basal conservation in the organization of some genes, but there are changes in the number, copies and size of them (Fig. **7-2**).

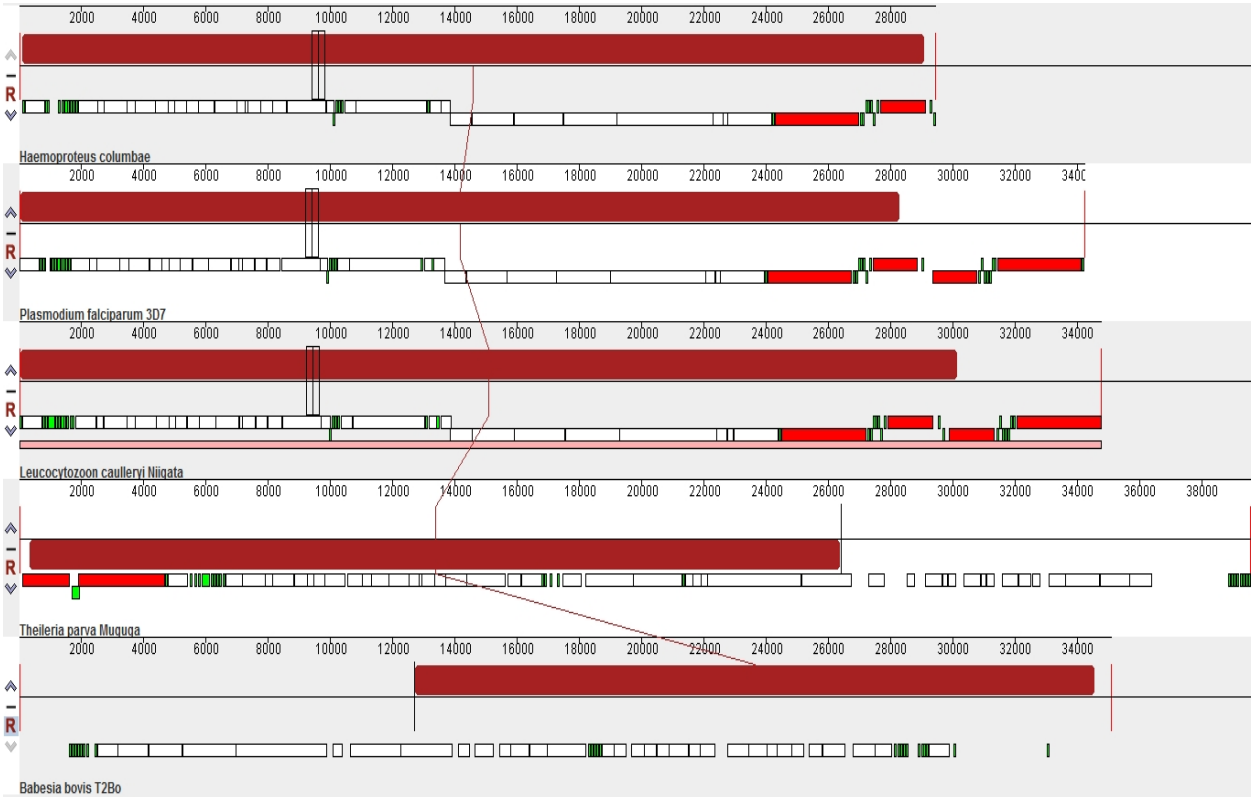


Figure 7-2.: Schematic representation of synteny in the ApiGenomes of different genera of apicomplexa phylum. Comparison was performed using Mauve. The burgundy color bars between DNA sequences represent regions highly conserved. The white, red and green bars indicate CDS, tRNAs and rRNAs, respectively

7.4.2. Phylogenetical hypotheses of ApiGenome within Haemosporida order

Based on alignment done with the ApiGenomes from Table 1S, index of substitution saturation for each CDS was calculated, allowing to unravel whether the sequence is useful or not for phylogenetic purposes. Therefore, for all sites, only 19 of 31 genes display little saturation (LS; $Iss < Iss.c, p < 0{,}05$) and the other genes have a low or very low phylogenetic signal (Table **7-2**).

Gene name	1st position							
	P (IS.)	**Iss**	**Iss.c**	**p***	**Iss.c Asym**	**p***	**DF**	**phylogenetic signal**
rps4	0,253	0,438	0,610	0,032	0,367	0,377	159	Little saturation
rpl4	0,177	0,993	0,625	0,00	0,374	0,00	193	Useless sequences
rpl23	0,150	1,079	0,418	0,00	0,362	0,00	73	Useless sequences
rpl2	0,4377	0,3115	0,6333	0	0,3785	0,3879	140	Little saturation
rps19	0,3992	0,4731	0,439	0,7558	0,3582	0,2968	54	Very poor signal
rps3	0,2	0,5571	0,6198	0,4654	0,3714	0,0315	181	Substantial saturation
rpl16	0,5569	0,5008	0,5242	0,791	0,3472	0,0862	58	Substantial saturation
rps17	-	-	-	-	-	-	-	-
rpl14	0,2403	0,2889	0,5041	0,0075	0,3473	0,4605	91	Little saturation
rps8	0,2182	0,2946	0,5194	0	0,3471	0,3467	101	Little saturation
rpl6	0,1598	0,6358	0,5854	0,5671	0,3577	0,0019	154	Very poor signal
rps5	0,1909	0,8285	0,644	0,0561	0,3849	0	217	Very poor signal
rpl36	-	-	-	-	-	-	-	-
rps11	0,3208	0,451	0,5303	0,4746	0,3475	0,3508	92	Substantial saturation
rps12	0,3655	0,3383	0,5178	0,1377	0,347	0,9418	81	Substantial saturation
rps7	0,019	0,7469	0,5443	0,0005	0,3488	0	143	Useless sequences
tufA	0,3075	0,1404	0,6943	0	0,4246	0	284	Little saturation
clpC	0,2218	0,3799	0,7537	0	0,4879	0,0051	615	Little saturation
rps2	0,4705	0,5819	0,6237	0,6382	0,3734	0,0202	123	Substantial saturation
*RPOC2B**	0,2476	0,6365	0,7093	0,2354	0,4464	0,0021	358	Substantial saturation
RPOC2A	0,2255	0,5085	0,726	0,0001	0,4537	0,3265	432	Little saturation
RPOC1	0,6049	0,3869	0,7303	0	0,4576	0,2419	230	Little saturation
RPOB	0,269	0,4346	0,7701	0	0,5198	0,0351	786	Little saturation
SufB	0,3012	0,1764	0,1087	0	0,4394	0	331	Useless sequences
*ORF91**	0,48747	0,8354	0,3998	0,0126	0,359	0,0067	42	Useless sequences
ORF78	0,1619	0,8349	0,4418	0,0203	0,3574	0,0052	77	Useless sequences
ORF129	0,3468	0,6021	0,5162	0,4261	0,347	0,0199	83	Very poor signal
ORF79	0,066	0,6151	0,0082	0,0108	0,7723	0,4936	25	Useless sequences
ORF105	0,1167	0,7358	0,4868	0,037	0,3486	0,0014	98	Useless sequences
ORF101	0,1976	1,2304	0,4966	0,0001	0,3477	0	93	Useless sequences
ORF51	0,1952	1,3648	0,3447	0	0,4	0	52	Useless sequences
Gene name	**2nd position**							
	P (IS.)	**Iss**	**Iss.c**	**p***	**Iss.c Asym**	**p***	**DF**	**phylogenetic signal**
rps4	0,269	0,372	0,610	0,004	0,367	0,947	155	Little saturation
rpl4	0,382	1,106	0,625	0,00	0,374	0,00	145	Useless sequences
rpl23	0,380	1,318	0,424	0,00	0,363	0,00	53	Useless sequences
rpl2	0,5765	0,3189	0,633	0	0,3758	0,515	105	Little saturation
rps19	0,4978	0,3644	0,439	0,5637	0,3582	0,9614	45	Substantial saturation
rps3	0,2603	0,4763	0,6198	0,1231	0,3714	0,259	168	Substantial saturation
rpl16	0,4502	0,3493	0,5242	0,0388	0,3472	0,9798	72	Little saturation
rps17	-	-	-	-	-	-	-	-
rpl14	0,2667	0,2137	0,5041	0	0,3473	0,0372	88	Little saturation
rps8	0,3566	0,302	0,5194	0,0012	0,3471	0,4905	83	Little saturation
rpl6	0,3932	0,8072	0,5854	0,0568	0,3577	0,0002	111	Very poor signal
rps5	0,2449	0,7448	0,644	0,3386	0,3849	0,0008	203	Very poor signal
rpl36	-	-	-	-	-	-	-	-
rps11	0,3483	0,3393	0,5303	0,0997	0,3475	0,9432	88	Substantial saturation
rps12	0,3594	0,2941	0,5178	0,0499	0,347	0,6388	82	Little saturation
rps7	0,0016	0,7549	0,5474	0,0001	0,3707	0	146	Useless sequences
tufA	0,3522	0,0843	0,6943	0	0,4246	0	265	Little saturation
clpC	0,3016	0,3366	0,7537	0	0,4879	0,0005	551	Little saturation
rps2	0,2703	0,2883	0,6237	0	0,3734	0,2188	170	Little saturation
*RPOC2B**	0,5483	0,9429	0,7093	0,0088	0,4464	0	214	Useless sequences

RPOC2A	0,2803	0,4711	0,726	0	0,4537	0,7761	401	Little saturation
RPOC1	0,2081	0,1634	0,7303	0	0,4576	0	461	Little saturation
RPOB	0,3318	0,3764	0,7701	0	0,5198	0,0013	719	Little saturation
SufB	0,3534	0,1249	0,7087	0	0,4394	0	306	Little saturation
*ORF91**	0,2581	0,4344	0,3998	0,7849	0,359	0,5528	61	Very poor signal
ORF78	0,3006	0,8656	0,4418	0,0333	0,3574	0,0113	64	Useless sequences
ORF129	0,2617	0,4646	0,5168	0,6035	0,347	0,2376	94	Substantial saturation
ORF79	-	-	-	-	-	-	-	-
ORF105	0,307	0,6356	0,4868	0,3178	0,3486	0,0561	77	Very poor signal
ORF101	0,2971	1,3162	0,4966	0,0001	0,3477	0	81	Useless sequences
ORF51	0,1262	1,1354	0,3447	0,0002	0,4	0,0004	57	Useless sequences
	1st and 2nd position							
Gene name	**P (IS.)**	**Iss**	**Iss.c**	**p***	**Iss.c Asym**	**p***	**DF**	**phylogenetic signal**
rps4	0,267	0,406	0,699	0,000	0,429	0,684	313	Little saturation
rpl4	0,210	0,938	0,708	0,01	0,439	0,00	372	Useless sequences
rpl23	0,175	1,050	0,576	0,00	0,355	0,00	142	Useless sequences
rpl2	0,6047	0,3899	0,7135	0	0,4447	0,4403	197	Little saturation
rps19	0,5391	0,4994	0,5854	0,3681	0,3577	0,1395	84	Substantial saturation
rps3	0,2414	0,5215	0,7048	0,0042	0,4352	0,176	345	Little saturation
rpl16	0,2557	0,2713	0,6419	0	0,3837	0,0155	197	Little saturation
rps17	-	-	-	-	-	-	-	-
rpl14	0,281	0,26	0,6286	0	0,376	0,0281	173	Little saturation
rps8	0,2279	0,275	0,6388	0	0,3817	0,0077	200	Little saturation
rpl6	0,1889	0,6333	0,6863	0,4524	0,4136	0,0008	297	Substantial saturation
rps5	0,2224	0,7859	0,7224	0,3733	0,4507	0	419	Very poor signal
rpl36	-	-	-	-	-	-	-	-
rps11	0,2132	0,3217	0,646	0	0,3862	0,3457	215	Little saturation
rps12	0,3499	0,3065	0,6377	0,0001	0,3811	0,3505	167	Little saturation
rps7	0,0091	0,7515	0,6552	0,0146	0,3924	0	290	Useless sequences
tufA	0,345	0,1134	0,7561	0	0,492	0	537	Little saturation
clpC	0,2805	0,3657	0,7844	0	0,5555	0	1137	Little saturation
rps2	0,2668	0,3436	0,7073	0	0,4379	0,0527	342	Little saturation
*RPOC2B**	0,2908	0,6327	0,7649	0,0041	0,5141	0,01	676	Little saturation
RPOC2A	0,2619	0,4948	0,7718	0	0,5236	0,4904	824	Little saturation
RPOC1	0,4771	0,2685	0,7737	0	0,5278	0	610	Little saturation
RPOB	0,3044	0,4058	0,7942	0	0,5782	0	1497	Little saturation
SufB	0,3309	0,1493	0,7641	0	0,5071	0	635	Little saturation
*ORF91**	0,5006	0,7321	0,5636	0,1689	0,3553	0,0026	82	Very poor signal
ORF78	0,4176	1,1106	0,5872	0,001	0,3583	0	107	Useless sequences
ORF129	0,2537	0,4935	0,6366	0,0396	0,3805	0,1036	190	Little saturation
ORF79	0,181	0,9041	0,2907	0,0047	0,4663	0,0282	45	Useless sequences
ORF105	0,3188	0,7857	0,6171	0,1056	0,3701	0	152	Useless sequences
ORF101	0,2826	1,3266	0,6237	0	0,3734	0	167	Useless sequences
ORF51	0,2683	1,4181	0,5226	0	0,3471	0	96	Useless sequences
	3rd position							
Gene name	**P (IS.)**	**Iss**	**Iss.c**	**p***	**Iss.c Asym**	**p***	**DF**	**phylogenetic signal**
rps4	0,1638	0,5912	0,6101	0,788	0,3669	0,0016	178	Substantial saturation
rpl4	0,1618	1,2577	0,6249	0,000	0,3740	0,0000	197	Useless sequences
rpl23	0,0769	1,1048	0,4311	0,000	0,3741	0,0000	79	Useless sequences
rpl2	0,1400	0,5060	0,6333	0,009	0,3785	0,0092	214	Little saturation
rps19	0,1797	0,4882	0,4328	0,484	0,3593	0,1060	74	Very poor signal
rps3	0,0889	0,6584	0,6198	0,606	0,3714	0,0000	207	Very poor signal

rpl16	0,0870	0,5295	0,5186	0,830	0,3474	0,0000	120	Very poor signal
rps17	-	-	-	-	-	-	-	-
rpl14	0,0914	0,4920	0,5041	0,827	0,3473	0,0101	109	Substantial saturation
rps8	0,0732	0,4497	0,5194	0,165	0,3471	0,0421	119	Substantial saturation
rpl6	0,1480	0,8139	0,5854	0,0085	0,3577	0,0000	156	Useless sequences
rps5	0,0630	0,9273	0,6440	0,000	0,3849	0,0000	252	Useless sequences
rpl36	-	-	-	-	-	-	-	-
rps11	0,0714	0,5478	0,5303	0,828	0,3475	0,0138	126	Very poor signal
rps12	0,1208	0,6612	0,5178	0,078	0,3470	0,0000	112	Very poor signal
rps7	0,013	0,9255	0,5443	0,000	0,349	0,00	144	Useless sequences
tufA	0,148	0,4374	0,6943	0,000	0,425	0,57	349	Little saturation
clpC	0,140	0,6271	0,7544	0,000	0,495	0,00	680	Little saturation
rps2	0,253	0,5699	0,6237	0,406	0,373	0,00	174	Substantial saturation
*RPOC2B**	0,175	0,8418	0,7093	0,016	0,446	0,00	393	Useless sequences
RPOC2A	0,175	0,7577	0,7260	0,535	0,454	0,00	460	Very poor signal
RPOC1	0,101	0,4093	0,7303	0,000	0,458	0,10	524	Little saturation
RPOB	0,165	0,6630	0,7701	0,002	0,520	0,00	899	Little saturation
SufB	0,1016	0,4135	0,7087	0,000	0,4394	0,3601	426	Little saturation
*ORF91**	0,129	0,6780	0,3998	0,022	0,359	0,01	71	Useless sequences
ORF78	0,340	1,0850	0,4418	0,002	0,357	0,00	60	Useless sequences
ORF129	0,284	0,7124	0,5162	0,043	0,347	0,00	91	Useless sequences
ORF79	0,063	0,7401	0,4150	0,021	0,366	0,01	78	Useless sequences
ORF105	0,266	0,8431	0,4868	0,010	0,349	0,00	81	Useless sequences
ORF101	0,060	1,2433	0,4966	0,000	0,348	0,00	109	Useless sequences
ORF51	0,047	1,1221	0,3447	0,000	0,400	0,00	62	Useless sequences
	All positions**							
Gene name	**P (IS.)**	**Iss**	**Iss.c**	**p***	**Iss.c Asym**	**p***	**DF**	**phylogenetic signal**
rps4	0,3400	0,5321	0,7387	0,000	0,4666	0,1921	423	Little saturation
rpl4	0,1883	1,0017	0,7463	0,000	0,4764	0,0000	574	Useless sequences
rpl23	0,1491	1,0494	0,6393	0,000	0,3821	0,0000	221	Useless sequences
rpl2	0,4880	0,4227	0,7503	0,000	0,4824	0,1822	383	Little saturation
rps19	0,4363	0,4973	0,6469	0,022	0,3868	0,0895	155	Little saturation
rps3	0,2166	0,5691	0,7437	0,000	0,4729	0,0544	535	Little saturation
rpl16	0,2070	0,3449	0,6911	0,000	0,4217	0,0288	315	Little saturation
rps17	0,1060	1,8598	0,6895	0,000	0,4201	0,0000	350	Useless sequences
rpl14	0,2319	0,3294	0,6807	0,000	0,4123	0,0361	278	Little saturation
rps8	0,1837	0,3368	0,6887	0,000	0,4194	0,0094	317	Little saturation
rpl6	0,1871	0,6894	0,7247	0,501	0,4526	0,0000	448	Substantial saturation
rps5	0,1957	0,8426	0,7552	0,115	0,4904	0,0000	651	Very poor signal
rpl36	0,3469	0,1144	0,4847	0,000	0,3488	0,0000	71	Little saturation
rps11	0,4506	0,5746	0,6943	0,105	0,4246	0,0425	225	Substantial saturation
rps12	0,2714	0,3787	0,6878	0,000	0,4186	0,4928	281	Little saturation
rps7	0,012	0,7923	0,7015	0,005	0,432	0,00	435	Useless sequences
tufA	0,281	0,1984	0,7760	0,000	0,533	0,00	885	Little saturation
clpC	0,240	0,4325	0,7978	0,000	0,584	0,00	1802	Little saturation
rps2	0,418	0,5142	0,7457	0,000	0,476	0,41	408	Little saturation
*RPOC2B**	0,251	0,6786	0,7825	0,003	0,553	0,00	1070	Little saturation
RPOC2A	0,244	0,5665	0,7856	0,000	0,560	0,85	1266	Little saturation
RPOC1	0,278	0,2756	0,7868	0,000	0,564	0,00	1265	Little saturation
RPOB	0,264	0,4713	0,8051	0,000	0,600	0,00	2378	Little saturation
SufB	0,2884	0,2295	0,7818	0,000	0,5461	0	1013	Little saturation
*ORF91**	0,229	0,5664	0,6313	0,374	0,383	0,01	191	Little saturation
ORF78	0,413	1,1346	0,6484	0,000	0,388	0,00	163	Useless sequences

ORF129	0,240	0,5378	0,6870	0,007	0,418	0,03	291	Little saturation
ORF79	0,330	0,8248	0,6344	0,070	0,379	0,00	168	Very poor signal
ORF105	0,299	0,7961	0,6718	0,130	0,405	0,00	235	Very poor signal
ORF101	0,173	1,2156	0,6769	0,000	0,409	0,00	289	Useless sequences
ORF51	0,197	1,3010	0,5976	0,000	0,362	0,00	158	Useless sequences

Table 7-2.: Entropy-based index of substitution saturation (**Iss**) for the first, second, third, first plus second, and all positions together of the apicoplast genes. Iss estimates were estimated with Dambe v6.4.81 (Xia, 2017). Analyses performed on gap-free sites only using a two-tailed test

Interestingly, depending on the method of phylogenetic inference the clade of *P. vivax* and *P. cynomolgi* (*P.v-P.c*) goes to a basal or derived position of the Asian primate malaria parasite clade. This result was evident for the complete ApiGenome, due to Bayesian inference result for *Pv-Pc* clade was a derived position and all branches support were statistically significant (equal to **1 PP**; Fig. **7-3**A). In contrast, using ML the position of *Pv-Pc* clade was basal, and the support of some branches were not statistically significant (Fig. **7-3**B).

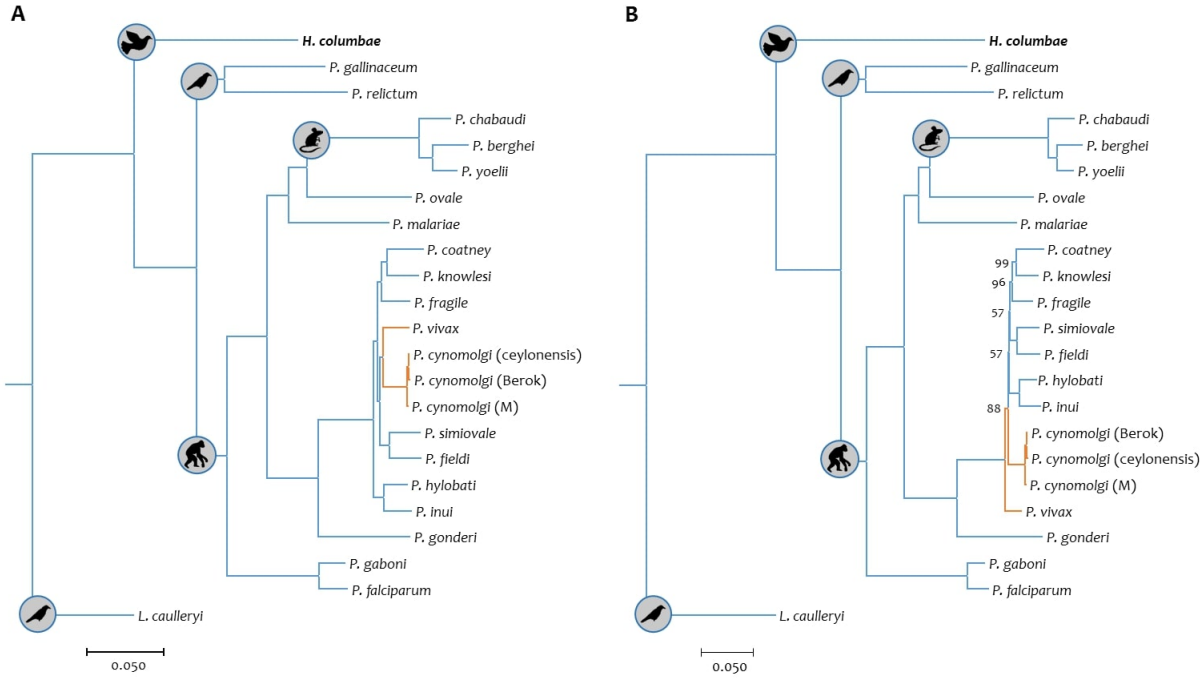


Figure 7-3.: Phylogenetic hypotheses of haemosporidian parasites based on complete ApiGenomes. A) Bayesian Inference. All values at the nodes are posterior probabilities equal 1. B) Maximum Likelihood hypothesis. All nodes are bootstrap values as a percentage obtained for 1,000 pseudoreplicates (nodes without value are equal to 100)

Likewise, the result found above was obtained anew independently of the method of phylogenetic inference but using different input data. For phylogenetic hypotheses of the 19 genes with little

saturation (Table **7-2**), the position of *Pv-Pc* clade was basal (Fig. **7-4**A), and for *clpC* gene *Pv-Pc* clade was a derived position (Fig. **7-4**B). However, for both hypotheses the support of some branches were not statistically significant.

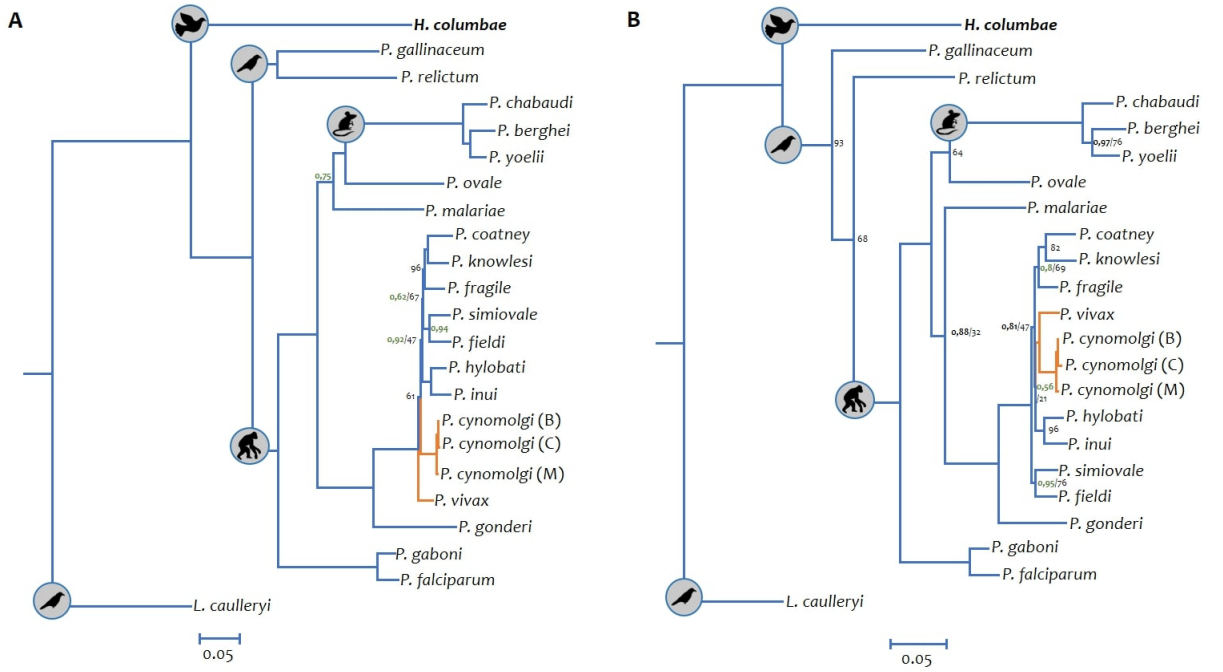


Figure 7-4.: Figure S2. Phylogenetic hypotheses of haemosporidian parasites based on difference approaches. The values at the nodes for Bayesian inference are posterior probabilities (green) together with bootstrap values (black) as a percentage obtained for 1,000 pseudoreplicates from a maximum likelihood tree with identical topology; the nodes without values are posterior probabilities equal 1 and/or bootstrap values equal to 100. A) Phylogenetic hypothesis based on CDS with little saturation for all positions (Table S2). B) Phylogenetic hypothesis based on clpC gene.)

Regardless of the topology calculated for complete Apicoplast using Bayesian inference (Fig. **7-4**A) or maximum likelihood (Fig. **7-4**B), the result obtained by RelTime was similar for each branch, which indicates that with respect to the root calculated by the RelTime method, none ApiGenoma does present evidence of accelerated relative evolutionary rates in any of the clades (Fig. **7-5**).

In addition, this analysis was carried out in order to know how was the *clpC* gene relative evolutionary rates. Nevertheless, the result was similar to apigenomas result, in which none terminal branch presented evidence does heterogeneous values (Fig. **7-6**).

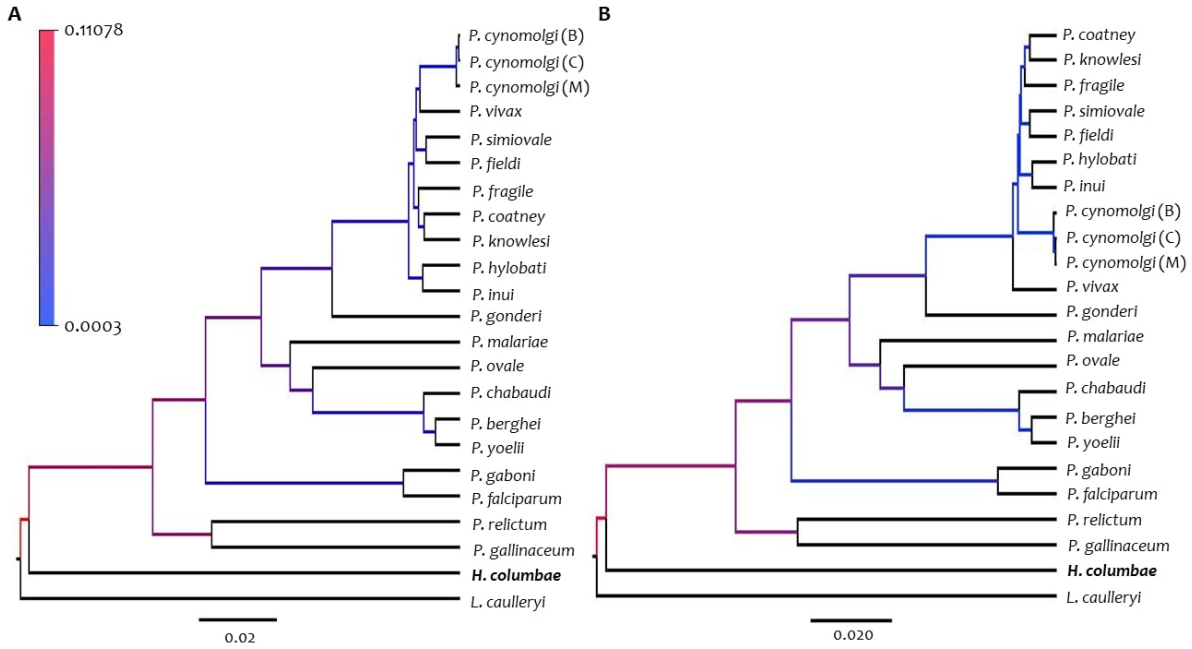


Figure 7-5.: Estimation of Relative evolutionary rates for whole ApiGenomes. Branches are colored according to their relative rates to the root rate (that is set to one) estimated from RelTime without calibration constraints. a) Topology calculated by inference bayesian, b) Topology calculated by Maximun likelihood

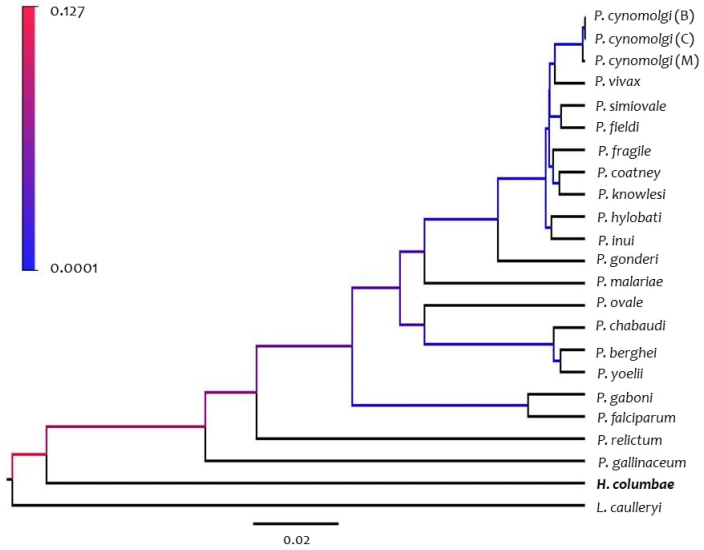


Figure 7-6.: Estimation of Relative evolutionary rates for *clpC* gene. Branches are colored according to their relative rates to the root rate (that is set to one) estimated from RelTime without calibration constraints

7.4.3. Molecular Evolution of ApiGenome

The analyses of codon usage in CDS of Haemosporidian parasites ApiGenome were determined and compared based on the **RSCU** and **ENc**. The RSCU values of each codon within the 21 haemosporidian ApiGenomes are provided in Table **7-3** (only Berok strain of *P. cynomolgi* was taken into account).

Almost all the 59 codons found (excluding **AUG** and **UGG**) were present in all species, except for **CGC** (absent in all species), **CUG** (extant only in *P. berghei*), and **GGC** (present in *P. berghei, P. hylobati* and *P. inui*). 29 codons were found with high-frequency (RSCU>1) codons, and the remaining were low-frequency codons (Fig. 3). It should be noted that codons with RSCU values < 1 (31 of 33 codons) have nucleotides G or C in the third position (Fig. **7-7**; Table **7-3**).

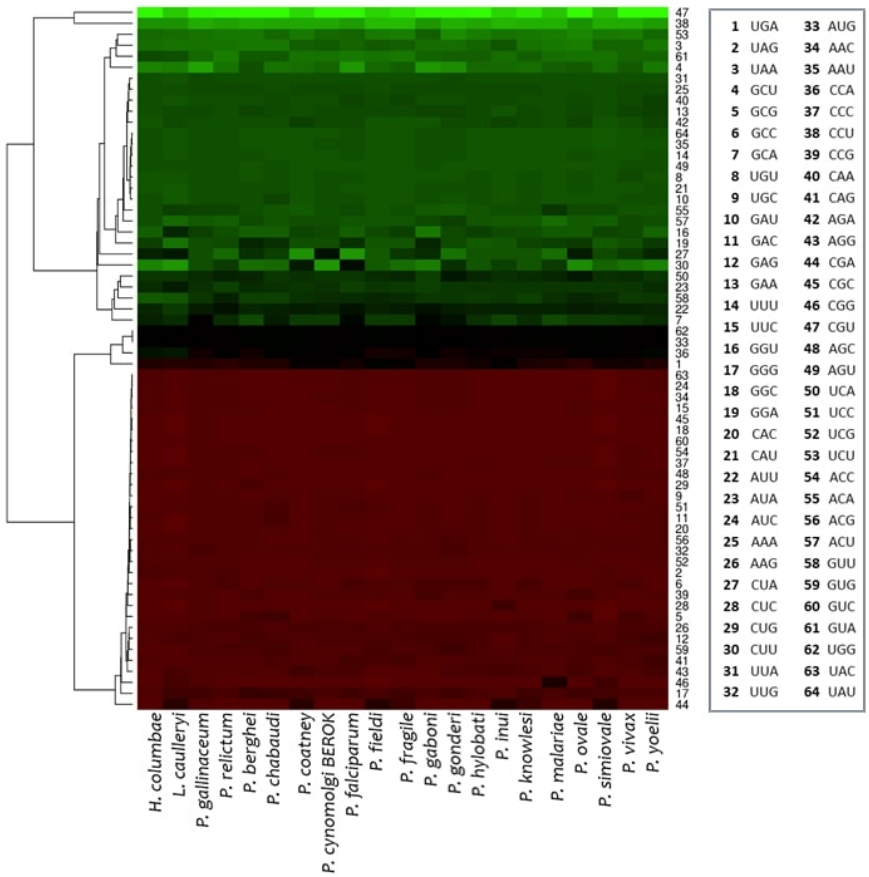


Figure 7-7.: Hierarchical cluster by average and heat map of the relative synonymous codon usage (RSCU) values of each codon in the CDS of Haemosporida ApiGenomes. Each square in the heat map represents the RSCU value of each codon (in rows) within the CDS of each Haemosporida ApiGenome (in columns). Colours indicate the magnitude of RSCU values: black, RSCU=1 (no bias in codon usage); green, RSCU>1; and red, RSCU<1

Haemosporida order											
Codon	**AA**	**Average**	**Codon**	**AA**	**Average**	**Codon**	**AA**	**Average**	**Codon**	**AA**	**Average**
CGC	R	0	GAC	D	0.0359	CCA	P	0.93015	GAU	D	1.9641
GGC	G	0.0037	UCC	S	0.0398	AUG	M	1	CAU	H	1.96465
CUG	L	0.00715	ACG	T	0.04715	UGG	W	1	UGU	C	1.9709
GUC	V	0.01055	CGG	R	0.04765	AUU	I	1.36025	AGU	S	1.9769
AUC	I	0.01325	GCC	A	0.0546	GCA	A	1.5265	UUU	F	1.97795
ACC	T	0.01395	CCG	P	0.0564	UCA	S	1.53055	GGU	G	1.9816
CUC	L	0.0143	UUG	L	0.0657	AUA	I	1.6265	AAU	N	1.9822
UAG	*	0.0148	GCG	A	0.0659	GUU	V	1.6276	UAU	Y	1.9831
UAC	Y	0.0169	AGG	R	0.08675	GGA	G	1.82425	ACU	T	1.9977
AAC	N	0.0178	CAG	Q	0.0956	GAA	E	1.88505	CUU	L	2.08285
CCC	P	0.02205	AAG	K	0.0981	CUA	L	1.8957	GUA	V	2.2469
UUC	F	0.02205	GAG	E	0.11495	AAA	K	1.9019	UAA	*	2.2606
AGC	S	0.0231	GUG	V	0.115	CAA	Q	1.9044	GCU	A	2.345285
UGC	C	0.0291	CGA	R	0.13955	AGA	R	1.91325	UCU	S	2.39395
CAC	H	0.03535	GGG	G	0.1904	UUA	L	1.9343	CCU	P	2.9913
UCG	S	0.03585	UGA	*	0.7246	ACA	T	1.9412	CGU	R	3.8128

Table 7-3.: Relative Synonymous Codon Usage for all CDS of each species Haemosporida.

ENc values < 35 indicate a strong codon usage bias. In the case of the complete ApiGenomes (ENc mean = 44,0510), there was not evidence of codon usage bias, including *P. semiovale* (39,8316) which presented the lowest value. Likewise, none ApiGenome had codon usage bias for all concatenated CDS (ENc mean = 45,5052). However, ENc value for each CDS ($> 500bp$) showed that *tufA* (ENc mean = 38,7004), *clpC* (ENc mean = 37,4491), *RPOC1* (ENc mean = 38,5471), *RPOB* (ENc mean = 36,7863) and *SufB* (ENc mean = 37,1294) genes present a weak codon usage bias (Table **7-4** and Fig. **7-8**).

	Genome	**All coding sequences**	**Genes (>500bp)**					
Species	**All sites**	**All sites**	***rps4***	***rpl4***	***rpl2***	***rps3***	***rpl6***	***rps5***
H. columbae	44,0937	45,5901	45,9846	45,3562	40,7819	44,4057	43,5827	43,2728
L. caulleryi	45,1315	46,0678	44,5586	46,0257	39,3818	42,5695	46,0577	42,6223
P. gallinaceum	43,4777	45,3313	45,7906	45,6638	41,5978	45,9676	44,3301	44,8189
P. relictum	40,8158	43,119	45,6930	47,4576	41,7052	44,4672	44,3301	40,2211
P. chabaudi	46,4004	46,0607	44,5013	45,5379	40,1670	43,1400	47,2211	45,5351
P. berghei	46,0049	46,2111	43,4043	44,2975	40,5598	43,2698	44,1322	47,3448
P. yoelii	42,8891	46,0143	42,5697	44,8437	40,1938	42,8178	43,4651	44,4212
P. malariae	45,9633	45,8019	46,5066	46,3395	40,9670	46,2415	47,7336	45,4561
P. ovale	44,3506	45,1703	47,1704	45,3188	42,8769	43,7700	47,4102	43,8854
P. coatney	43,1765	45,3005	44,0087	45,2382	40,4985	42,9050	47,5001	44,2561
P. vivax	43,1536	45,9492	42,1970	45,6770	40,3981	45,3489	46,8748	44,8106
P. gaboni	46,3629	45,6521	44,5586	44,1388	41,3846	45,2641	46,2008	43,0339
P. falciparum	43,5582	45,0301	46,2940	46,3815	40,7994	44,9087	45,7180	42,9662
P. gonderi	44,895	45,2158	46,6815	44,1804	42,1492	45,6174	46,0711	43,9969
P. knowlesi	45,1681	45,8178	44,4049	44,4962	40,8808	45,1708	47,2240	45,8544
P. hylobati	43,6414	45,7098	44,9011	44,7260	39,6851	44,8041	47,2096	45,4220
P. inui	41,87	45,2535	46,7679	45,0444	40,0019	44,9485	48,1840	44,6839
P. fragile	44,0092	45,5171	44,3737	45,6816	39,8478	44,6269	46,2855	42,5300
P. cynomolgi	44,236	45,2885	43,0522	45,1256	40,9141	43,8949	46,9189	45,3560
P. simiovale	39,8316	45,775	43,1239	44,4347	40,0410	43,1930	48,2142	44,6251
P. fieldi	46,0407	45,7332	44,4523	43,0497	40,9222	44,6398	46,2963	44,4468
Mean	**44,0510**	**45,5052**	**44,8093**	**45,1912**	**40,7502**	**44,3796**	**46,2362**	**44,2647**
STD	**1,7641**	**0,6426**	**1,4529**	**0,9583**	**0,8484**	**1,0831**	**1,4738**	**1,5001**
	Genes (>500bp)							
Species	***tufA***	***clpC***	***rps2***	***RPOC2B****	***RPOC2A***	***RPOC1***	***RPOB***	***SufB***
H. columbae	39,5065	38,4586	41,5279	36,9303	39,5449	39,2162	36,1315	39,1230
L. caulleryi	38,4079	38,6779	42,1754	42,2108	39,4961	38,0151	37,4530	37,8445
P. gallinaceum	37,7025	36,2636	42,8209	41,0350	39,3930	37,0921	36,2069	36,9303
P. relictum	37,8581	38,4678	44,6005	41,5812	40,3259	37,8297	36,1991	37,7992
P. chabaudi	41,2144	37,7574	43,1117	39,8816	40,7929	38,9928	35,8677	38,9328
P. berghei	40,1227	37,9299	42,7699	38,3891	40,5025	39,6089	37,7616	38,7264
P. yoelii	39,7131	37,4784	44,2184	39,9332	38,4332	38,8700	37,7833	38,8776
P. malariae	38,5134	37,3911	44,1670	-	39,9834	37,1343	37,4425	37,3597
P. ovale	38,1134	38,6858	43,1228	39,9673	41,1991	40,1509	36,7091	36,5959

P. coatney	39,1612	36,9281	44,2080	39,4618	40,1479	38,8018	36,3672	36,3438
P. vivax	38,6866	37,3985	44,3275	41,0037	41,4615	38,4472	36,4854	36,1366
P. gaboni	38,1641	37,4593	44,5784	41,1827	39,2263	38,4791	36,5173	36,9869
P. falciparum	38,2454	38,0362	44,7080	40,6640	40,3711	39,8738	37,0674	36,9551
P. gonderi	38,5664	35,2345	44,6169	39,0215	39,9865	39,0239	36,8050	36,5991
P. knowlesi	38,5894	38,4476	43,7241	40,6629	41,7128	39,3493	36,8799	35,6807
P. hylobati	36,8661	37,1169	44,6744	39,8819	42,2715	38,4097	35,7299	37,0309
P. inui	37,4610	37,2845	43,8409	39,4972	39,8928	38,2912	36,8283	36,0681
P. fragile	38,6452	36,3106	44,8542	41,5317	41,1427	37,4204	36,6060	35,6278
P. cynomolgi	39,2342	37,1245	42,9148	39,4488	41,0647	38,2654	36,9586	37,2560
P. simiovale	38,9842	37,1493	44,6418	41,4842	41,5604	38,0100	37,4716	36,5839
P. fieldi	38,9530	36,8314	44,1574	40,8182	41,8366	38,2074	37,2420	36,2598
Mean	**38,7004**	**37,4491**	**43,7981**	**40,2294**	**40,4927**	**38,5471**	**36,7863**	**37,1294**
STD	**0,9555**	**0,8764**	**0,9421**	**1,2517**	**0,9965**	**0,8339**	**0,5943**	**1,0645**

Table 7-4.: Effective Codon number for all ApiGenomes, all CDS and each gene (>500bp)

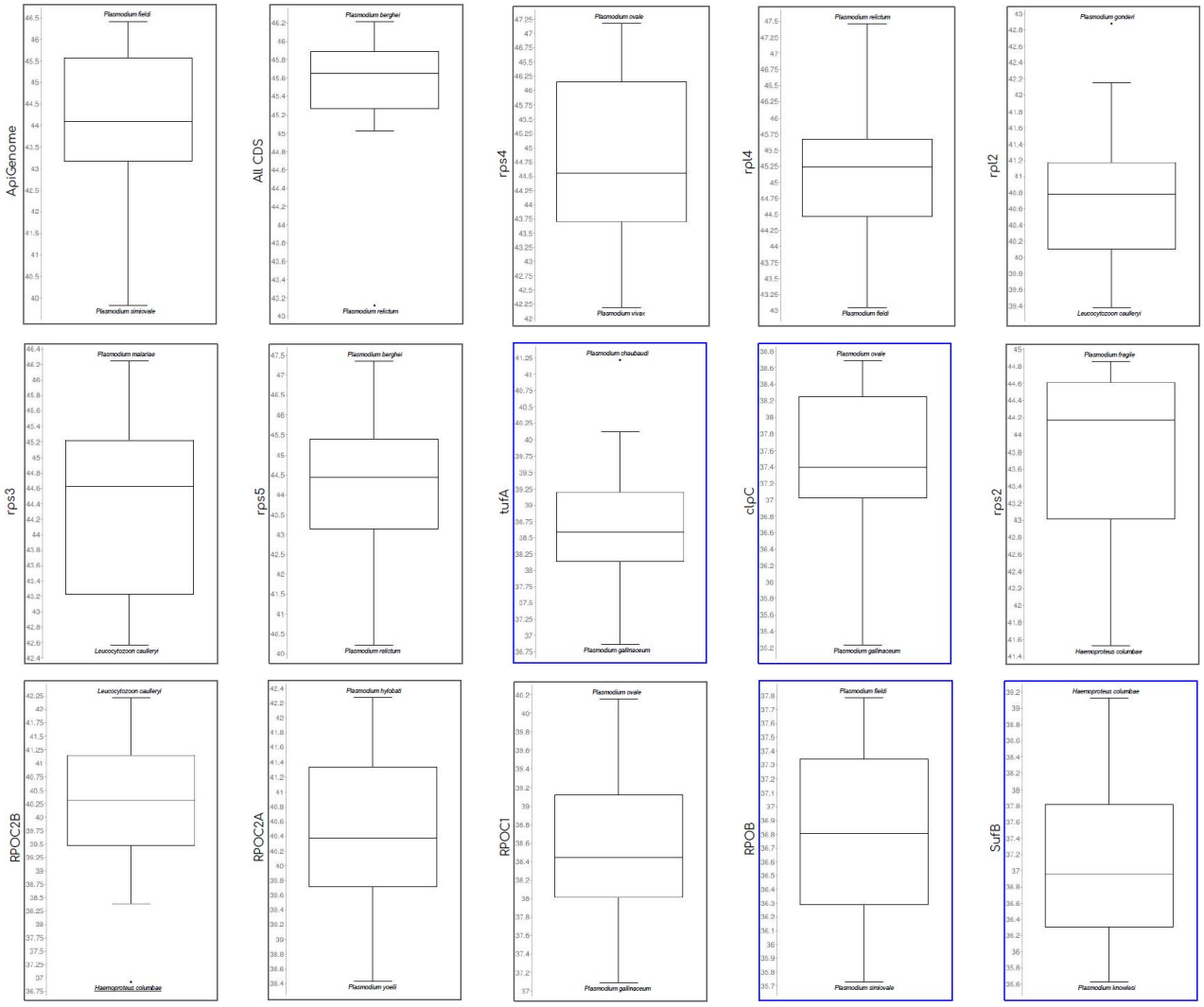


Figure 7-8.: Box-plots of the effective number of codons (**ENc**) in Haemosporida parasites. Box-plots with blue border indicate weak codon usage bias

In addition, comparisons were done of ENc values in orthologous genes present in the apicoplast or chloroplast genomes of ancestral organisms to haemosporidian parasites (Table **7-5**). Based on ENc values for each orthologous CDS ($> 500bp$) were able to evidence that Haemoporida ApiGenomes have a lower average value in comparison to ancestral genomes (apicoplast or chloroplast; Table **7-6** and Fig. **7-9**).

In *tufA*, *clpC*, *RPOC1* and *RPOB* genes, there is a clear trend that reflect in which the more derivative is the genome, the lower the ENc value. However, for the remaining genes (*rps4, rpl4, rpl2, rps3, rpl6 rps5* and *rps2*), although they do not present a bias in ENc value, there is a decrease in these values is evidenced.

				Genome features		
NCBI code	**Phylum**	**Ancestral species of Haemosporida order**	**Strain**	**Organelle**	**Length (bp)**	**GC content**
AB002583	Rhodophyta	*Cyanidioschyzon merolae*	10D	Chloroplast	149987	37,6
AY673996	Rhodophyta	*Gracilaria tenuistipitata*	-	Chloroplast	183883	29,2
EF508371	Cryptophyta	*Rhodomonas salina*	CCMP1319	Chloroplast	135854	34,8
EF067920	Bacillariophyta	*Phaeodactylum tricornutum*	-	Chloroplast	117369	32,6
EF067921	Bacillariophyta	*Thalassiosira pseudonana*	-	Chloroplast	128814	30,7
JN039300	Miozoa	*Karlodinium veneficum*	-	Chloroplast	142981	24,1
KX897545	Cercozoa	*Paulinella micropora*	KR01	Chloroplast	976991	39,9
HM222968	Incertae sedis	*Alveolata sp.*	CCMP3155	Chloroplast	85535	47,7
MH557086	Apicomplexa	*Hepatozoon canis*	-	Apicoplast	31869	23,2
NC_001799	Apicomplexa	*Toxoplasma gondii*	RH	Apicoplast	34996	21,4
AAXT01000007	Apicomplexa	*Babesia bovis*	T2Bo	Apicoplast	35107	22
AAGK01000009	Apicomplexa	*Theileria parva*	Muguga	Apicoplast	39579	19,5
Table S1	Apicomplexa	Haemosporida order	-	Apicoplast	28968 to 34779	13,73

Table 7-5.: Chloroplast or Aplicoplast GC content through ancestral species to the order Haemosporida. From top to bottom, the species are organized according to the divergence time proposed by Janouškovec *et al.*, 2010

	Genes (>500bp)										
Species	rps4	rpl4	*rpl2*	*rps3*	*rpl6*	*rps5*	*tufA*	*clpC*	*rps2*	*RPOC1*	*RPOB*
Cyanidioschyzon merolae	48,3191	50,0353	45,6495	47,6462	50,416	48,3743	47,7442	47,859	50,9348	48,3591	44,766
Gracilaria tenuistipitata	47,557	47,9797	44,3309	47,2115	47,985	46,1588	43,1284	44,7342	47,8594	45,1139	43,4458
Rhodomonas salina	46,2685	49,9209	46,0045	45,3818	50,0855	49,6058	45,7361	45,2329	47,9491	52,0214	48,8231
Phaeodactylum tricornutum	50,6057	46,4934	44,201	46,3058	47,6894	47,4815	41,4216	44,8388	48,0567	44,2019	45,4294
Thalassiosira pseudonana	47,7901	44,224	41,637	44,9999	44,1494	42,8263	39,4702	40,6241	41,8293	44,3216	42,3084
Karlodinium veneficum	45,8456	45,8456	46,1231	46,5013	51,9584	43,7978	44,7591	49,4886	50,326	45,2145	50,7141
Paulinella micropora	53,3846	54,2886	50,5065	52,1942	50,9144	51,3592	52,1549	-	51,4888	52,8128	53,2355
Alveolata sp.	57,0505	54,5388	56,9442	54,7167	56,0774	58,5419	57,2272	58,693	56,5881	55,6155	57,7025
Hepatozoon canis	46,5145	45,7237	46,6423	42,4118	45,404	46,3080	42,7869	44,2088	45,9528	44,3638	42,2351
Toxoplasma gondii	41,3767	43,3228	39,8742	42,2604	44,1624	41,4843	37,834	36,7877	41,5902	-	-
Babesia bovis	-	-	46,5369	46,2384	-	-	43,1117	-	46,5659	42,5736	41,4177
Theileria parva	39,4682	40,1394	41,5886	40,2725	43,7889	41,3356	39,8953	-	41,3781	-	-
L. caulleryi	44,5586	46,0257	39,3818	42,5695	46,0577	42,6223	38,4079	38,6779	42,1754	38,0151	37,4530
H. columbae	45,9846	45,3562	40,7819	44,4057	43,5827	43,2728	39,5065	38,4586	41,5279	39,2162	36,1315
P. gallinaceum	45,7906	45,6638	41,5978	45,9676	44,3301	44,8189	37,7025	36,2636	42,8209	37,0921	36,2069
P. vivax	42,1970	45,6770	40,3981	45,3489	46,8748	44,8106	38,6866	37,3985	44,3275	38,4472	36,4854
P. falciparum	46,2940	46,3815	40,7994	44,9087	45,7180	42,9662	38,2454	38,0362	44,7080	39,8738	37,0674

Table 7-6.: ENc for each orthologue gene (>500bp) through ancestral species to the order Haemosporidae. From top to bottom, the species are organized according to the divergence time proposed by Janouškovec *et al.*, 2010.

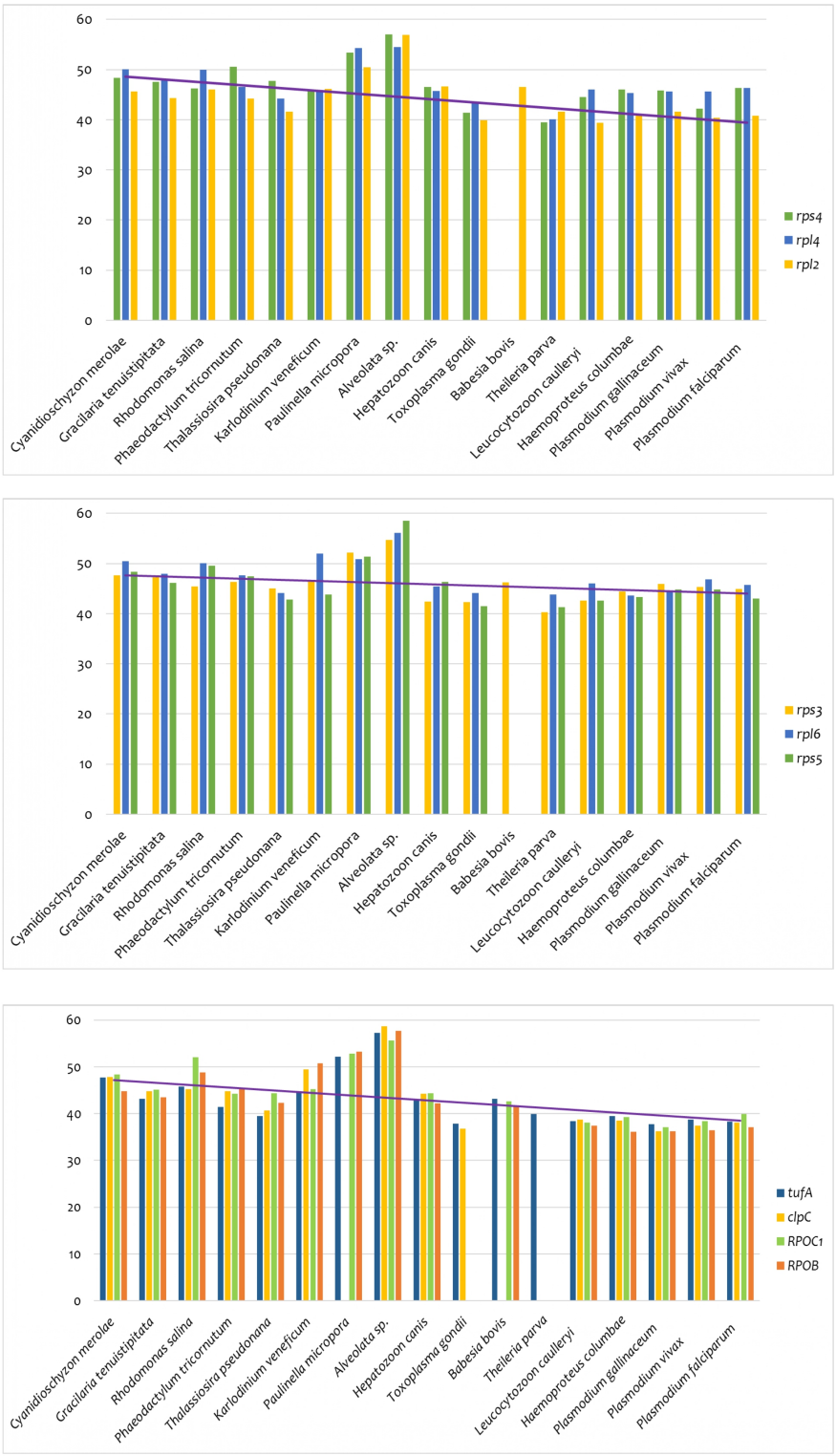


Figure 7-9.: Bar charts comparing Enc for each orthologue gene through ancestral species to the order Haemosporidae. From left to right, the species are organized according to the divergence time proposed by Janouškovec *et al.*, 2010. The line shows the trend of the data.

The previous results are correlated with genomic GC content, in which the tendency is that Apicomplexa phylum presents a drastic decrease in GC content (Table **7-5** and **7-10**). Therefore, the bias presented by ENc and RSCU values is understood in the context that the lower the content of GC, the greater the probability of increasing the frequency of codons rich in A-T.

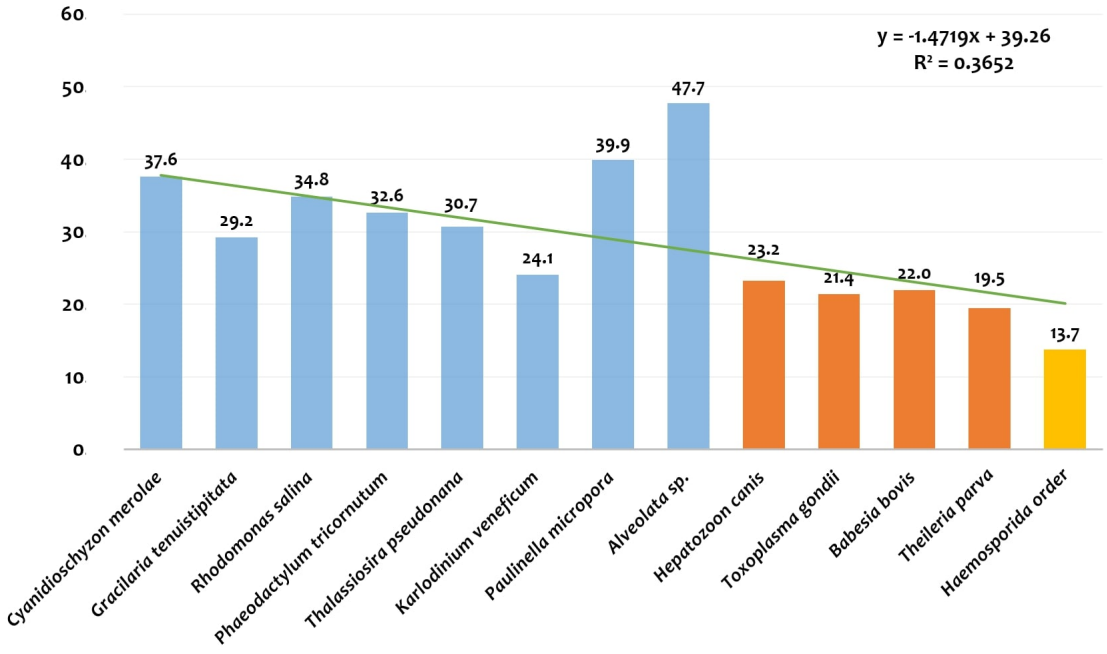


Figure 7-10.: Bar chart comparing Chloroplast or Aplicoplast GC content through ancestral species to the order Haemosporidae. From left to right, the species are organized according to the divergence time proposed by Janouškovec *et al.*, 2010. The line shows the trend of the data.

7.4.4. Design of *clpC* primers

The previous results indicated that *clpC* gene contains little saturation, has a weak codon usage bias and does not present a heterogeneous relative evolutionary rate, indicating that this gene is an excellent candidate to be considered in the phylogenetic hypotheses within order Haemosporida. Therefore, the following primers were designed for a nested PCR protocol (Table **7-7**), and they are capable of amplifying a sequence of approximately 1758 bp.

Code/PCR	Primers sequences (5'-3')	Size (bp)	GC (%)
NT-01F/outer	ATG ATC TTA TAT AAT ATW TAY WGT AC	26	15-19
NT-02R/outer	TCT TTT TAA WGG ACG WGC HCC	21	43-48
NT-03F/inner	TTA TGC CAA TTC ATT TAW TRT TAR G	25	20-28
NT-04R/inner	TTA GTT AAT CTA TTY AAT AAT TCW GG	26	19-32

Table 7-7.: primers designed to amplify the *clpC* gene

7.5. Discussion

In this study, the first ApiGenome of a species belonging to the genus *Haemoproteus*, subgenus *Haemoproteus* (*H. columbae*) is reported. The size, organization and structure of this ApiGenome is similar to the others reported in the order Haemosporida (Arisue *et al.*, 2012; Imura *et al.*, 2014; Arisue *et al.*, 2019); which it is interesting despite the differences in their vectors and hosts. Consequently, the metabolic pathways involved of this organelle should be also conserved, allowing the use of an experimental model of *Haemoproteus* infection (Cepeda *et al.*, 2019) in the future for the assessments of therapeutic targets present in Apicoplasts.

Within eukaryotic organisms, the Apicomplexa phylum is characterized by a nuclear genome with a GC content bias (with the exception of some species such as *P. vivax*; Weber, 1987; Gardner *et al.*, 2002; Hamilton *et al.*, 2016; Böhme *et al.*, 2018). Nevertheless, in the apicoplast genome, this bias is even more marked, and *H. columbae* ApiGenome shows the highest bias found until now in these parasites. We compared the GC content between chloroplast genome (the ancestral organelle of Apicoplast) and Apicoplast genome (**7-10**). Interestingly, the trend found was a continuous decrease in GC content through the ancestral clades (free-living organisms) to the derived clades (parasitic organisms), which consistently showed the highest bias. Similar results have been demonstrated the same tendency, eg. bacterias which harbor plasmids or nuclear genome with high AT richness, are more successful in infecting and surviving in their hosts than others (Rocha and Danchin, 2002; Dietel *et al.*, 2019).

Pacheco *et al.,* (2017) postulated a phylogenetic hypothesis for Haemosporida parasites using the entire mitochondrial genome, and the same topology was rescued using the complete ApiGenome by Bayesian inference (Fig. **7-3**A) and the single *clpC* gene (Fig. **7-4**B). On the other hand, the phylogenetic hypothesis inferred by ML using the whole genome (Fig. **7-3**B) and CDS's without saturation (Fig. **7-4**A) achieved to rescue the phylogeny proposed by Ariuse *et al.,* (2019). These results are in agreement with the main conclusion of Borner *et al.*, (2016), where topologies can vary based on the number of taxa analyzed and molecular marker used as observed for *P. vivax* and *P. cynomolgy* (Fig. **7-3** and Fig. **7-4**). However, the robustness of hypothesis presented in Fig. **7-3**A, showed that tRNA and rRNA genes are resolving the deep and terminal nodes of phylogeny. In addition, these genomes did not show an accelerated relative evolutionary rate, which are promising data for phylogenetic purposes. Therefore, it is necessary to expand the number of ApiGenomes of the order Haemosporida, in order to improve the phylogenetic relationship between the parasites and genera.

The AT richness in ApiGenome could explain the saturation of 12 ApiGenes and the problems around phylogenetic hypothesis (Dávalos and Perkins, 2008). This result has been found in other organisms with a bias in the GC genomic content (Yoder *et al.*, 1996; Blouin *et al.*, 1998; Breinholt and Kawahara, 2013). Nevertheless, for some ApiGenes, the third position presents a high satura-

tion compared to positions 1st and 2nd; therefore, whoever who uses those genes for phylogenetic hypotheses, the third position should be excluded (**7-2**). Due to the GC content bias, the result obtained for ApiGenome a similar result to Saul and Battisttutta (1988), Musto *et. al.*, (1999), and Peixoto, Fernández and Musto (2004), who used the nuclear genes of *P. falciparum*, in which the codon usage is strongly biased toward A and T at the third position. Therefore, it was found a decrease in the RSCU and ENc values in codons with predominantly GC content. In fact, the frequency of some codons was non-existent or very rare. This codon usage bias was found in previous studies in nuclear genes of *P. falciparum* (Saul and Battistutta, 1988).

Our evidence on ApiGenomes, together with previous evolutionary studies in nuclear and mitochondrial genomes (Musto *et al.*, 1999; Taylor *et al.*, 2013;Pacheco *et al.*, 2017), demonstrates that for Haemosporida parasites there is an integrated evolutionary dynamic (not independent) among the organelles of order. For, the evolutionary processes in each organelle are similar to each other.
Recently, Valkiūnas *et. al.*, (2019) highlights the importance of generating useful molecular markers from Apicoplast for taxonomic and phylogenetic purposes, and our evidence obtained through index of substitution saturation (**7-2**), phylogeny (**7-4**B), codon usage (Fig. **7-7**, Fig. **7-7**), and relative evolution rate (Fig. **7-6**), indicated that the *clpC* gene is a good molecular marker to be used in phylogenetic analyses. Despite the existence of primers that amplify a $641pb$ sequence of the *clpC* gene (Rathore *et al.*, 2001), we proposed primers capable of amplifying a sequence 3 times longer, therefore, more informative sites will be obtained in order to improve the phylogenetic signal from this gene.

7.6. Conclusions

We sequenced, annotated and characterized the first ApiGenome of Haemoproteus genus, which evidenced a structural and functional conservation of the Apigenomes at the order level. Likewise, it was useful to evaluate and analyze phylogenetic relationships, molecular evolutionary features within the Haemosporida order as well as at the ancestral level, allowing to provide important evidence for evolution, phylogenetic and medical purposes in the development of antimalarial drugs using the Apicoplast as target. Finally, we proposed, under the evidence obtained, new primers for the taxonomic and phylogenetic purposes of the *clpC* gene.

Acknowledgments

The authors would like to appreciate the Professor Andrés Pinzón for his valuable help by giving us access to the Instituto de Genética server, to carry out a large part of the bioinformatics analyzes, and Dr. Juan Pablo Isaza for his advice and recommendations at bioinformatic level.

References

Arisue, N. and Hashimoto, T. (2015). Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitology international*, 64(3):254–259.

Arisue, N., Hashimoto, T., Kawai, S., Honma, H., Kume, K., and Horii, T. (2019). Apicoplast phylogeny reveals the position of *Plasmodium vivax* basal to the asian primate malaria parasite clade. *Scientific reports*, 9(1):7274.

Arisue, N., Hashimoto, T., Mitsui, H., Palacpac, N. M., Kaneko, A., Kawai, S., Hasegawa, M., Tanabe, K., and Horii, T. (2012). The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Molecular biology and evolution*, 29(9):2095–2099.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., et al. (2008). Plasmodb: a functional genomic database for malaria parasites. *Nucleic acids research*, 37(suppl_1):D539–D543.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012). Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5):455–477.

Bioinformatics, B. (2017). Fastqc a quality control tool for high throughput sequence data 2016. *URL http://www. bioinformatics. babraham. ac. uk/projects/fastqc*.

Blouin, M. S., Yowell, C. A., Courtney, C. H., and Dame, J. B. (1998). Substitution bias, rapid saturation, and the use of mtdna for nematode systematics. *Molecular biology and evolution*, 15(12):1719–1727.

Böhme, U., Otto, T. D., Cotton, J. A., Steinbiss, S., Sanders, M., Oyola, S. O., Nicot, A., Gandon, S., Patra, K. P., Herd, C., et al. (2018). Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome research*, 28(4):547–560.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Borner, J., Pick, C., Thiede, J., Kolawole, O. M., Kingsley, M. T., Schulze, J., Cottontail, V. M., Wellinghausen, N., Schmidt-Chanasit, J., Bruchhaus, I., et al. (2016). Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. *Molecular phylogenetics and evolution*, 94:221–231.

Breinholt, J. W. and Kawahara, A. Y. (2013). Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome biology and evolution*, 5(11):2082–2092.

Cepeda, A. S., Lotta, I. A., Pinto Osorio, D. F., Macías Zapata, J., Valkiūnas, G., Barato, P., and Matta, N. E. (2019). The experimental characterization of complete life cycle of *Haemoproteus columbae*, with description of natural host-parasite system to study this infection. *International Journal for Parasitology*, submitted.

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6):e11147.

Dávalos, L. M. and Perkins, S. L. (2008). Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. *Genomics*, 91(5):433–442.

Dietel, A.-K., Merker, H., Kaltenpoth, M., and Kost, C. (2019). Selective advantages favour high genomic at-contents in intracellular elements. *PLoS genetics*, 15(4):e1007778.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906):498.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding dna sequences. *Molecular biology and evolution*, 11(5):725–736.

Gouy, M., Guindon, S., and Gascuel, O. (2009). Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2):221–224.

Hamilton, W. L., Claessens, A., Otto, T. D., Kekre, M., Fairhurst, R. M., Rayner, J. C., and Kwiatkowski, D. (2016). Extreme mutation bias and high at content in *Plasmodium falciparum*. *Nucleic acids research*, 45(4):1889–1901.

Hellgren, O., Waldenström, J., and Bensch, S. (2004). A new pcr assay for simultaneous studies of *LeucocytozoonI*, *Plasmodium*, and *Haemoproteus* from avian blood. *Journal of Parasitology*, 90(4):797–803.

Imura, T., Sato, S., Sato, Y., Sakamoto, D., Isobe, T., Murata, K., Holder, A. A., and Yukawa, M. (2014). The apicoplast genome of *Leucocytozoon caulleryi*, a pathogenic apicomplexan parasite of the chicken. *Parasitology research*, 113(3):823–828.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., and Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences*, 107(24):10949–10954.

Kumar, S., Stecher, G., and Tamura, K. (2016). Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7):1870–1874.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, 26(5):589–595.

Madeira, F., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A., Potter, S. C., Finn, R. D., Lopez, R., et al. (2019). The embl-ebi search and sequence analysis tools apis in 2019. *Nucleic acids research*.

Magoč, T. and Salzberg, S. L. (2011). Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957–2963.

Martinsen, E. S., Perkins, S. L., and Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Molecular phylogenetics and evolution*, 47(1):261–273.

McFadden, G. I. (2011). The apicoplast. *Protoplasma*, 248(4):641–650.

Mehlhorn, H. (2016). Encyclopedia of parasitology (2016 edition).

Musto, H., Romero, H., Zavala, A., Jabbari, K., and Bernardi, G. (1999). Synonymous codon choices in the extremely gc-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. *Journal of molecular evolution*, 49(1):27–35.

Okamoto, N., Spurck, T. P., Goodman, C. D., and McFadden, G. I. (2009). Apicoplast and mitochondrion in gametocytogenesis of plasmodium falciparum. *Eukaryotic cell*, 8(1):128–132.

Pacheco, M. A., Matta, N. E., Valkiūnas, G., Parker, P. G., Mello, B., Stanley Jr, C. E., Lentino, M., Garcia-Amado, M. A., Cranfield, M., Kosakovsky Pond, S. L., et al. (2017). Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. *Molecular biology and evolution*, 35(2):383–403.

Peixoto, L., Fernandez, V., and Musto, H. (2004). The effect of expression levels on codon usage in *Plasmodium falciparum*. *Parasitology*, 128(3):245–251.

Ralph, S. A., D'Ombrain, M. C., and McFadden, G. I. (2001). The apicoplast as an antimalarial drug target. *Drug Resistance Updates*, 4(3):145–151.

Rathore, D., Wahl, A. M., Sullivan, M., and McCutchan, T. F. (2001). A phylogenetic comparison of gene trees constructed from plastid, mitochondrial and genomic dna of plasmodium species. *Molecular and biochemical parasitology*, 114(1):89–94.

Rocha, E. P. and Danchin, A. (2002). Base composition bias might result from competition for metabolic resources. *TRENDS in Genetics*, 18(6):291–294.

Ronquist, F. and Huelsenbeck, J. P. (2003). Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574.

Saul, A. and Battistutta, D. (1988). Codon usage in *Plasmodium falciparum*. *Molecular and biochemical parasitology*, 27(1):35–42.

Shears, M. J., Botté, C. Y., and McFadden, G. I. (2015). Fatty acid metabolism in the plasmodium apicoplast: Drugs, doubts and knockouts. *Molecular and biochemical parasitology*, 199(1-2):34–50.

Sigala, P. A. and Goldberg, D. E. (2014). The peculiarities and paradoxes of plasmodium heme metabolism. *Annual review of microbiology*, 68:259–278.

Soldati, D. (1999). The apicoplast as a potential therapeutic target in *Toxoplasma* and other apicomplexan parasites. *Parasitology Today*, 15(1):5–7.

Srimath P., R., S Kusuma, S., Nammi, D., and RR Neelapu, N. (2017). Apicoplast import protein tic20 a promising therapeutic molecular target for plasmodium falciparum: An in silico approach for therapeutic intervention. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)*, 17(3):199–222.

Stothard, P. and Wishart, D. S. (2004). Circular genome visualization and exploration using cgview. *Bioinformatics*, 21(4):537–539.

Taylor, J. E., Pacheco, M. A., Bacon, D. J., Beg, M. A., Machado, R. L., Fairhurst, R. M., Herrera, S., Kim, J.-Y., Menard, D., Póvoa, M. M., et al. (2013). The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: parasite genetic diversity in the americas. *Molecular biology and evolution*, 30(9):2050–2064.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-iq-tree: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research*, 44(W1):W232–W235.

Valkiūnas, G. (2005). Avian malaria parasites and other haemosporidia CRC press. *Florida, Boca Raton*.

Valkiūnas, G., Ilgūnas, M., Bukauskaitė, D., Chagas, C. R. F., Bernotienė, R., Himmel, T., Harl, J., Weissenböck, H., and Iezhova, T. A. (2019). Molecular characterization of six widespread avian haemoproteids, with description of three new *Haemoproteus* species. *Acta tropica*, page 105051.

Weber, J. L. (1987). Analysis of sequences from the extremely a+ t-rich genome of *Plasmodium falciparum*. *Gene*, 52(1):103–109.

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, 87(1):23–29.

Xia, X. (2017). Dambe6: new tools for microbial genomics, phylogenetics, and molecular evolution. *Journal of Heredity*, 108(4):431–437.

Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y. (2003). An index of substitution saturation and its application. *Molecular phylogenetics and evolution*, 26(1):1–7.

Yeh, E. and DeRisi, J. L. (2011). Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage plasmodium falciparum. *PLoS biology*, 9(8):e1001138.

Yoder, A. D., Vilgalys, R., and Ruvolo, M. (1996). Molecular evolutionary dynamics of *cytochrome b* in strepsirrhine primates: The phylogenetic significance of third-position transversions. *Molecular Biology and Evolution*, 13(10):1339–1350.

Draft Genome Sequence of *Haemoproteus columbae* (lineage HAECOL1): first steps in a long way to complete a parasitic life cycle.

Axl S. Cepeda[a], Juan F. Alzate[b], and Nubia E. Matta[a]

a. Departamento de Biología, Grupo de Investigación Caracterización Genética e Inmunología, Sede Bogotá-Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia.

b. Centro Nacional de Secuenciación Genómica – CNSG, SIU, Grupo de Parasitología, Facultad de Medicina, Universidad de Antioquia, Medellín, Antioquia, Colombia.

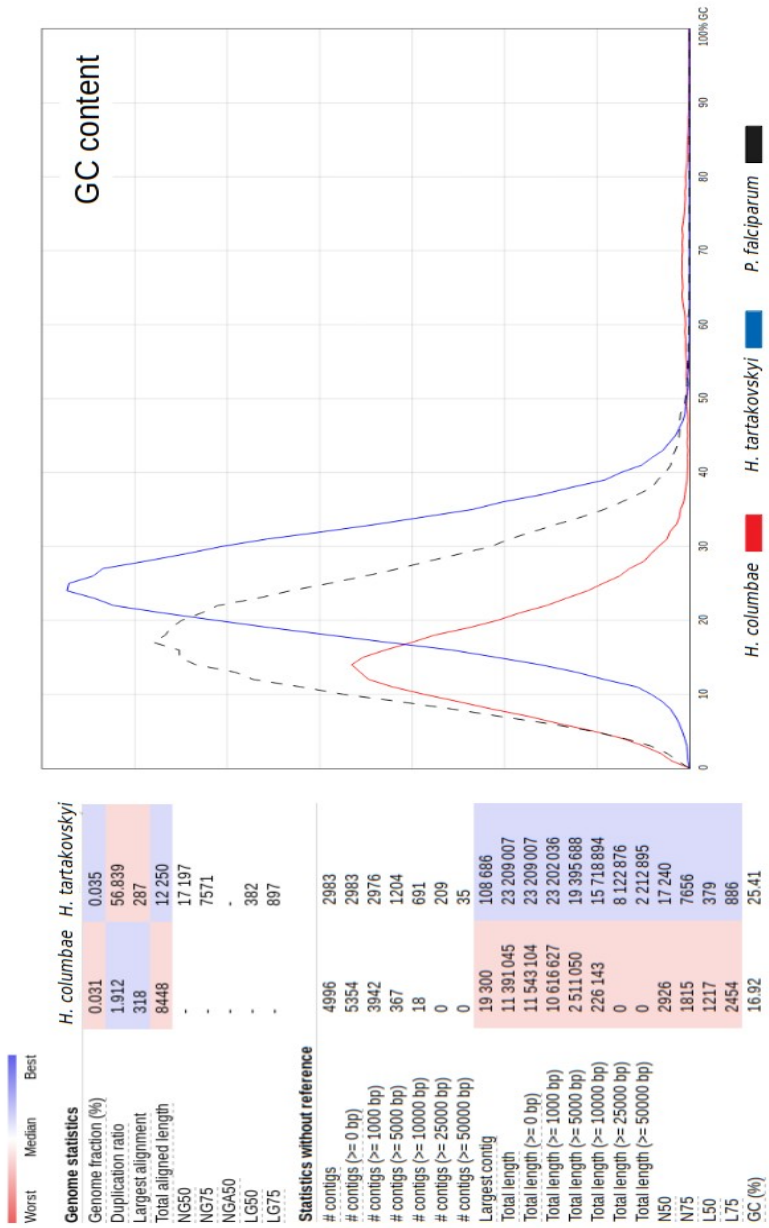
8.1. Abstract

Haemoproteus (*Haemoproteus*) *columbae*, lineage HAECOL1 was obtained from a feral Piegon Rock (*Columba livia*) naturally infected with a parasitaemia of 70.8 %. The sample was obtained from sexual stages present in peripheral blood. This is the first draft genome of a *Haemoproteus* subgenus. Considering the importance of Haemosporidian parasites, and the necessity of new information coming from different taxa, this information could be useful for the design of new molecular markers and detailed comparative genomic analyses.

8.2. Announcement

Haemoproteus (*Haemoproteus*) *columbae* is a widely distributed haemoparasite belonging to the phylum Apicomplexa. It is commonly found infecting pigeons mainly in the tropics (Valkiūnas, 2005). This genus is closely related to the *Plasmodium* genus, and they share several essential features in their life cycles; such as blood-sucking insects transmit them, the sexual phase occurs in the vector, develop meronts in tissues and liberate gamonts into the peripheral blood. However, there are some significant differences, such as *Haemoproteus* spp. infect only birds, the vectors are louse flies, instead of mosquitos as it occurs in *Plasmodium*, and in the peripheral blood are not present meronts, for that reason, it is very low probability to get an infection by injection of infected blood. These critical features shared between the two parasite genus, make that *Haemoproteus* parasites be appropriate in the study of evolution, metabolism, clinical aspects of infection and epizootiological topics The parasite was isolated from a feral Rock Pigeon captured in Bogotá. Total genomic DNA was subjected to Truseq library preparation and sequencing on Illumina HiSeqX 150-bp technology, resulting in a total of 628′859,636 pair-end reads. Quality trimming was performed using Trimmomatic software (Bolger *et al.*, 2014). These reads were aligned to the *Columba livia* reference genome (accession number GCA_001887795.1) using Burrows-Wheeler Aligner long-read alignment (BWA-mem; Li and Durbin, 2010). 11′515,634 unaligned reads were used for *de novo* assembly processes. FLASh software (Fast Length Adjustment of SHort reads; Magoč and Salzberg, 2011) was used to extend reads when possible. *De novo* assembly was done using St. Petersburg genome assembler (SPAdes; Nurk *et al.*, 2013) with default parameters with kmers of 33, 55, 77 and 99. Extended reads (with FLASh) as well as cleaned PE reads reads were used for this process. The final assembly for nuclear genome consisted of 5, 354 contigs. The average coverage was 2,3X, (minimum length, 251 bp; maximum, 19, 300 bp; N50, 2, 926; Fig. **8-1**). For mitochondrial (6, 087bp) and apicoplast (29, 798bp) genomes, a single contig was obtained for each one with a coverage of 68X and 15X respectively. Genome annotation was performed using Companion server (Steinbiss *et al.*, 2016).The final draft nuclear genome consists of 11,5 **Mb** with 16,92 % GC content (lowest content found in

Haemoporida order; this result was also consistent with AgiGenome; 2,982 putative coding and orthologous sequences with *Plasmodium falcipaum* (Table. **8-1**) and 994 singleton coding sequences (Table. **8-2**); for a total of 3,976 genes annotated in this genome; which indicates that approximately 65 to 75 % of the genes expected for Haemosporida order were found (Gardner *et al.*, 2002; Bensch *et al.*, 2016; Böhme *et al.*, 2018). In addition, sequences corresponding to 10 ltr-retrotransposons were identified by LTRharvest (Ellinghaus *et al.*, 2008) and LTRdigest (Steinbiss *et al.*, 2010) softwares (Table. **8-3**). The above is in agreement with the result found by Böhme *et al.,* 2018, in which only these kind of sequences have been found in Haemosporidians that infect birds.



	H. columbae	*H. tartakovskyi*
Genome statistics		
Genome fraction (%)	0.031	0.035
Duplication ratio	1.912	56.839
Largest alignment	318	287
Total aligned length	8448	12 250
NG50	-	17 197
NG75	-	7571
NGA50	-	-
LG50	-	382
LG75	-	897
Statistics without reference		
# contigs	4996	2983
# contigs (>= 0 bp)	5354	2983
# contigs (>= 1000 bp)	3942	2976
# contigs (>= 5000 bp)	367	1204
# contigs (>= 10000 bp)	18	691
# contigs (>= 25000 bp)	0	209
# contigs (>= 50000 bp)	0	35
Largest contig	19 300	108 686
Total length	11 391 045	23 209 007
Total length (>= 0 bp)	11 543 104	23 209 007
Total length (>= 1000 bp)	10 616 627	23 202 036
Total length (>= 5000 bp)	2 511 050	19 395 688
Total length (>= 10000 bp)	226 143	15 718 894
Total length (>= 25000 bp)	0	8 122 876
Total length (>= 50000 bp)	0	2 212 895
N50	2926	17 240
N75	1815	7656
L50	1217	379
L75	2454	886
GC (%)	16.92	25.41

Figure 8-1.: Genomic assembly statistics compared with *Haemorpoteus* (*Parahaemoproteus*) *tartakovskyi* and *P. falciparum* as reference. All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g.,"# contigs (>= 0 bp).[a]nd "Total length (>= 0 bp)ïnclude all contigs)

Table 8-1.: Genes summary of *Haemoproteus columbae*. Orthology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000000100	aspartate–tRNA ligase	PF3D7_0102900
Hcol_000000200	vacuolar protein sorting-associated protein 51, putative	PF3D7_0103100
Hcol_000000300	vacuolar protein sorting-associated protein 51, putative	PF3D7_0103100
Hcol_000000400	zinc-carboxypeptidase, putative	PF3D7_0103400
Hcol_000000500	conserved *Plasmodium* protein, unknown function	PF3D7_0103500
Hcol_000000600	actin-related protein	PF3D7_0103800

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000000700	parasite-infected erythrocyte surface protein	PF3D7_0103900
Hcol_000000800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0104100
Hcol_000000900	ubiquitin carboxyl-terminal hydrolase 1, putative	PF3D7_0104300
Hcol_000001000	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	PF3D7_0104400
Hcol_000001100	conserved protein, unknown function	PF3D7_0104500
Hcol_000001200	conserved *Plasmodium* protein, unknown function	PF3D7_0104600
Hcol_000001300	conserved *Plasmodium* protein, unknown function	PF3D7_0105100
Hcol_000001400	RAP protein, putative	PF3D7_0105200
Hcol_000001500	cyclase-associated protein	PF3D7_0105300
Hcol_000001600	conserved *Plasmodium* protein, unknown function	PF3D7_0105500
Hcol_000001700	conserved *Plasmodium* protein, unknown function	PF3D7_0105600
Hcol_000001800	conserved *Plasmodium* protein, unknown function	PF3D7_0105800
Hcol_000001900	conserved *Plasmodium* protein, unknown function	PF3D7_0105800
Hcol_000002000	DNA binding protein, putative	PF3D7_0105900
Hcol_000002100	conserved *Plasmodium* protein, unknown function	PF3D7_0106000
Hcol_000002200	V-type proton ATPase subunit C, putative	PF3D7_0106100
Hcol_000002300	pre-rRNA-processing protein TSR2, putative	PF3D7_0106400
Hcol_000002400	conserved *Plasmodium* protein, unknown function	PF3D7_0106500
Hcol_000002500	small ribosomal subunit assembling AARP2 protein	PF3D7_0106700
Hcol_000002600	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, putative	PF3D7_0106900
Hcol_000002700	conserved *Plasmodium* protein, unknown function	PF3D7_0107400
Hcol_000002800	serine/threonine protein kinase, putative	PF3D7_0107600
Hcol_000002900	double-strand break repair protein MRE11	PF3D7_0107800
Hcol_000003000	conserved *Plasmodium* protein, unknown function	PF3D7_0107900
Hcol_000003100	proteasome subunit beta type-3, putative	PF3D7_0108000
Hcol_000003200	conserved *Plasmodium* protein, unknown function	PF3D7_0108600
Hcol_000003300	secreted ookinete protein, putative	PF3D7_0108700
Hcol_000003400	conserved *Plasmodium* protein, unknown function	PF3D7_0108800
Hcol_000003500	photosensitized INA-labeled protein PHIL1, putative	PF3D7_0109000
Hcol_000003600	mRNA cleavage factor-like protein, putative	PF3D7_0109200
Hcol_000003700	fatty acid elongation protein, GNS1/SUR4 family, putative	PF3D7_0109300
Hcol_000003800	tubulin-specific chaperone a, putative	PF3D7_0109400
Hcol_000003900	N-acetyltransferase, putative	PF3D7_0109500
Hcol_000004000	cold-shock protein, putative	PF3D7_0109600
Hcol_000004100	conserved *Plasmodium* protein, unknown function	PF3D7_0110000
Hcol_000004200	selenocysteine-specific elongation factor selB homologue, putative	PF3D7_0110100
Hcol_000004300	FAD-linked sulfhydryl oxidase ERV1, putative	PF3D7_0110200
Hcol_000004400	bromodomain protein, putative	PF3D7_0110500
Hcol_000004500	phosphatidylinositol-4-phosphate 5-kinase	PF3D7_0110600
Hcol_000004600	chromatin assembly factor 1 protein WD40 domain, putative	PF3D7_0110700
Hcol_000004700	adenylate kinase-like protein 1	PF3D7_0110900
Hcol_000004800	replication factor c protein, putative	PF3D7_0111300
Hcol_000004900	replication factor c protein, putative	PF3D7_0111300
Hcol_000005000	conserved *Plasmodium* protein, unknown function	PF3D7_0111400
Hcol_000005100	UMP-CMP kinase, putative	PF3D7_0111500
Hcol_000005200	conserved protein, unknown function	PF3D7_0111900
Hcol_000005300	octaprenyl pyrophosphate synthase	PF3D7_0202700
Hcol_000005400	repetitive organellar protein, putative	PF3D7_0203000
Hcol_000005500	conserved *Plasmodium* protein, unknown function	PF3D7_0203200
Hcol_000005600	ERCC1 nucleotide excision repair protein, putative	PF3D7_0203300
Hcol_000005700	conserved *Plasmodium* protein, unknown function	PF3D7_0203400
Hcol_000005800	conserved *Plasmodium* protein, unknown function	PF3D7_0203800
Hcol_000005900	5'-3' exonuclease, N-terminal resolvase-like domain, putative	PF3D7_0203900
Hcol_000006000	conserved *Plasmodium* protein, unknown function	PF3D7_0204100
Hcol_000006100	conserved *Plasmodium* protein, unknown function	PF3D7_0204100
Hcol_000006200	conserved *Plasmodium* protein, unknown function	PF3D7_0204200
Hcol_000006300	aspartate transaminase	PF3D7_0204500
Hcol_000006400	3'-5' exonuclease, putative	PF3D7_0204800
Hcol_000006500	ubiE/COQ5 methyltransferase, putative	PF3D7_0204900
Hcol_000006600	PCI domain-containing protein, putative	PF3D7_0205400
Hcol_000006700	DNA-directed RNA polymerase II 16 kDa subunit, putative	PF3D7_0205500
Hcol_000006800	conserved *Plasmodium* protein, unknown function	PF3D7_0205600
Hcol_000006900	conserved *Plasmodium* protein, unknown function	PF3D7_0205700
Hcol_000007000	conserved *Plasmodium* protein, unknown function	PF3D7_0205800
Hcol_000007100	26S proteasome regulatory subunit RPN1, putative	PF3D7_0205900
Hcol_000007200	DNA repair protein RAD2, putative	PF3D7_0206000
Hcol_000007300	pantothenate transporter	PF3D7_0206200
Hcol_000007400	conserved *Plasmodium* protein, unknown function	PF3D7_0206500
Hcol_000007500	DNA-directed RNA polymerase III subunit RPC10, putative	PF3D7_0206600
Hcol_000007600	serine repeat antigen 8	PF3D7_0207300
Hcol_000007700	serine repeat antigen 7	PF3D7_0207400
Hcol_000007800	conserved *Plasmodium* protein, unknown function	PF3D7_0208100

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000007900	conserved *Plasmodium* protein, unknown function	PF3D7_0208300
Hcol_000008000	ribosome-recycling factor	PF3D7_0208600
Hcol_000008100	conserved *Plasmodium* protein, unknown function	PF3D7_0208700
Hcol_000008200	conserved *Plasmodium* protein, unknown function	PF3D7_0208700
Hcol_000008300	conserved *Plasmodium* protein, unknown function	PF3D7_0208800
Hcol_000008400	phospholipase A2, putative	PF3D7_0209100
Hcol_000008500	phospholipase A2, putative	PF3D7_0209100
Hcol_000008600	conserved protein, unknown function	PF3D7_0209400
Hcol_000008700	ATP-dependent RNA helicase UAP56	PF3D7_0209800
Hcol_000008800	secretory complex protein 61 gamma subunit	PF3D7_0210000
Hcol_000008900	60S ribosomal protein L37ae, putative	PF3D7_0210100
Hcol_000009000	conserved *Plasmodium* protein, unknown function	PF3D7_0210200
Hcol_000009100	monocarboxylate transporter, putative	PF3D7_0210300
Hcol_000009200	syntaxin, Qa-SNARE family	PF3D7_0210700
Hcol_000009300	ubiquinol-cytochrome-c reductase complex assembly factor 1, putative	PF3D7_0211300
Hcol_000009400	beta-ketoacyl-ACP synthase III	PF3D7_0211400
Hcol_000009500	tyrosine kinase-like protein, putative	PF3D7_0211700
Hcol_000009600	asparagine–tRNA ligase	PF3D7_0211800
Hcol_000009700	mitochondrial ribosomal protein L12 precursor, putative	PF3D7_0212200
Hcol_000009800	peptide chain release factor subunit 1, putative	PF3D7_0212300
Hcol_000009900	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0212400
Hcol_000010000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0212400
Hcol_000010100	conserved *Plasmodium* protein, unknown function	PF3D7_0212500
Hcol_000010200	conserved *Plasmodium* protein, unknown function	PF3D7_0212500
Hcol_000010300	secreted protein with altered thrombospondin repeat domain	PF3D7_0212600
Hcol_000010400	Leu/Phe-tRNA protein transferase, putative	PF3D7_0212900
Hcol_000010500	conserved *Plasmodium* protein, unknown function	PF3D7_0213000
Hcol_000010600	protein SIS1	PF3D7_0213100
Hcol_000010700	conserved *Plasmodium* protein, unknown function	PF3D7_0213300
Hcol_000010800	protein kinase 7	PF3D7_0213400
Hcol_000010900	conserved *Plasmodium* protein, unknown function	PF3D7_0213600
Hcol_000011000	conserved *Plasmodium* protein, unknown function	PF3D7_0213600
Hcol_000011100	conserved *Plasmodium* protein, unknown function	PF3D7_0213800
Hcol_000011200	T-complex protein 1 subunit theta	PF3D7_0214000
Hcol_000011300	protein transport protein SEC31	PF3D7_0214100
Hcol_000011400	conserved *Plasmodium* protein, unknown function	PF3D7_0214400
Hcol_000011500	conserved *Plasmodium* protein, unknown function	PF3D7_0214500
Hcol_000011600	serine/threonine protein kinase, putative	PF3D7_0214600
Hcol_000011700	conserved *Plasmodium* protein, unknown function	PF3D7_0214700
Hcol_000011800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0214800
Hcol_000011900	RING zinc finger protein, putative	PF3D7_0215100
Hcol_000012000	conserved *Plasmodium* protein, unknown function	PF3D7_0215200
Hcol_000012100	conserved *Plasmodium* protein, unknown function	PF3D7_0215500
Hcol_000012200	DNA-directed RNA polymerase II subunit RPB2, putative	PF3D7_0215700
Hcol_000012300	DNA-directed RNA polymerase II subunit RPB2, putative	PF3D7_0215700
Hcol_000012400	conserved *Plasmodium* protein, unknown function	PF3D7_0216100
Hcol_000012500	conserved *Plasmodium* protein, unknown function	PF3D7_0216200
Hcol_000012600	vacuolar protein sorting-associated protein 45, putative	PF3D7_0216400
Hcol_000012700	MtN3-like protein	PF3D7_0216600
Hcol_000012800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0217000
Hcol_000012900	ATP synthase F1, alpha subunit	PF3D7_0217100
Hcol_000013000	conserved *Plasmodium* protein, unknown function	PF3D7_0217200
Hcol_000013100	conserved *Plasmodium* protein, unknown function	PF3D7_0217400
Hcol_000013200	calcium-dependent protein kinase 1	PF3D7_0217500
Hcol_000013300	conserved *Plasmodium* protein, unknown function	PF3D7_0217600
Hcol_000013400	40S ribosomal protein S26	PF3D7_0217800
Hcol_000013500	replication factor C subunit 2, putative	PF3D7_0218000
Hcol_000013600	ATP-dependent rRNA helicase RRP3, putative	PF3D7_0218400
Hcol_000013700	pre-mRNA-processing protein 45, putative	PF3D7_0218700
Hcol_000013800	conserved *Plasmodium* protein, unknown function	PF3D7_0301900
Hcol_000013900	serine/threonine protein kinase	PF3D7_0302100
Hcol_000014000	CDGSH iron-sulfur domain-containing protein, putative	PF3D7_0302700
Hcol_000014100	exportin-1, putative	PF3D7_0302900
Hcol_000014200	N-ethylmaleimide-sensitive fusion protein	PF3D7_0303000
Hcol_000014300	conserved *Plasmodium* protein, unknown function	PF3D7_0303100
Hcol_000014400	HAD superfamily protein, putative	PF3D7_0303200
Hcol_000014500	DNA-directed RNA polymerases I, II, and III subunit RPABC2, putative	PF3D7_0303300
Hcol_000014600	palmitoyltransferase	PF3D7_0303400
Hcol_000014700	spindle pole body protein, putative	PF3D7_0303500
Hcol_000014800	IBR domain protein, putative	PF3D7_0303800
Hcol_000014900	inner membrane complex protein 1e, putative	PF3D7_0304100
Hcol_000015000	EH domain-containing protein	PF3D7_0304200

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000015100	conserved *Plasmodium* protein, unknown function	PF3D7_0304300
Hcol_000015200	60S ribosomal protein L44	PF3D7_0304400
Hcol_000015300	conserved *Plasmodium* protein, unknown function	PF3D7_0304900
Hcol_000015400	elongation factor Ts	PF3D7_0305000
Hcol_000015500	conserved *Plasmodium* protein, unknown function	PF3D7_0305100
Hcol_000015600	conserved *Plasmodium* protein, unknown function	PF3D7_0305200
Hcol_000015700	conserved *Plasmodium* protein, unknown function	PF3D7_0305200
Hcol_000015800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0305300
Hcol_000015900	AP endonuclease (DNA-[apurinic or apyrimidinic site] lyase), putative	PF3D7_0305600
Hcol_000016000	P-loop containing nucleoside triphosphate hydrolase, putative	PF3D7_0305800
Hcol_000016100	conserved *Plasmodium* protein, unknown function	PF3D7_0306100
Hcol_000016200	activator of Hsp90 ATPase	PF3D7_0306200
Hcol_000016300	glutaredoxin 1	PF3D7_0306300
Hcol_000016400	FAD-dependent glycerol-3-phosphate dehydrogenase, putative	PF3D7_0306400
Hcol_000016500	conserved *Plasmodium* protein, unknown function	PF3D7_0306500
Hcol_000016600	conserved *Plasmodium* protein, unknown function	PF3D7_0306600
Hcol_000016700	T-complex protein 1 subunit beta	PF3D7_0306800
Hcol_000016800	60S ribosomal protein L7, putative	PF3D7_0307200
Hcol_000016900	EB1 homolog, putative	PF3D7_0307300
Hcol_000017000	ATP-dependent Clp protease proteolytic subunit	PF3D7_0307400
Hcol_000017100	conserved *Plasmodium* protein, unknown function	PF3D7_0307800
Hcol_000017200	conserved *Plasmodium* protein, unknown function	PF3D7_0307900
Hcol_000017300	DNA polymerase delta small subunit, putative	PF3D7_0308000
Hcol_000017400	T-complex protein 1 subunit eta	PF3D7_0308200
Hcol_000017500	pre-mRNA-processing factor 19, putative	PF3D7_0308600
Hcol_000017600	conserved *Plasmodium* protein, unknown function	PF3D7_0308700
Hcol_000017700	splicing factor 3B subunit 1, putative	PF3D7_0308900
Hcol_000017800	dual specificity protein phosphatase	PF3D7_0309000
Hcol_000017900	conserved *Plasmodium* protein, unknown function	PF3D7_0309100
Hcol_000018000	serine/threonine protein kinase, putative	PF3D7_0309200
Hcol_000018100	N2227-like protein, putative	PF3D7_0309300
Hcol_000018200	N2227-like protein, putative	PF3D7_0309300
Hcol_000018300	asparagine synthetase, putative	PF3D7_0309500
Hcol_000018400	SECIS-binding protein 2, putative	PF3D7_0309700
Hcol_000018500	YTH domain-containing protein, putative	PF3D7_0309800
Hcol_000018600	conserved *Plasmodium* protein, unknown function	PF3D7_0309900
Hcol_000018700	calcium-dependent protein kinase 3	PF3D7_0310100
Hcol_000018800	phd finger protein, putative	PF3D7_0310200
Hcol_000018900	phosphoglycerate mutase, putative	PF3D7_0310300
Hcol_000019000	parasite-infected erythrocyte surface protein	PF3D7_0310400
Hcol_000019100	eukaryotic translation initiation factor 3 subunit K, putative	PF3D7_0310600
Hcol_000019200	trafficking protein particle complex subunit 4, putative	PF3D7_0310700
Hcol_000019300	conserved *Plasmodium* protein, unknown function	PF3D7_0310800
Hcol_000019400	valine–tRNA ligase, putative	PF3D7_0311200
Hcol_000019500	valine–tRNA ligase, putative	PF3D7_0311200
Hcol_000019600	phosphatidylinositol 3- and 4-kinase, putative	PF3D7_0311300
Hcol_000019700	conserved protein, unknown function	PF3D7_0311800
Hcol_000019800	conserved *Plasmodium* protein, unknown function	PF3D7_0311900
Hcol_000019900	E3 ubiquitin-protein ligase, putative	PF3D7_0312100
Hcol_000020000	TPR domain containing protein	PF3D7_0312200
Hcol_000020100	26S proteasome regulatory subunit RPN12, putative	PF3D7_0312300
Hcol_000020200	glycogen synthase kinase 3	PF3D7_0312400
Hcol_000020300	60S ribosomal protein L26, putative	PF3D7_0312800
Hcol_000020400	ubiquitin-protein ligase, putative	PF3D7_0313100
Hcol_000020500	ubiquitin-protein ligase, putative	PF3D7_0313100
Hcol_000020600	conserved *Plasmodium* protein, unknown function	PF3D7_0313200
Hcol_000020700	conserved *Plasmodium* protein, unknown function	PF3D7_0313400
Hcol_000020800	GTP-binding protein EngA, putative	PF3D7_0313500
Hcol_000020900	conserved *Plasmodium* protein, unknown function	PF3D7_0313700
Hcol_000021000	conserved *Plasmodium* protein, unknown function	PF3D7_0313900
Hcol_000021100	vesicle transport v-SNARE protein, putative	PF3D7_0314100
Hcol_000021200	conserved *Plasmodium* protein, unknown function	PF3D7_0314200
Hcol_000021300	DER1-like protein	PF3D7_0314300
Hcol_000021400	conserved *Plasmodium* protein, unknown function	PF3D7_0314500
Hcol_000021500	conserved *Plasmodium* protein, unknown function	PF3D7_0314600
Hcol_000021600	conserved *Plasmodium* protein, unknown function	PF3D7_0314800
Hcol_000021700	circumsporozoite- and TRAP-related protein	PF3D7_0315200
Hcol_000021800	conserved *Plasmodium* protein, unknown function	PF3D7_0315300
Hcol_000021900	mitochondrial ribosomal protein L29/L47 precursor, putative	PF3D7_0315500
Hcol_000022000	conserved *Plasmodium* protein, unknown function	PF3D7_0315600
Hcol_000022100	zinc finger protein, putative	PF3D7_0315800
Hcol_000022200	microneme associated antigen	PF3D7_0316000

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000022300	inorganic pyrophosphatase, putative	PF3D7_0316300
Hcol_000022400	conserved *Plasmodium* protein, unknown function	PF3D7_0316400
Hcol_000022500	kinetochore protein NUF2, putative	PF3D7_0316500
Hcol_000022600	formate-nitrite transporter	PF3D7_0316600
Hcol_000022700	HVA22/TB2/DP1 family protein, putative	PF3D7_0316700
Hcol_000022800	E3 ubiquitin-protein ligase, putative	PF3D7_0316900
Hcol_000022900	proteasome subunit alpha type-3, putative	PF3D7_0317000
Hcol_000023000	cdc2-related protein kinase 4	PF3D7_0317200
Hcol_000023100	conserved *Plasmodium* protein, unknown function	PF3D7_0317300
Hcol_000023200	conserved *Plasmodium* protein, unknown function	PF3D7_0317400
Hcol_000023300	kinesin-5	PF3D7_0317500
Hcol_000023400	40S ribosomal protein S11, putative	PF3D7_0317600
Hcol_000023500	CPSF (cleavage and polyadenylation specific factor), subunit A, putative	PF3D7_0317700
Hcol_000023600	proteasome regulatory protein, putative	PF3D7_0317800
Hcol_000023700	conserved *Plasmodium* protein, unknown function	PF3D7_0318000
Hcol_000023800	stomatin-like protein	PF3D7_0318100
Hcol_000023900	DNA-directed RNA polymerase II subunit RPB1	PF3D7_0318200
Hcol_000024000	conserved *Plasmodium* protein, unknown function	PF3D7_0318500
Hcol_000024100	cleavage and polyadenylation specificity factor, putative	PF3D7_0318600
Hcol_000024200	trophozoite stage antigen	PF3D7_0318700
Hcol_000024300	conserved protein, unknown function	PF3D7_0318900
Hcol_000024400	P-type ATPase, putative	PF3D7_0319000
Hcol_000024500	P-type ATPase, putative	PF3D7_0319000
Hcol_000024600	endonuclease/exonuclease/phosphatase family protein, putative	PF3D7_0319200
Hcol_000024700	kinesin-8, putative	PF3D7_0319400
Hcol_000024800	RNA-binding protein, putative	PF3D7_0319500
Hcol_000024900	elongation factor 1 (EF-1), putative	PF3D7_0319600
Hcol_000025000	conserved *Plasmodium* protein, unknown function	PF3D7_0319800
Hcol_000025100	T-complex protein 1 subunit epsilon	PF3D7_0320300
Hcol_000025200	nicotinamidase, putative	PF3D7_0320500
Hcol_000025300	ATP-dependent RNA helicase DDX6	PF3D7_0320800
Hcol_000025400	histone H2A variant, putative	PF3D7_0320900
Hcol_000025500	conserved *Plasmodium* protein, unknown function	PF3D7_0321000
Hcol_000025600	UDP-N-acetylglucosamine–dolichyl-phosphate N-acetylglucosaminephosphotransferase, putative	PF3D7_0321200
Hcol_000025700	protein kinase, putative	PF3D7_0321400
Hcol_000025800	peptidase, putative	PF3D7_0321500
Hcol_000025900	microtubule and actin binding protein, putative	PF3D7_0321700
Hcol_000026000	cyclic amine resistance locus protein	PF3D7_0321900
Hcol_000026100	peptidyl-prolyl cis-trans isomerase	PF3D7_0322000
Hcol_000026200	diacylglycerol O-acyltransferase	PF3D7_0322300
Hcol_000026300	regulator of initiation factor 2 (eIF2)	PF3D7_0322400
Hcol_000026400	conserved *Plasmodium* protein, unknown function	PF3D7_0322600
Hcol_000026500	conserved *Plasmodium* protein, unknown function	PF3D7_0322700
Hcol_000026600	conserved *Plasmodium* protein, unknown function	PF3D7_0322800
Hcol_000026700	40S ribosomal protein S3A, putative	PF3D7_0322900
Hcol_000026800	conserved *Plasmodium* protein, unknown function	PF3D7_0323200
Hcol_000026900	Rh5 interacting protein	PF3D7_0323400
Hcol_000027000	survival motor neuron-like protein	PF3D7_0323500
Hcol_000027100	conserved *Plasmodium* protein, unknown function	PF3D7_0323600
Hcol_000027200	U4/U6.U5 tri-snRNP-associated protein 1, putative	PF3D7_0323700
Hcol_000027300	U4/U6.U5 tri-snRNP-associated protein 1, putative	PF3D7_0323700
Hcol_000027400	conserved *Plasmodium* protein, unknown function	PF3D7_0323800
Hcol_000027500	conserved *Plasmodium* protein, unknown function	PF3D7_0323800
Hcol_000027600	conserved *Plasmodium* protein, unknown function	PF3D7_0403400
Hcol_000027700	alpha/beta hydrolase, putative	PF3D7_0403800
Hcol_000027800	SET domain protein, putative	PF3D7_0403900
Hcol_000027900	transcription factor with AP2 domain(s), putative	PF3D7_0404100
Hcol_000028000	conserved *Plasmodium* protein, unknown function	PF3D7_0404200
Hcol_000028100	6-cysteine protein	PF3D7_0404500
Hcol_000028200	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0404600
Hcol_000028300	ATP-dependent RNA helicase, putative	PF3D7_0405000
Hcol_000028400	protein transport protein Sec24B	PF3D7_0405100
Hcol_000028500	ag-1 blood stage membrane protein homologue	PF3D7_0405200
Hcol_000028600	6-cysteine protein	PF3D7_0405300
Hcol_000028700	lysine decarboxylase, putative	PF3D7_0405700
Hcol_000028800	conserved *Plasmodium* protein, unknown function	PF3D7_0405800
Hcol_000028900	conserved *Plasmodium* protein, unknown function	PF3D7_0406000
Hcol_000029000	conserved *Plasmodium* protein, unknown function	PF3D7_0406500
Hcol_000029100	conserved *Plasmodium* protein, unknown function	PF3D7_0406600
Hcol_000029200	conserved *Plasmodium* protein, unknown function	PF3D7_0406800
Hcol_000029300	conserved *Plasmodium* protein, unknown function	PF3D7_0406900

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000029400	conserved *Plasmodium* protein, unknown function	PF3D7_0407400
Hcol_000029500	conserved *Plasmodium* protein, unknown function	PF3D7_0407700
Hcol_000029600	conserved *Plasmodium* protein, unknown function	PF3D7_0407800
Hcol_000029700	conserved *Plasmodium* protein, unknown function	PF3D7_0408100
Hcol_000029800	zinc finger, RAN binding protein, putative	PF3D7_0408300
Hcol_000029900	conserved *Plasmodium* protein, unknown function	PF3D7_0408400
Hcol_000030000	flap endonuclease 1	PF3D7_0408500
Hcol_000030100	sporozoite micronemal protein essential for cell traversal	PF3D7_0408700
Hcol_000030200	tRNA N6-adenosine threonylcarbamoyltransferase	PF3D7_0408900
Hcol_000030300	conserved *Plasmodium* protein, unknown function	PF3D7_0409300
Hcol_000030400	heat shock protein 40	PF3D7_0409400
Hcol_000030500	replication protein A1, large subunit	PF3D7_0409600
Hcol_000030600	actin-like protein, putative	PF3D7_0409900
Hcol_000030700	conserved *Plasmodium* protein, unknown function	PF3D7_0410600
Hcol_000030800	conserved *Plasmodium* protein, unknown function	PF3D7_0410800
Hcol_000030900	conserved *Plasmodium* protein, unknown function	PF3D7_0411000
Hcol_000031000	conserved *Plasmodium* protein, unknown function	PF3D7_0411100
Hcol_000031100	conserved *Plasmodium* protein, unknown function	PF3D7_0411300
Hcol_000031200	DEAD box ATP-dependent RNA helicase, putative	PF3D7_0411400
Hcol_000031300	conserved *Plasmodium* protein, unknown function	PF3D7_0411800
Hcol_000031400	DNA polymerase alpha catalytic subunit A	PF3D7_0411900
Hcol_000031500	conserved *Plasmodium* protein, unknown function	PF3D7_0412200
Hcol_000031600	26S protease regulatory subunit 6B, putative	PF3D7_0413600
Hcol_000031700	50S ribosomal protein L10, putative	PF3D7_0413800
Hcol_000031800	ubiquitin carboxyl-terminal hydrolase 13, putative	PF3D7_0413900
Hcol_000031900	structural maintenance of chromosomes protein 3, putative	PF3D7_0414000
Hcol_000032000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0414100
Hcol_000032100	calmodulin-like protein	PF3D7_0414200
Hcol_000032200	conserved protein, unknown function	PF3D7_0414600
Hcol_000032300	GTP-binding protein, putative	PF3D7_0414700
Hcol_000032400	arsenical pump-driving ATPase, putative	PF3D7_0415000
Hcol_000032500	conserved *Plasmodium* protein, unknown function	PF3D7_0415200
Hcol_000032600	cdc2-related protein kinase 3	PF3D7_0415300
Hcol_000032700	coatomer subunit zeta, putative	PF3D7_0415400
Hcol_000032800	nuclear cap-binding protein, putative	PF3D7_0415500
Hcol_000032900	GTP:AMP phosphotransferase	PF3D7_0415600
Hcol_000033000	conserved *Plasmodium* protein, unknown function	PF3D7_0415700
Hcol_000033100	RING zinc finger protein, putative	PF3D7_0415800
Hcol_000033200	60S ribosomal protein L15, putative	PF3D7_0415900
Hcol_000033300	glutamyl-tRNA(Gln) amidotransferase subunit A	PF3D7_0416100
Hcol_000033400	DNA helicase MCM9, putative	PF3D7_0416300
Hcol_000033500	repressor of RNA polymerase III transcription MAF1, putative	PF3D7_0416500
Hcol_000033600	prohibitin-like protein, putative	PF3D7_0416600
Hcol_000033700	CDGSH iron-sulfur domain-containing protein, putative	PF3D7_0416700
Hcol_000033800	small GTP-binding protein sar1	PF3D7_0416800
Hcol_000033900	conserved *Plasmodium* protein, unknown function	PF3D7_0416900
Hcol_000034000	mRNA-binding protein PUF2	PF3D7_0417100
Hcol_000034100	bifunctional dihydrofolate reductase-thymidylate synthase	PF3D7_0417200
Hcol_000034200	LETM1-like protein, putative	PF3D7_0417300
Hcol_000034300	conserved *Plasmodium* protein, unknown function	PF3D7_0417400
Hcol_000034400	conserved *Plasmodium* protein, unknown function	PF3D7_0417700
Hcol_000034500	cdc2-related protein kinase 1	PF3D7_0417800
Hcol_000034600	conserved *Plasmodium* protein, unknown function	PF3D7_0417900
Hcol_000034700	conserved *Plasmodium* protein, unknown function	PF3D7_0418000
Hcol_000034800	conserved *Plasmodium* protein, unknown function	PF3D7_0418000
Hcol_000034900	eukaryotic translation initiation factor 3 subunit M, putative	PF3D7_0418200
Hcol_000035000	conserved *Plasmodium* protein, unknown function	PF3D7_0418300
Hcol_000035100	conserved *Plasmodium* protein, unknown function	PF3D7_0418400
Hcol_000035200	regulator of chromosome condensation, putative	PF3D7_0418600
Hcol_000035300	RNA-binding protein NOB1, putative	PF3D7_0418700
Hcol_000035400	conserved *Plasmodium* protein, unknown function	PF3D7_0418900
Hcol_000035500	conserved *Plasmodium* protein, unknown function	PF3D7_0419000
Hcol_000035600	conserved *Plasmodium* protein, unknown function	PF3D7_0419100
Hcol_000035700	conserved *Plasmodium* protein, unknown function	PF3D7_0419300
Hcol_000035800	conserved *Plasmodium* protein, unknown function	PF3D7_0419400
Hcol_000035900	zinc finger protein, putative	PF3D7_0420000
Hcol_000036000	serine/threonine protein kinase RIO2	PF3D7_0420100
Hcol_000036100	holo-[acyl-carrier-protein] synthase, putative	PF3D7_0420200
Hcol_000036200	transcription factor with AP2 domain(s)	PF3D7_0420300
Hcol_000036300	transcription factor with AP2 domain(s)	PF3D7_0420300
Hcol_000036400	ribosome-recycling factor	PF3D7_0420400
Hcol_000036500	conserved *Plasmodium* protein, unknown function	PF3D7_0420500

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000036600	conserved *Plasmodium* protein, unknown function	PF3D7_0420600
Hcol_000036700	steroid dehydrogenase, putative	PF3D7_0422000
Hcol_000036800	erythrocyte membrane-associated antigen	PF3D7_0422200
Hcol_000036900	alpha tubulin 2	PF3D7_0422300
Hcol_000037000	40S ribosomal protein S19	PF3D7_0422400
Hcol_000037100	pre-mRNA-splicing helicase BRR2, putative	PF3D7_0422500
Hcol_000037200	conserved *Plasmodium* protein, unknown function	PF3D7_0422600
Hcol_000037300	eukaryotic initiation factor 4A-III, putative	PF3D7_0422700
Hcol_000037400	AP-4 complex subunit sigma, putative	PF3D7_0423100
Hcol_000037500	BSD-domain protein, putative	PF3D7_0423200
Hcol_000037600	glideosome associated protein with multiple membrane spans 2	PF3D7_0423500
Hcol_000037700	mature parasite-infected erythrocyte surface antigen	PF3D7_0500800
Hcol_000037800	anaphase promoting complex subunit, putative	PF3D7_0501700
Hcol_000037900	chromosome assembly factor 1	PF3D7_0501800
Hcol_000038000	conserved protein, unknown function	PF3D7_0501900
Hcol_000038100	vacuolar protein sorting-associated protein 11, putative	PF3D7_0502000
Hcol_000038200	HCNGP-like protein	PF3D7_0502100
Hcol_000038300	conserved *Plasmodium* protein, unknown function	PF3D7_0502500
Hcol_000038400	DnaJ protein, putative	PF3D7_0502800
Hcol_000038500	conserved *Plasmodium* protein, unknown function	PF3D7_0503200
Hcol_000038600	actin-depolymerizing factor 1	PF3D7_0503400
Hcol_000038700	protein kinase, putative	PF3D7_0503500
Hcol_000038800	myosin B	PF3D7_0503600
Hcol_000038900	cation transporting P-ATPase	PF3D7_0504000
Hcol_000039000	ATP-dependent helicase, putative	PF3D7_0504400
Hcol_000039100	conserved *Plasmodium* protein, unknown function	PF3D7_0504500
Hcol_000039200	2-oxoisovalerate dehydrogenase subunit beta, mitochondrial, putative	PF3D7_0504600
Hcol_000039300	conserved *Plasmodium* protein, unknown function	PF3D7_0504800
Hcol_000039400	phd finger protein, putative	PF3D7_0505000
Hcol_000039500	trafficking protein particle complex subunit 8, putative	PF3D7_0505100
Hcol_000039600	actin-like protein, putative	PF3D7_0505200
Hcol_000039700	UDP-N-acetyl glucosamine:UMP antiporter	PF3D7_0505300
Hcol_000039800	conserved *Plasmodium* protein, unknown function	PF3D7_0505400
Hcol_000039900	conserved *Plasmodium* protein, unknown function	PF3D7_0505600
Hcol_000040000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0505700
Hcol_000040100	TATA-box-binding protein	PF3D7_0506200
Hcol_000040200	conserved *Plasmodium* protein, unknown function	PF3D7_0506300
Hcol_000040300	conserved *Plasmodium* protein, unknown function	PF3D7_0506400
Hcol_000040400	conserved *Plasmodium* protein, unknown function	PF3D7_0506500
Hcol_000040500	rhomboid protease ROM4	PF3D7_0506900
Hcol_000040600	60S ribosomal protein L4	PF3D7_0507100
Hcol_000040700	subtilisin-like protease 1	PF3D7_0507500
Hcol_000040800	nuclear protein localization protein 4, putative	PF3D7_0507700
Hcol_000040900	6-cysteine protein	PF3D7_0508000
Hcol_000041000	SET domain protein, putative	PF3D7_0508100
Hcol_000041100	longevity-assurance (LAG1) protein, putative	PF3D7_0508200
Hcol_000041200	triose phosphate transporter	PF3D7_0508300
Hcol_000041300	transcription factor IIb, putative	PF3D7_0508400
Hcol_000041400	guanidine nucleotide exchange factor	PF3D7_0508500
Hcol_000041500	conserved *Plasmodium* protein, unknown function	PF3D7_0508600
Hcol_000041600	pre-mRNA-processing ATP-dependent RNA helicase PRP5, putative	PF3D7_0508700
Hcol_000041700	single-stranded DNA-binding protein	PF3D7_0508800
Hcol_000041800	SNAP protein (soluble N-ethylmaleimide-sensitive factor attachment protein), putative	PF3D7_0509000
Hcol_000041900	chromosome condensation protein, putative	PF3D7_0509100
Hcol_000042000	chromosome condensation protein, putative	PF3D7_0509100
Hcol_000042100	leucine-rich repeat protein	PF3D7_0509200
Hcol_000042200	RNA polymerase I	PF3D7_0509400
Hcol_000042300	RNA polymerase I	PF3D7_0509400
Hcol_000042400	conserved *Plasmodium* protein, unknown function	PF3D7_0509500
Hcol_000042500	asparagine–tRNA ligase	PF3D7_0509600
Hcol_000042600	phosphatidylinositol 4-kinase	PF3D7_0509800
Hcol_000042700	phosphatidylinositol 4-kinase	PF3D7_0509800
Hcol_000042800	conserved *Plasmodium* protein, unknown function	PF3D7_0510000
Hcol_000042900	conserved *Plasmodium* protein, unknown function	PF3D7_0510100
Hcol_000043000	peptidyl-prolyl cis-trans isomerase	PF3D7_0510200
Hcol_000043100	stripes inner membrane complex protein, putative	PF3D7_0510300
Hcol_000043200	conserved *Plasmodium* protein, unknown function	PF3D7_0510400
Hcol_000043300	WD repeat-containing protein, putative	PF3D7_0510800
Hcol_000043400	WD repeat-containing protein, putative	PF3D7_0510800
Hcol_000043500	translationally-controlled tumor protein homolog	PF3D7_0511000

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000043600	stearoyl-CoA desaturase	PF3D7_0511200
Hcol_000043700	MORN repeat protein, putative	PF3D7_0511300
Hcol_000043800	RNA pseudouridylate synthase, putative	PF3D7_0511500
Hcol_000043900	RNA pseudouridylate synthase, putative	PF3D7_0511500
Hcol_000044000	EKC/KEOPS complex subunit CGI121	PF3D7_0511700
Hcol_000044100	conserved *Plasmodium* protein, unknown function	PF3D7_0512100
Hcol_000044200	glutathione synthetase	PF3D7_0512200
Hcol_000044300	CDK-activating kinase assembly factor	PF3D7_0512300
Hcol_000044400	conserved *Plasmodium* protein, unknown function	PF3D7_0512500
Hcol_000044500	ras-related protein Rab-1B	PF3D7_0512600
Hcol_000044600	conserved *Plasmodium* protein, unknown function	PF3D7_0512800
Hcol_000044700	conserved *Plasmodium* protein, unknown function	PF3D7_0513000
Hcol_000044800	conserved *Plasmodium* protein, unknown function	PF3D7_0513100
Hcol_000044900	purine nucleoside phosphorylase	PF3D7_0513300
Hcol_000045000	GTP-binding protein, putative	PF3D7_0513400
Hcol_000045100	mitochondrial import inner membrane translocase subunit TIM16, putative	PF3D7_0513500
Hcol_000045200	deoxyribodipyrimidine photo-lyase, putative	PF3D7_0513600
Hcol_000045300	secreted ookinete protein, putative	PF3D7_0513700
Hcol_000045400	conserved *Plasmodium* protein, unknown function	PF3D7_0513900
Hcol_000045500	tubulin–tyrosine ligase, putative	PF3D7_0514000
Hcol_000045600	ATP-dependent DNA helicase UvrD	PF3D7_0514100
Hcol_000045700	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0514500
Hcol_000045800	ribose 5-phosphate epimerase, putative	PF3D7_0514600
Hcol_000045900	conserved *Plasmodium* protein, unknown function	PF3D7_0514900
Hcol_000046000	pre-mRNA-splicing factor CWC2, putative	PF3D7_0515000
Hcol_000046100	rhomboid protease ROM9	PF3D7_0515100
Hcol_000046200	conserved *Plasmodium* protein, unknown function	PF3D7_0515200
Hcol_000046300	phosphatidylinositol 3-kinase	PF3D7_0515300
Hcol_000046400	conserved *Plasmodium* protein, unknown function	PF3D7_0515400
Hcol_000046500	conserved *Plasmodium* protein, unknown function	PF3D7_0515600
Hcol_000046600	BolA-like protein, putative	PF3D7_0515800
Hcol_000046700	RAP protein, putative	PF3D7_0516000
Hcol_000046800	cation-transporting ATPase 1	PF3D7_0516100
Hcol_000046900	cation-transporting ATPase 1	PF3D7_0516100
Hcol_000047000	40S ribosomal protein S11	PF3D7_0516200
Hcol_000047100	tRNA pseudouridine synthase, putative	PF3D7_0516300
Hcol_000047200	tRNA pseudouridine synthase, putative	PF3D7_0516300
Hcol_000047300	metabolite/drug transporter, putative	PF3D7_0516500
Hcol_000047400	sporozoite surface antigen MB2	PF3D7_0516600
Hcol_000047500	ubiquitin carboxyl-terminal hydrolase 2, putative	PF3D7_0516700
Hcol_000047600	transcription factor with AP2 domain(s)	PF3D7_0516800
Hcol_000047700	60S ribosomal protein L2	PF3D7_0516900
Hcol_000047800	60S ribosomal protein L12, putative	PF3D7_0517000
Hcol_000047900	conserved *Plasmodium* protein, unknown function	PF3D7_0517100
Hcol_000048000	conserved *Plasmodium* protein, unknown function	PF3D7_0517200
Hcol_000048100	FACT complex subunit SPT16, putative	PF3D7_0517400
Hcol_000048200	UTP–glucose-1-phosphate uridylyltransferase, putative	PF3D7_0517500
Hcol_000048300	UTP–glucose-1-phosphate uridylyltransferase, putative	PF3D7_0517500
Hcol_000048400	eukaryotic translation initiation factor 3 subunit B, putative	PF3D7_0517700
Hcol_000048500	zinc finger protein, putative	PF3D7_0517900
Hcol_000048600	RAP protein, putative	PF3D7_0518100
Hcol_000048700	cyclin	PF3D7_0518400
Hcol_000048800	ATP-dependent RNA helicase DDX23, putative	PF3D7_0518500
Hcol_000048900	WD repeat-containing protein 26, putative	PF3D7_0518600
Hcol_000049000	mRNA-binding protein PUF1	PF3D7_0518700
Hcol_000049100	secreted ookinete protein, putative	PF3D7_0518800
Hcol_000049200	conserved *Plasmodium* protein, unknown function	PF3D7_0518900
Hcol_000049300	histone RNA hairpin-binding protein, putative	PF3D7_0519000
Hcol_000049400	mitochondrial ribosomal protein L14 precursor, putative	PF3D7_0519100
Hcol_000049500	cytochrome c oxidase assembly protein (heme A: farnesyltransferase), putative	PF3D7_0519300
Hcol_000049600	40S ribosomal protein S24	PF3D7_0519400
Hcol_000049700	carbon catabolite repressor protein 4, putative	PF3D7_0519500
Hcol_000049800	carbon catabolite repressor protein 4, putative	PF3D7_0519500
Hcol_000049900	zinc finger protein, putative	PF3D7_0519600
Hcol_000050000	conserved *Plasmodium* protein, unknown function	PF3D7_0519700
Hcol_000050100	conserved protein, unknown function	PF3D7_0519800
Hcol_000050200	conserved *Plasmodium* protein, unknown function	PF3D7_0519900
Hcol_000050300	protein phosphatase PPM9, putative	PF3D7_0520100
Hcol_000050400	conserved *Plasmodium* protein, unknown function	PF3D7_0520200
Hcol_000050500	U6 snRNA-associated Sm-like protein LSm2, putative	PF3D7_0520300

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000050600	conserved *Plasmodium* protein, unknown function	PF3D7_0520400
Hcol_000050700	phosphomethylpyrimidine kinase, putative	PF3D7_0520500
Hcol_000050800	conserved *Plasmodium* protein, unknown function	PF3D7_0520700
Hcol_000050900	conserved *Plasmodium* protein, unknown function	PF3D7_0520800
Hcol_000051000	S-adenosyl-L-homocysteine hydrolase	PF3D7_0520900
Hcol_000051100	conserved *Plasmodium* protein, unknown function	PF3D7_0521200
Hcol_000051200	conserved *Plasmodium* protein, unknown function	PF3D7_0521300
Hcol_000051300	conserved *Plasmodium* protein, unknown function	PF3D7_0521300
Hcol_000051400	conserved *Plasmodium* protein, unknown function	PF3D7_0521400
Hcol_000051500	ribosomal large subunit pseudouridylate synthase, putative	PF3D7_0521500
Hcol_000051600	DEAD/DEAH box ATP-dependent RNA helicase, putative	PF3D7_0521700
Hcol_000051700	nucleotide binding protein, putative	PF3D7_0521800
Hcol_000051800	conserved *Plasmodium* protein, unknown function	PF3D7_0522100
Hcol_000051900	conserved *Plasmodium* protein, unknown function	PF3D7_0522400
Hcol_000052000	conserved *Plasmodium* protein, unknown function	PF3D7_0522400
Hcol_000052100	50S ribosomal protein L17, apicoplast, putative	PF3D7_0522500
Hcol_000052200	inner membrane complex protein	PF3D7_0522600
Hcol_000052300	iron-sulfur assembly protein, putative	PF3D7_0522700
Hcol_000052400	zinc finger protein, putative	PF3D7_0522900
Hcol_000052500	multidrug resistance protein 1	PF3D7_0523000
Hcol_000052600	mitochondrial processing peptidase alpha subunit, putative	PF3D7_0523100
Hcol_000052700	conserved *Plasmodium* protein, unknown function	PF3D7_0523200
Hcol_000052800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0523700
Hcol_000052900	transporter, putative	PF3D7_0523800
Hcol_000053000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0523900
Hcol_000053100	karyopherin beta	PF3D7_0524000
Hcol_000053200	conserved *Plasmodium* protein, unknown function	PF3D7_0524300
Hcol_000053300	ribosome-interacting GTPase 1, putative	PF3D7_0524400
Hcol_000053400	50S ribosomal protein L12, apicoplast, putative	PF3D7_0524600
Hcol_000053500	mitochondrial import receptor subunit TOM22, putative	PF3D7_0524700
Hcol_000053600	ubiquitin fusion degradation protein 1, putative	PF3D7_0524800
Hcol_000053700	zinc finger protein, putative	PF3D7_0525000
Hcol_000053800	structural maintenance of chromosomes protein 6, putative	PF3D7_0525200
Hcol_000053900	conserved *Plasmodium* protein, unknown function	PF3D7_0525300
Hcol_000054000	G-protein coupled receptor, putative	PF3D7_0525400
Hcol_000054100	WD repeat-containing protein, putative	PF3D7_0525500
Hcol_000054200	RNA methyltransferase, putative	PF3D7_0525600
Hcol_000054300	conserved *Plasmodium* protein, unknown function	PF3D7_0525700
Hcol_000054400	RAP protein, putative	PF3D7_0526000
Hcol_000054500	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0526100
Hcol_000054600	conserved *Plasmodium* protein, unknown function	PF3D7_0526400
Hcol_000054700	conserved *Plasmodium* protein, unknown function	PF3D7_0526500
Hcol_000054800	conserved *Plasmodium* protein, unknown function	PF3D7_0526700
Hcol_000054900	conserved *Plasmodium* protein, unknown function	PF3D7_0526800
Hcol_000055000	transmembrane emp24 domain-containing protein, putative	PF3D7_0526900
Hcol_000055100	DNA replication licensing factor MCM3, putative	PF3D7_0527000
Hcol_000055200	ubiquitin-conjugating enzyme E2 N, putative	PF3D7_0527100
Hcol_000055300	ubiquitin carboxyl-terminal hydrolase 14, putative	PF3D7_0527200
Hcol_000055400	methionine aminopeptidase 1a, putative	PF3D7_0527300
Hcol_000055500	conserved *Plasmodium* protein, unknown function	PF3D7_0527600
Hcol_000055600	proteasome maturation factor UMP1, putative	PF3D7_0528000
Hcol_000055700	AP-1 complex subunit beta, putative	PF3D7_0528100
Hcol_000055800	eukaryotic translation initiation factor 3 subunit E, putative	PF3D7_0528200
Hcol_000055900	conserved *Plasmodium* protein, unknown function	PF3D7_0528300
Hcol_000056000	palmitoyltransferase	PF3D7_0528400
Hcol_000056100	nucleolar preribosomal GTPase, putative	PF3D7_0528800
Hcol_000056200	conserved *Plasmodium* protein, unknown function	PF3D7_0528900
Hcol_000056300	conserved *Plasmodium* protein, unknown function	PF3D7_0529000
Hcol_000056400	apicoplast TIC22 protein	PF3D7_0529300
Hcol_000056500	conserved *Plasmodium* protein, unknown function	PF3D7_0529400
Hcol_000056600	conserved *Plasmodium* protein, unknown function	PF3D7_0529400
Hcol_000056700	conserved *Plasmodium* protein, unknown function	PF3D7_0529800
Hcol_000056800	RING zinc finger protein, putative	PF3D7_0529900
Hcol_000056900	SNARE protein, putative	PF3D7_0530100
Hcol_000057000	phosphoenolpyruvate/phosphate translocator	PF3D7_0530200
Hcol_000057100	XAP-5 DNA binding protein, putative	PF3D7_0530600
Hcol_000057200	conserved *Plasmodium* protein, unknown function	PF3D7_0530700
Hcol_000057300	dolichol-phosphate mannosyltransferase subunit 3, putative	PF3D7_0530750
Hcol_000057400	formin 1	PF3D7_0530900
Hcol_000057500	formin 1	PF3D7_0530900
Hcol_000057600	conserved *Plasmodium* protein, unknown function	PF3D7_0531000
Hcol_000057700	mitochondrial ribosomal protein S16 precursor, putative	PF3D7_0531200

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000057800	conserved *Plasmodium* protein, unknown function	PF3D7_0531400
Hcol_000057900	ATP-dependent RNA helicase, putative	PF3D7_0602100
Hcol_000058000	MYND finger protein, putative	PF3D7_0602200
Hcol_000058100	liver merozoite formation protein, putative	PF3D7_0602300
Hcol_000058200	elongation factor G	PF3D7_0602400
Hcol_000058300	geranylgeranyltransferase, putative	PF3D7_0602500
Hcol_000058400	conserved *Plasmodium* protein, unknown function	PF3D7_0602600
Hcol_000058500	conserved *Plasmodium* protein, unknown function	PF3D7_0602600
Hcol_000058600	conserved *Plasmodium* protein, unknown function	PF3D7_0602700
Hcol_000058700	conserved *Plasmodium* protein, unknown function	PF3D7_0603000
Hcol_000058800	conserved *Plasmodium* protein, unknown function	PF3D7_0603100
Hcol_000058900	mitochondrial chaperone BCS1, putative	PF3D7_0603200
Hcol_000059000	dihydroorotate dehydrogenase	PF3D7_0603300
Hcol_000059100	trophozoite exported protein 1	PF3D7_0603400
Hcol_000059200	trophozoite exported protein 1	PF3D7_0603400
Hcol_000059300	cation/H+ antiporter	PF3D7_0603500
Hcol_000059400	phenylalanine–tRNA ligase	PF3D7_0603700
Hcol_000059500	centrosomal protein CEP76, putative	PF3D7_0603800
Hcol_000059600	conserved *Plasmodium* protein, unknown function	PF3D7_0603900
Hcol_000059700	transcription factor with AP2 domain(s)	PF3D7_0604100
Hcol_000059800	conserved *Plasmodium* protein, unknown function	PF3D7_0604300
Hcol_000059900	conserved *Plasmodium* protein, unknown function	PF3D7_0604500
Hcol_000060000	glyoxalase I	PF3D7_0604700
Hcol_000060100	RAP protein, putative	PF3D7_0604800
Hcol_000060200	50S ribosomal protein L24, putative	PF3D7_0605000
Hcol_000060300	RNA-binding protein, putative	PF3D7_0605100
Hcol_000060400	serine/threonine protein kinase	PF3D7_0605300
Hcol_000060500	calcium-binding protein, putative	PF3D7_0605400
Hcol_000060600	nucleoside diphosphate kinase, putative	PF3D7_0605600
Hcol_000060700	conserved *Plasmodium* protein, unknown function	PF3D7_0605700
Hcol_000060800	DNA repair protein RAD50, putative	PF3D7_0605800
Hcol_000060900	conserved *Plasmodium* protein, unknown function	PF3D7_0606000
Hcol_000061000	RNA-binding protein, putative	PF3D7_0606100
Hcol_000061100	ubiquitin-conjugating enzyme E2, putative	PF3D7_0606200
Hcol_000061200	polypyrimidine tract binding protein, putative	PF3D7_0606500
Hcol_000061300	conserved *Plasmodium* protein, unknown function	PF3D7_0606600
Hcol_000061400	glutaredoxin-like protein	PF3D7_0606900
Hcol_000061500	translation initiation factor IF-2, putative	PF3D7_0607000
Hcol_000061600	MYND finger protein, putative	PF3D7_0607100
Hcol_000061700	RING zinc finger protein, putative	PF3D7_0607200
Hcol_000061800	para-hydroxybenzoate–polyprenyltransferase, putative	PF3D7_0607500
Hcol_000061900	spindle assembly abnormal protein 6, putative	PF3D7_0607600
Hcol_000062000	conserved *Plasmodium* protein, unknown function	PF3D7_0607700
Hcol_000062100	GTP-binding protein, putative	PF3D7_0607800
Hcol_000062200	diphthine methyltransferase, putative	PF3D7_0608000
Hcol_000062300	conserved *Plasmodium* protein, unknown function	PF3D7_0608100
Hcol_000062400	conserved *Plasmodium* protein, unknown function	PF3D7_0608200
Hcol_000062500	conserved *Plasmodium* protein, unknown function	PF3D7_0608300
Hcol_000062600	conserved *Plasmodium* protein, unknown function	PF3D7_0608300
Hcol_000062700	sorting assembly machinery 50 kDa subunit, putative	PF3D7_0608310
Hcol_000062800	proteasome subunit alpha type-2, putative	PF3D7_0608500
Hcol_000062900	conserved *Plasmodium* protein, unknown function	PF3D7_0608600
Hcol_000063000	T-complex protein 1 subunit zeta	PF3D7_0608700
Hcol_000063100	conserved *Plasmodium* protein, unknown function	PF3D7_0608900
Hcol_000063200	conserved *Plasmodium* protein, unknown function	PF3D7_0609000
Hcol_000063300	conserved *Plasmodium* protein, unknown function	PF3D7_0609000
Hcol_000063400	Zn2+ or Fe2+ permease	PF3D7_0609100
Hcol_000063500	citrate synthase-like protein, putative	PF3D7_0609200
Hcol_000063600	mitochondrial cardiolipin synthase, putative	PF3D7_0609400
Hcol_000063700	hypothetical protein, conserved	PF3D7_0609600
Hcol_000063800	conserved *Plasmodium* protein, unknown function	PF3D7_0609900
Hcol_000063900	mitochondrial ribosomal protein L19 precursor, putative	PF3D7_0610000
Hcol_000064000	pre-mRNA-splicing factor SLU7, putative	PF3D7_0610100
Hcol_000064100	conserved *Plasmodium* protein, unknown function	PF3D7_0610200
Hcol_000064200	histone H3	PF3D7_0610400
Hcol_000064300	peptidyl-tRNA hydrolase PTRHD1, putative	PF3D7_0610500
Hcol_000064400	transketolase	PF3D7_0610800
Hcol_000064500	SNARE associated Golgi protein, putative	PF3D7_0611000
Hcol_000064600	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin	PF3D7_0611400
Hcol_000064700	conserved *Plasmodium* protein, unknown function	PF3D7_0611500
Hcol_000064800	conserved *Plasmodium* protein, unknown function	PF3D7_0611600

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000064900	conserved *Plasmodium* protein, unknown function	PF3D7_0611800
Hcol_000065000	lsm12, putative	PF3D7_0611900
Hcol_000065100	eukaryotic translation initiation factor 3 subunit L, putative	PF3D7_0612100
Hcol_000065200	leucine-rich repeat protein	PF3D7_0612200
Hcol_000065300	conserved protein, unknown function	PF3D7_0612300
Hcol_000065400	conserved *Plasmodium* protein, unknown function	PF3D7_0612400
Hcol_000065500	conserved *Plasmodium* protein, unknown function	PF3D7_0612500
Hcol_000065600	cytoplasmic tRNA 2-thiolation protein 1, putative	PF3D7_0612600
Hcol_000065700	6-cysteine protein	PF3D7_0612800
Hcol_000065800	6-cysteine protein	PF3D7_0612800
Hcol_000065900	nucleolar GTP-binding protein 1, putative	PF3D7_0612900
Hcol_000066000	rhoptry protein ROP14	PF3D7_0613300
Hcol_000066100	apicoplast ribosomal protein L18 precursor, putative	PF3D7_0613400
Hcol_000066200	AP-3 complex subunit beta, putative	PF3D7_0613500
Hcol_000066300	AP-3 complex subunit beta, putative	PF3D7_0613500
Hcol_000066400	syntaxin binding protein, putative	PF3D7_0613700
Hcol_000066500	myosin E, putative	PF3D7_0613900
Hcol_000066600	thiamin-phosphate pyrophosphorylase, putative	PF3D7_0614000
Hcol_000066700	conserved *Plasmodium* protein, unknown function	PF3D7_0614100
Hcol_000066800	cytosolic Fe-S cluster assembly factor NAR1, putative	PF3D7_0614200
Hcol_000066900	conserved protein, unknown function	PF3D7_0614400
Hcol_000067000	60S ribosomal protein L19	PF3D7_0614500
Hcol_000067100	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0614900
Hcol_000067200	enoyl-acyl carrier reductase	PF3D7_0615100
Hcol_000067300	conserved *Plasmodium* protein, unknown function	PF3D7_0615200
Hcol_000067400	GPI-anchored wall transfer protein 1, putative	PF3D7_0615300
Hcol_000067500	ribonuclease, putative	PF3D7_0615400
Hcol_000067600	cdc2-related protein kinase 5	PF3D7_0615500
Hcol_000067700	conserved *Plasmodium* protein, unknown function	PF3D7_0615600
Hcol_000067800	conserved *Plasmodium* protein, unknown function	PF3D7_0615800
Hcol_000067900	conserved *Plasmodium* protein, unknown function	PF3D7_0615800
Hcol_000068000	conserved *Plasmodium* protein, unknown function	PF3D7_0615900
Hcol_000068100	conserved *Plasmodium* protein, unknown function	PF3D7_0616300
Hcol_000068200	conserved *Plasmodium* protein, unknown function	PF3D7_0616400
Hcol_000068300	TRAP-like protein	PF3D7_0616500
Hcol_000068400	conserved *Plasmodium* protein, unknown function	PF3D7_0616600
Hcol_000068500	ras GTPAse, putative	PF3D7_0616700
Hcol_000068600	malate:quinone oxidoreductase, putative	PF3D7_0616800
Hcol_000068700	conserved *Plasmodium* protein, unknown function	PF3D7_0616900
Hcol_000068800	mitochondrial import receptor subunit TOM40, putative	PF3D7_0617000
Hcol_000068900	AP-2 complex subunit alpha, putative	PF3D7_0617100
Hcol_000069000	conserved *Plasmodium* protein, unknown function	PF3D7_0617200
Hcol_000069100	conserved *Plasmodium* protein, unknown function	PF3D7_0617300
Hcol_000069200	histone H2A	PF3D7_0617800
Hcol_000069300	histone H3 variant, putative	PF3D7_0617900
Hcol_000069400	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0618000
Hcol_000069500	conserved *Plasmodium* protein, unknown function	PF3D7_0618100
Hcol_000069600	60S ribosomal protein L27a, putative	PF3D7_0618300
Hcol_000069700	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0618400
Hcol_000069800	malate dehydrogenase	PF3D7_0618500
Hcol_000069900	conserved *Plasmodium* protein, unknown function	PF3D7_0619000
Hcol_000070000	conserved *Plasmodium* protein, unknown function	PF3D7_0619100
Hcol_000070100	conserved *Plasmodium* protein, unknown function	PF3D7_0619200
Hcol_000070200	conserved *Plasmodium* protein, unknown function	PF3D7_0619300
Hcol_000070300	cell division cycle protein 48 homologue, putative	PF3D7_0619400
Hcol_000070400	acyl-CoA synthetase	PF3D7_0619500
Hcol_000070500	conserved *Plasmodium* protein, unknown function	PF3D7_0619600
Hcol_000070600	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0619800
Hcol_000070700	conserved *Plasmodium* protein, unknown function	PF3D7_0620100
Hcol_000070800	conserved *Plasmodium* protein, unknown function	PF3D7_0620200
Hcol_000070900	conserved *Plasmodium* protein, unknown function	PF3D7_0620300
Hcol_000071000	merozoite surface protein 10	PF3D7_0620400
Hcol_000071100	conserved *Plasmodium* protein, unknown function	PF3D7_0620600
Hcol_000071200	DnaJ protein, putative	PF3D7_0620700
Hcol_000071300	conserved *Plasmodium* protein, unknown function	PF3D7_0621100
Hcol_000071400	Pf77 protein	PF3D7_0621400
Hcol_000071500	ribonuclease P/MRP protein subunit RPP1, putative	PF3D7_0621500
Hcol_000071600	conserved *Plasmodium* protein, unknown function	PF3D7_0622100
Hcol_000071700	conserved Apicomplexan protein, unknown function	PF3D7_0622400
Hcol_000071800	RNA methyltransferase, putative	PF3D7_0622500
Hcol_000071900	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0622700
Hcol_000072000	nuclear polyadenylated RNA-binding protein NAB2, putative	PF3D7_0623100

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000072100	ferredoxin–NADP reductase	PF3D7_0623200
Hcol_000072200	superoxide dismutase [Fe]	PF3D7_0623500
Hcol_000072300	transcription or splicing factor-like protein, putative	PF3D7_0623600
Hcol_000072400	ATP dependent DEAD-box helicase, putative	PF3D7_0623700
Hcol_000072500	tyrosine kinase-like protein, putative	PF3D7_0623800
Hcol_000072600	conserved *Plasmodium* protein, unknown function	PF3D7_0624100
Hcol_000072700	conserved *Plasmodium* protein, unknown function	PF3D7_0624400
Hcol_000072800	sphingomyelin synthase 1, putative	PF3D7_0625000
Hcol_000072900	DNA polymerase 1, putative	PF3D7_0625300
Hcol_000073000	poly(A) polymerase PAP, putative	PF3D7_0625600
Hcol_000073100	conserved *Plasmodium* protein, unknown function	PF3D7_0625700
Hcol_000073200	conserved *Plasmodium* protein, unknown function	PF3D7_0626000
Hcol_000073300	conserved *Plasmodium* protein, unknown function	PF3D7_0626000
Hcol_000073400	conserved *Plasmodium* protein, unknown function	PF3D7_0626200
Hcol_000073500	3-oxoacyl-acyl-carrier protein synthase I/II	PF3D7_0626300
Hcol_000073600	Sec14 domain containing protein	PF3D7_0626400
Hcol_000073700	conserved *Plasmodium* protein, unknown function	PF3D7_0626500
Hcol_000073800	pyruvate kinase	PF3D7_0626800
Hcol_000073900	mitochondrial ribosomal protein L46 precursor, putative	PF3D7_0626900
Hcol_000074000	ankyrin-repeat protein, putative	PF3D7_0627100
Hcol_000074100	mitochondrial import inner membrane translocase subunit TIM22, putative	PF3D7_0627400
Hcol_000074200	protein DJ-1	PF3D7_0627500
Hcol_000074300	conserved *Plasmodium* protein, unknown function	PF3D7_0627600
Hcol_000074400	transportin	PF3D7_0627700
Hcol_000074500	acetyl-CoA synthetase, putative	PF3D7_0627800
Hcol_000074600	ribonuclease P protein subunit p29, putative	PF3D7_0627900
Hcol_000074700	HECT-domain (ubiquitin-transferase), putative	PF3D7_0628100
Hcol_000074800	protein kinase PK4	PF3D7_0628200
Hcol_000074900	conserved *Plasmodium* protein, unknown function	PF3D7_0628500
Hcol_000075000	conserved *Plasmodium* protein, unknown function	PF3D7_0628600
Hcol_000075100	conserved *Plasmodium* protein, unknown function	PF3D7_0628700
Hcol_000075200	glutamyl-tRNA(Gln) amidotransferase subunit B	PF3D7_0628800
Hcol_000075300	RAP protein, putative	PF3D7_0628900
Hcol_000075400	N-acetyltransferase, putative	PF3D7_0629000
Hcol_000075500	nicotinate phosphoribosyltransferase, putative	PF3D7_0629100
Hcol_000075600	DnaJ protein, putative	PF3D7_0629200
Hcol_000075700	phospholipase, putative	PF3D7_0629300
Hcol_000075800	amino acid transporter, putative	PF3D7_0629500
Hcol_000075900	conserved *Plasmodium* protein, unknown function	PF3D7_0629600
Hcol_000076000	SET domain protein, putative	PF3D7_0629700
Hcol_000076100	SET domain protein, putative	PF3D7_0629700
Hcol_000076200	cullin-like protein, putative	PF3D7_0629800
Hcol_000076300	cullin-like protein, putative	PF3D7_0629800
Hcol_000076400	conserved *Plasmodium* protein, unknown function	PF3D7_0630100
Hcol_000076500	DNA polymerase epsilon catalytic subunit A, putative	PF3D7_0630300
Hcol_000076600	DNA polymerase epsilon catalytic subunit A, putative	PF3D7_0630300
Hcol_000076700	DNA polymerase epsilon catalytic subunit A, putative	PF3D7_0630300
Hcol_000076800	ATP-dependent RNA helicase HAS1	PF3D7_0630900
Hcol_000076900	centrin, putative	PF3D7_0702900
Hcol_000077000	conserved *Plasmodium* protein, unknown function	PF3D7_0703000
Hcol_000077100	conserved *Plasmodium* protein, unknown function	PF3D7_0703000
Hcol_000077200	conserved *Plasmodium* protein, unknown function	PF3D7_0703100
Hcol_000077300	conserved *Plasmodium* protein, unknown function	PF3D7_0703200
Hcol_000077400	conserved *Plasmodium* protein, unknown function	PF3D7_0703400
Hcol_000077500	erythrocyte membrane-associated antigen	PF3D7_0703500
Hcol_000077600	conserved *Plasmodium* protein, unknown function	PF3D7_0703700
Hcol_000077700	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0703900
Hcol_000077800	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0704000
Hcol_000077900	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0704000
Hcol_000078000	phosphoinositide-binding protein, putative	PF3D7_0704400
Hcol_000078100	serine/threonine protein kinase, putative	PF3D7_0704500
Hcol_000078200	E3 ubiquitin-protein ligase	PF3D7_0704600
Hcol_000078300	E3 ubiquitin-protein ligase	PF3D7_0704600
Hcol_000078400	phosphopantetheine adenylyltransferase, putative	PF3D7_0704700
Hcol_000078500	phosphopantetheine adenylyltransferase, putative	PF3D7_0704700
Hcol_000078600	peptide chain release factor, putative	PF3D7_0704900
Hcol_000078700	methyltransferase, putative	PF3D7_0705000
Hcol_000078800	methyltransferase, putative	PF3D7_0705000
Hcol_000078900	conserved *Plasmodium* protein, unknown function	PF3D7_0705100
Hcol_000079000	conserved *Plasmodium* protein, unknown function	PF3D7_0705200
Hcol_000079100	DNA replication licensing factor MCM7	PF3D7_0705400

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000079200	inositol-phosphate phosphatase, putative	PF3D7_0705500
Hcol_000079300	inositol-phosphate phosphatase, putative	PF3D7_0705500
Hcol_000079400	RNA helicase, putative	PF3D7_0705600
Hcol_000079500	cysteine-rich secretory protein, putative	PF3D7_0705800
Hcol_000079600	ATP synthase subunit C, putative	PF3D7_0705900
Hcol_000079700	importin-7, putative	PF3D7_0706000
Hcol_000079800	conserved *Plasmodium* protein, unknown function	PF3D7_0706100
Hcol_000079900	conserved *Plasmodium* protein, unknown function	PF3D7_0706500
Hcol_000080000	conserved *Plasmodium* protein, unknown function	PF3D7_0707200
Hcol_000080100	conserved *Plasmodium* protein, unknown function	PF3D7_0707500
Hcol_000080200	conserved *Plasmodium* protein, unknown function	PF3D7_0707500
Hcol_000080300	conserved *Plasmodium* protein, unknown function	PF3D7_0707600
Hcol_000080400	E3 ubiquitin-protein ligase, putative	PF3D7_0707700
Hcol_000080500	ribosomal protein S8e, putative	PF3D7_0707900
Hcol_000080600	cytoskeleton associated protein, putative	PF3D7_0708000
Hcol_000080700	conserved *Plasmodium* protein, unknown function	PF3D7_0708200
Hcol_000080800	EKC/KEOPS complex subunit BUD32	PF3D7_0708300
Hcol_000080900	heat shock protein 90	PF3D7_0708400
Hcol_000081000	heat shock protein 86 family protein	PF3D7_0708500
Hcol_000081100	inner membrane complex protein 1d, putative	PF3D7_0708600
Hcol_000081200	Cg8 protein	PF3D7_0708700
Hcol_000081300	heat shock protein 110	PF3D7_0708800
Hcol_000081400	Cg3 protein	PF3D7_0708900
Hcol_000081500	Cg1 protein	PF3D7_0709100
Hcol_000081600	Cg2 protein	PF3D7_0709300
Hcol_000081700	Cg7 protein	PF3D7_0709400
Hcol_000081800	conserved *Plasmodium* protein, unknown function	PF3D7_0709500
Hcol_000081900	ribonucleases P/MRP protein subunit POP1, putative	PF3D7_0709600
Hcol_000082000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0709900
Hcol_000082100	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0709900
Hcol_000082200	60S ribosomal protein L34	PF3D7_0710600
Hcol_000082300	50S ribosomal protein L1, mitochondrial, putative	PF3D7_0710900
Hcol_000082400	AAA family ATPase, CDC48 subfamily	PF3D7_0711000
Hcol_000082500	conserved *Plasmodium* protein, unknown function	PF3D7_0711200
Hcol_000082600	histone deacetylase complex subunit SAP18, putative	PF3D7_0711400
Hcol_000082700	regulator of chromosome condensation, putative	PF3D7_0711500
Hcol_000082800	negative elongation factor A, putative	PF3D7_0713800
Hcol_000082900	conserved *Plasmodium* protein, unknown function	PF3D7_0713900
Hcol_000083000	conserved *Plasmodium* protein, unknown function	PF3D7_0714100
Hcol_000083100	conserved *Plasmodium* protein, unknown function	PF3D7_0714200
Hcol_000083200	palmitoyltransferase, putative	PF3D7_0714300
Hcol_000083300	palmitoyltransferase, putative	PF3D7_0714300
Hcol_000083400	calmodulin, putative	PF3D7_0714400
Hcol_000083500	transcription elongation factor s-II, putative	PF3D7_0714500
Hcol_000083600	conserved *Plasmodium* protein, unknown function	PF3D7_0714600
Hcol_000083700	conserved *Plasmodium* protein, unknown function	PF3D7_0715200
Hcol_000083800	calcium/calmodulin-dependent protein kinase, putative	PF3D7_0715300
Hcol_000083900	mitochondrial ATP synthase F1, epsilon subunit, putative	PF3D7_0715500
Hcol_000084000	GTP-binding translation elongation factor tu family protein, putative	PF3D7_0715600
Hcol_000084100	conserved *Plasmodium* protein, unknown function	PF3D7_0715700
Hcol_000084200	zinc transporter, putative	PF3D7_0715900
Hcol_000084300	RNA-binding protein, putative	PF3D7_0716000
Hcol_000084400	protein SDA1, putative	PF3D7_0716100
Hcol_000084500	protein SDA1, putative	PF3D7_0716100
Hcol_000084600	conserved *Plasmodium* protein, unknown function	PF3D7_0716200
Hcol_000084700	eukaryotic translation initiation factor 3 subunit I, putative	PF3D7_0716800
Hcol_000084800	conserved *Plasmodium* protein, unknown function	PF3D7_0717100
Hcol_000084900	conserved *Plasmodium* protein, unknown function	PF3D7_0717200
Hcol_000085000	transcription initiation factor IIE subunit alpha, putative	PF3D7_0717300
Hcol_000085100	calcium-dependent protein kinase 4	PF3D7_0717500
Hcol_000085200	conserved *Plasmodium* protein, unknown function	PF3D7_0717600
Hcol_000085300	conserved *Plasmodium* protein, unknown function	PF3D7_0717800
Hcol_000085400	thioredoxin-like protein	PF3D7_0717900
Hcol_000085500	dynein heavy chain, putative	PF3D7_0718000
Hcol_000085600	dynein heavy chain, putative	PF3D7_0718000
Hcol_000085700	dynein heavy chain, putative	PF3D7_0718000
Hcol_000085800	exported serine/threonine protein kinase	PF3D7_0718100
Hcol_000085900	conserved *Plasmodium* protein, unknown function	PF3D7_0718600
Hcol_000086000	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0718700
Hcol_000086100	conserved *Plasmodium* protein, unknown function	PF3D7_0718800
Hcol_000086200	conserved *Plasmodium* protein, unknown function	PF3D7_0718900
Hcol_000086300	conserved *Plasmodium* protein, unknown function	PF3D7_0719000

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000086400	actin-related protein, putative	PF3D7_0719300
Hcol_000086500	40S ribosomal protein S10, putative	PF3D7_0719700
Hcol_000086600	conserved *Plasmodium* protein, unknown function	PF3D7_0719800
Hcol_000086700	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0719900
Hcol_000086800	exosome complex component CSL4, putative	PF3D7_0720000
Hcol_000086900	conserved *Plasmodium* protein, unknown function	PF3D7_0720300
Hcol_000087000	ferrodoxin reductase-like protein	PF3D7_0720400
Hcol_000087100	phosphoinositide-binding protein, putative	PF3D7_0720700
Hcol_000087200	Ham1-like protein, putative	PF3D7_0720800
Hcol_000087300	conserved *Plasmodium* protein, unknown function	PF3D7_0721100
Hcol_000087400	conserved *Plasmodium* protein, unknown function	PF3D7_0721200
Hcol_000087500	ATP-dependent RNA helicase DBP7, putative	PF3D7_0721300
Hcol_000087600	40S ribosomal protein S5, putative	PF3D7_0721600
Hcol_000087700	secreted ookinete protein, putative	PF3D7_0721700
Hcol_000087800	PelOta protein homologue, putative	PF3D7_0722100
Hcol_000087900	ubiquitin carboxyl-terminal hydrolase, putative	PF3D7_0722300
Hcol_000088000	GTP-binding protein, putative	PF3D7_0722400
Hcol_000088100	pre-mRNA-splicing factor CWC15, putative	PF3D7_0722500
Hcol_000088200	U3 small nucleolar RNA-associated protein 7, putative	PF3D7_0722600
Hcol_000088300	tRNAHis guanylyltransferase, putative	PF3D7_0723000
Hcol_000088400	conserved *Plasmodium* protein, unknown function	PF3D7_0723100
Hcol_000088500	conserved *Plasmodium* protein, unknown function	PF3D7_0723300
Hcol_000088600	conserved *Plasmodium* protein, unknown function	PF3D7_0723400
Hcol_000088700	dynactin subunit 5, putative	PF3D7_0723500
Hcol_000088800	conserved *Plasmodium* protein, unknown function	PF3D7_0723600
Hcol_000088900	metallo-hydrolase/oxidoreductase, putative	PF3D7_0723700
Hcol_000089000	conserved *Plasmodium* protein, unknown function	PF3D7_0723800
Hcol_000089100	conserved *Plasmodium* protein, unknown function	PF3D7_0723800
Hcol_000089200	conserved *Plasmodium* protein, unknown function	PF3D7_0723900
Hcol_000089300	Rab GTPase activator and protein kinase, putative	PF3D7_0724000
Hcol_000089400	type 2A phosphatase-associated protein 42, putative	PF3D7_0724200
Hcol_000089500	3-demethylubiquinone-9 3-methyltransferase, putative	PF3D7_0724300
Hcol_000089600	conserved *Plasmodium* protein, unknown function	PF3D7_0724700
Hcol_000089700	conserved *Plasmodium* protein, unknown function	PF3D7_0724700
Hcol_000089800	kinesin-19, putative	PF3D7_0724900
Hcol_000089900	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0725100
Hcol_000090000	DNA mismatch repair protein PMS1, putative	PF3D7_0726300
Hcol_000090100	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0726400
Hcol_000090200	ubiquitin carboxyl-terminal hydrolase, putative	PF3D7_0726500
Hcol_000090300	ubiquitin carboxyl-terminal hydrolase, putative	PF3D7_0726500
Hcol_000090400	conserved *Plasmodium* protein, unknown function	PF3D7_0726700
Hcol_000090500	dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit DAD1, putative	PF3D7_0726800
Hcol_000090600	mitochondrial import inner membrane translocase subunit TIM50, putative	PF3D7_0726900
Hcol_000090700	conserved *Plasmodium* protein, unknown function	PF3D7_0727100
Hcol_000090800	DNA (cytosine-5)-methyltransferase	PF3D7_0727300
Hcol_000090900	proteasome subunit alpha type-5, putative	PF3D7_0727400
Hcol_000091000	conserved *Plasmodium* protein, unknown function	PF3D7_0727500
Hcol_000091100	conserved *Plasmodium* protein, unknown function	PF3D7_0727700
Hcol_000091200	cation transporting ATPase, putative	PF3D7_0727800
Hcol_000091300	eukaryotic translation initiation factor 2 alpha subunit, putative	PF3D7_0728000
Hcol_000091400	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0728100
Hcol_000091500	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0728100
Hcol_000091600	actin-like protein, putative	PF3D7_0728200
Hcol_000091700	conserved *Plasmodium* protein, unknown function	PF3D7_0728300
Hcol_000091800	conserved *Plasmodium* protein, unknown function	PF3D7_0728400
Hcol_000091900	zinc finger, C3HC4 type, putative	PF3D7_0728600
Hcol_000092000	zinc finger, C3HC4 type, putative	PF3D7_0728600
Hcol_000092100	RNA-binding protein, putative	PF3D7_0728900
Hcol_000092200	conserved *Plasmodium* protein, unknown function	PF3D7_0729100
Hcol_000092300	1-cys peroxiredoxin	PF3D7_0729200
Hcol_000092400	60S ribosomal export protein NMD3, putative	PF3D7_0729300
Hcol_000092500	ribosome biogenesis protein BRX1 homolog, putative	PF3D7_0729400
Hcol_000092600	mRNA (N6-adenosine)-methyltransferase, putative	PF3D7_0729500
Hcol_000092700	conserved *Plasmodium* protein, unknown function	PF3D7_0729600
Hcol_000092800	conserved *Plasmodium* protein, unknown function	PF3D7_0729700
Hcol_000092900	dynein light chain, putative	PF3D7_0729800
Hcol_000093000	dynein heavy chain, putative	PF3D7_0729900

Table 8-1 Genes summary of *H. columbae*. Ortology with *P. falciparum*

H. columbae ID	Protein annotation	Orthologue in *P. falciparum*
Hcol_000093100	conserved *Plasmodium* protein, unknown function	PF3D7_0730000
Hcol_000093200	tRNA pseudouridine synthase D, putative	PF3D7_0730100
Hcol_000093300	AP-4 complex subunit beta, putative	PF3D7_0730200
Hcol_000093400	conserved *Plasmodium* protein, unknown function	PF3D7_0730500
Hcol_000093500	probable protein, unknown function	PF3D7_0801400
Hcol_000093600	conserved *Plasmodium* protein, unknown function	PF3D7_0801600
Hcol_000093700	sentrin-specific protease 2, putative	PF3D7_0801700
Hcol_000093800	conserved *Plasmodium* protein, unknown function	PF3D7_0801900
Hcol_000093900	glutamate dehydrogenase, putative	PF3D7_0802000
Hcol_000094000	1-cys peroxiredoxin	PF3D7_0802200
Hcol_000094100	conserved *Plasmodium* protein, unknown function	PF3D7_0802400
Hcol_000094200	inositol 5-phosphatase, putative	PF3D7_0802500
Hcol_000094300	conserved *Plasmodium* protein, unknown function	PF3D7_0802700
Hcol_000094400	serine/threonine protein phosphatase 2B catalytic subunit A	PF3D7_0802800
Hcol_000094500	conserved *Plasmodium* protein, unknown function	PF3D7_0802900
Hcol_000094600	peptidyl-prolyl cis-trans isomerase	PF3D7_0803000
Hcol_000094700	filament assembling protein, putative	PF3D7_0803200
Hcol_000094800	mitogen-activated protein kinase organizer 1, putative	PF3D7_0803300
Hcol_000094900	DNA repair and recombination protein RAD54, putative	PF3D7_0803400
Hcol_000095000	conserved *Plasmodium* protein, unknown function	PF3D7_0803600
Hcol_000095100	tubulin gamma chain	PF3D7_0803700
Hcol_000095200	cactin homolog, putative	PF3D7_0804000
Hcol_000095300	methionine aminopeptidase 1c, putative	PF3D7_0804400
Hcol_000095400	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0804500
Hcol_000095500	peptidyl-prolyl cis-trans isomerase	PF3D7_0804800
Hcol_000095600	conserved *Plasmodium* protein, unknown function	PF3D7_0805100
Hcol_000095700	conserved *Plasmodium* protein, unknown function	PF3D7_0805300
Hcol_000095800	conserved *Plasmodium* protein, unknown function	PF3D7_0805500
Hcol_000095900	PAP2-like protein, putative	PF3D7_0805600
Hcol_000096000	serine/threonine protein kinase, FIKK family	PF3D7_0805700
Hcol_000096100	conserved *Plasmodium* protein, unknown function	PF3D7_0805800
Hcol_000096200	AAA family ATPase, putative	PF3D7_0806000
Hcol_000096300	conserved *Plasmodium* protein, unknown function	PF3D7_0806100
Hcol_000096400	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0806200
Hcol_000096500	ferlin, putative	PF3D7_0806300
Hcol_000096600	ferlin, putative	PF3D7_0806300
Hcol_000096700	glycosyltransferase family 28 protein, putative	PF3D7_0806400
Hcol_000096800	kinesin-like protein, putative	PF3D7_0806600
Hcol_000096900	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0806700
Hcol_000097000	vacuolar proton translocating ATPase subunit A, putative	PF3D7_0806800
Hcol_000097100	conserved *Plasmodium* protein, unknown function	PF3D7_0806900
Hcol_000097200	RNA helicase, putative	PF3D7_0807100
Hcol_000097300	conserved *Plasmodium* membrane protein, unknown function	PF3D7_0807200
Hcol_000097400	26S proteasome regulatory subunit RPN10, putative	PF3D7_0807800
Hcol_000097500	tyrosine–tRNA ligase	PF3D7_0807900
Hcol_000097600	conserved *Plasmodium* protein, unknown function	PF3D7_0808000
Hcol_000097700	AP-3 complex subunit delta, putative	PF3D7_0808100
Hcol_000097800	AP-3 complex subunit delta, putative	PF3D7_0808100
Hcol_000097900	plasmepsin X	PF3D7_0808200
Hcol_000098000	ubiquitin regulatory protein, putative	PF3D7_0808300
Hcol_000098100	coatomer subunit epsilon, putative	PF3D7_0808400
Hcol_000098200	asparagine-rich antigen Pfa55-14	PF3D7_0809200
Hcol_000098300	peptidase family C50, putative	PF3D7_0809600
Hcol_000098400	peptidase family C50, putative	PF3D7_0809600
Hcol_000098500	RuvB-like helicase 1	PF3D7_0809700
Hcol_000098600	JmjC domain containing protein	PF3D7_0809900
Hcol_000098700	ABC1 family, putative	PF3D7_0810200
Hcol_000098800	protein phosphatase PPM5, putative	PF3D7_0810300
Hcol_000098900	RNA helicase, putative	PF3D7_0810600
Hcol_000099000	hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase	PF3D7_0810800
Hcol_000099100	hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase	PF3D7_0810800
Hcol_000099200	conserved *Plasmodium* protein, unknown function	PF3D7_0810900
Hcol_000099300		PF3D7_0811000
Hcol_000099400	ER membrane protein complex subunit 1, putative	PF3D7_0811200
Hcol_000099500	CCR4-associated factor 1	PF3D7_0811300
Hcol_000099600	conserved protein, unknown function	PF3D7_0811400
Hcol_000099700	conserved *Plasmodium* protein, unknown function	PF3D7_0811600
Hcol_000099800	conserved *Plasmodium* protein, unknown function	PF3D7_0811700
Hcol_000099900	conserved *Plasmodium* protein, unknown function	PF3D7_0811800
Hcol_000100000	RNA-binding protein, putative	PF3D7_0811900

Table 8-2.: Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100000100	hypothetical protein
Hcol_100000200	TCP-1/cpn60 chaperonin family, putative
Hcol_100000300	hypothetical protein
Hcol_100000400	hypothetical protein
Hcol_100000500	hypothetical protein
Hcol_100000600	hypothetical protein
Hcol_100000700	hypothetical protein
Hcol_100000800	hypothetical protein
Hcol_100000900	hypothetical protein
Hcol_100001000	hypothetical protein
Hcol_100001100	hypothetical protein
Hcol_100001200	Transcription factor S-II (TFIIS), putative
Hcol_100001300	hypothetical protein
Hcol_100001400	Vacuolar sorting protein 39 domain 2, putative
Hcol_100001500	hypothetical protein
Hcol_100001600	hypothetical protein
Hcol_100001700	hypothetical protein
Hcol_100001800	GcpE protein, putative
Hcol_100001900	hypothetical protein
Hcol_100002000	hypothetical protein
Hcol_100002100	hypothetical protein
Hcol_100002200	ASF1 like histone chaperone, putative
Hcol_100002300	hypothetical protein
Hcol_100002400	FAD binding domain of DNA photolyase, putative
Hcol_100002500	hypothetical protein
Hcol_100002600	hypothetical protein
Hcol_100002700	hypothetical protein
Hcol_100002800	hypothetical protein
Hcol_100002900	hypothetical protein
Hcol_100003000	hypothetical protein
Hcol_100003100	hypothetical protein
Hcol_100003200	hypothetical protein
Hcol_100003300	hypothetical protein
Hcol_100003400	hypothetical protein
Hcol_100003500	hypothetical protein
Hcol_100003600	Reverse transcriptase (RNA-dependent DNA polymerase), putative
Hcol_100003700	Reverse transcriptase (RNA-dependent DNA polymerase), putative
Hcol_100003800	hypothetical protein
Hcol_100003900	hypothetical protein
Hcol_100004000	hypothetical protein
Hcol_100004100	hypothetical protein
Hcol_100004200	hypothetical protein
Hcol_100004300	hypothetical protein
Hcol_100004400	Coproporphyrinogen III oxidase, putative
Hcol_100004500	hypothetical protein
Hcol_100004600	Thrombospondin type 1 domain containing protein, putative
Hcol_100004700	Thrombospondin type 1 domain containing protein, putative
Hcol_100004800	hypothetical protein
Hcol_100004900	hypothetical protein
Hcol_100005000	hypothetical protein
Hcol_100005100	hypothetical protein
Hcol_100005200	hypothetical protein
Hcol_100005300	AAA domain/Viral (Superfamily 1) RNA helicase, putative
Hcol_100005400	AAA domain containing protein, putative
Hcol_100005500	hypothetical protein
Hcol_100005600	hypothetical protein
Hcol_100005700	hypothetical protein
Hcol_100005800	hypothetical protein
Hcol_100005900	hypothetical protein
Hcol_100006000	ATPase family associated with various cellular activities (AAA), putative
Hcol_100006100	hypothetical protein
Hcol_100006200	hypothetical protein
Hcol_100006300	hypothetical protein
Hcol_100006400	hypothetical protein
Hcol_100006500	hypothetical protein
Hcol_100006600	hypothetical protein
Hcol_100006700	hypothetical protein
Hcol_100006800	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100006900	ATPase family associated with various cellular activities (AAA)/AAA domain (dynein-related subfamily)/Sigma-54 interaction domain/AAA domain (Cdc48 subfamily)/C-terminal, D2-small domain, of ClpB protein, putative
Hcol_100007000	hypothetical protein
Hcol_100007100	hypothetical protein
Hcol_100007200	hypothetical protein
Hcol_100007300	hypothetical protein
Hcol_100007400	hypothetical protein
Hcol_100007500	hypothetical protein
Hcol_100007600	E1-E2 ATPase, putative
Hcol_100007700	Sel1 repeat, putative
Hcol_100007800	hypothetical protein
Hcol_100007900	hypothetical protein
Hcol_100008000	hypothetical protein
Hcol_100008100	hypothetical protein
Hcol_100008200	hypothetical protein
Hcol_100008300	hypothetical protein
Hcol_100008400	hypothetical protein
Hcol_100008500	hypothetical protein
Hcol_100008600	hypothetical protein
Hcol_100008700	BTB/POZ domain containing protein, putative
Hcol_100008800	ABC1 family/Lipopolysaccharide core biosynthesis protein (WaaY), putative
Hcol_100008900	SBDS protein C-terminal domain containing protein, putative
Hcol_100009000	hypothetical protein
Hcol_100009100	hypothetical protein
Hcol_100009200	hypothetical protein
Hcol_100009300	hypothetical protein
Hcol_100009400	hypothetical protein
Hcol_100009500	hypothetical protein
Hcol_100009600	hypothetical protein
Hcol_100009700	hypothetical protein
Hcol_100009800	hypothetical protein
Hcol_100009900	IQ calmodulin-binding motif containing protein, putative
Hcol_100010000	hypothetical protein
Hcol_100010100	hypothetical protein
Hcol_100010200	hypothetical protein
Hcol_100010300	hypothetical protein
Hcol_100010400	Alcohol dehydrogenase GroES-like domain containing protein, putative
Hcol_100010500	hypothetical protein
Hcol_100010600	Endonuclease/Exonuclease/phosphatase family, putative
Hcol_100010700	hypothetical protein
Hcol_100010800	Plasmodium falciparum domain of unknown function (CPW_WPC), putative
Hcol_100010900	hypothetical protein
Hcol_100011000	hypothetical protein
Hcol_100011100	Protein kinase domain/Protein tyrosine kinase, putative
Hcol_100011200	hypothetical protein
Hcol_100011300	hypothetical protein
Hcol_100011400	hypothetical protein
Hcol_100011500	hypothetical protein
Hcol_100011600	hypothetical protein
Hcol_100011700	hypothetical protein
Hcol_100011800	hypothetical protein
Hcol_100011900	hypothetical protein
Hcol_100012000	hypothetical protein
Hcol_100012100	hypothetical protein
Hcol_100012200	hypothetical protein
Hcol_100012300	hypothetical protein
Hcol_100012400	hypothetical protein
Hcol_100012500	hypothetical protein
Hcol_100012600	hypothetical protein
Hcol_100012700	hypothetical protein
Hcol_100012800	hypothetical protein
Hcol_100012900	hypothetical protein
Hcol_100013000	hypothetical protein
Hcol_100013100	hypothetical protein
Hcol_100013200	AAA domain/Part of AAA domain containing protein, putative
Hcol_100013300	hypothetical protein
Hcol_100013400	hypothetical protein
Hcol_100013500	hypothetical protein
Hcol_100013600	hypothetical protein
Hcol_100013700	hypothetical protein
Hcol_100013800	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100013900	hypothetical protein
Hcol_100014000	hypothetical protein
Hcol_100014100	Ubiquitin carboxyl-terminal hydrolase, putative
Hcol_100014200	hypothetical protein
Hcol_100014300	hypothetical protein
Hcol_100014400	hypothetical protein
Hcol_100014500	hypothetical protein
Hcol_100014600	hypothetical protein
Hcol_100014700	hypothetical protein
Hcol_100014800	hypothetical protein
Hcol_100014900	hypothetical protein
Hcol_100015000	hypothetical protein
Hcol_100015100	U-box domain containing protein, putative
Hcol_100015200	U-box domain containing protein, putative
Hcol_100015300	hypothetical protein
Hcol_100015400	hypothetical protein
Hcol_100015500	Ubiquitin elongating factor core, putative
Hcol_100015600	Ubiquitin elongating factor core, putative
Hcol_100015700	hypothetical protein
Hcol_100015800	hypothetical protein
Hcol_100015900	hypothetical protein
Hcol_100016000	hypothetical protein
Hcol_100016100	hypothetical protein
Hcol_100016200	hypothetical protein
Hcol_100016300	hypothetical protein
Hcol_100016400	Hsp90 protein, putative
Hcol_100016500	hypothetical protein
Hcol_100016600	hypothetical protein
Hcol_100016700	hypothetical protein
Hcol_100016800	hypothetical protein
Hcol_100016900	hypothetical protein
Hcol_100017000	Histone acetyl transferase HAT1 N-terminus, putative
Hcol_100017100	hypothetical protein
Hcol_100017200	Zinc-binding dehydrogenase, putative
Hcol_100017300	hypothetical protein
Hcol_100017400	hypothetical protein
Hcol_100017500	hypothetical protein
Hcol_100017600	hypothetical protein
Hcol_100017700	hypothetical protein
Hcol_100017800	hypothetical protein
Hcol_100017900	hypothetical protein
Hcol_100018000	hypothetical protein
Hcol_100018100	hypothetical protein
Hcol_100018200	hypothetical protein
Hcol_100018300	hypothetical protein
Hcol_100018400	ATPase MipZ/Anion-transporting ATPase/CobQ/CobB/ MinD/ParA nucleotide binding domain/ParA/MinD ATPase like, putative
Hcol_100018500	hypothetical protein
Hcol_100018600	hypothetical protein
Hcol_100018700	hypothetical protein
Hcol_100018800	hypothetical protein
Hcol_100018900	hypothetical protein
Hcol_100019000	hypothetical protein
Hcol_100019100	hypothetical protein
Hcol_100019200	RNA binding activity-knot of a chromodomain, putative
Hcol_100019300	Starch synthase catalytic domain containing protein, putative
Hcol_100019400	hypothetical protein
Hcol_100019500	RING-type zinc-finger containing protein, putative
Hcol_100019600	hypothetical protein
Hcol_100019700	hypothetical protein
Hcol_100019800	Helicase conserved C-terminal domain containing protein, putative
Hcol_100019900	hypothetical protein
Hcol_100020000	hypothetical protein
Hcol_100020100	Protein tyrosine kinase/Protein kinase domain containing protein, putative
Hcol_100020200	hypothetical protein
Hcol_100020300	hypothetical protein
Hcol_100020400	hypothetical protein
Hcol_100020500	SNF2 family N-terminal domain/Helicase conserved C-terminal domain/SLIDE, putative
Hcol_100020600	SNF2 family N-terminal domain/Helicase conserved C-terminal domain/SLIDE, putative
Hcol_100020700	COP9 signalosome, subunit CSN8/PCI domain containing protein, putative

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100020800	hypothetical protein
Hcol_100020900	hypothetical protein
Hcol_100021000	hypothetical protein
Hcol_100021100	Thymidylate synthase, putative
Hcol_100021200	hypothetical protein
Hcol_100021300	hypothetical protein
Hcol_100021400	ENTH domain containing protein, putative
Hcol_100021500	hypothetical protein
Hcol_100021600	Cathepsin C exclusion domain containing protein, putative
Hcol_100021700	hypothetical protein
Hcol_100021800	hypothetical protein
Hcol_100021900	hypothetical protein
Hcol_100022000	hypothetical protein
Hcol_100022100	hypothetical protein
Hcol_100022200	hypothetical protein
Hcol_100022300	hypothetical protein
Hcol_100022400	hypothetical protein
Hcol_100022500	hypothetical protein
Hcol_100022600	hypothetical protein
Hcol_100022700	hypothetical protein
Hcol_100022800	hypothetical protein
Hcol_100022900	hypothetical protein
Hcol_100023000	hypothetical protein
Hcol_100023100	hypothetical protein
Hcol_100023200	hypothetical protein
Hcol_100023300	hypothetical protein
Hcol_100023400	hypothetical protein
Hcol_100023500	hypothetical protein
Hcol_100023600	hypothetical protein
Hcol_100023700	hypothetical protein
Hcol_100023800	hypothetical protein
Hcol_100023900	Spc97 / Spc98 family, putative
Hcol_100024000	Spc97 / Spc98 family, putative
Hcol_100024100	hypothetical protein
Hcol_100024200	Ankyrin repeats (3 copies)/Ankyrin repeats (many copies)/Ankyrin repeat, putative
Hcol_100024300	hypothetical protein
Hcol_100024400	hypothetical protein
Hcol_100024500	hypothetical protein
Hcol_100024600	hypothetical protein
Hcol_100024700	hypothetical protein
Hcol_100024800	hypothetical protein
Hcol_100024900	Helix-hairpin-helix motif containing protein, putative
Hcol_100025000	hypothetical protein
Hcol_100025100	hypothetical protein
Hcol_100025200	hypothetical protein
Hcol_100025300	hypothetical protein
Hcol_100025400	hypothetical protein
Hcol_100025500	hypothetical protein
Hcol_100025600	hypothetical protein
Hcol_100025700	hypothetical protein
Hcol_100025800	hypothetical protein
Hcol_100025900	hypothetical protein
Hcol_100026000	hypothetical protein
Hcol_100026100	hypothetical protein
Hcol_100026200	hypothetical protein
Hcol_100026300	hypothetical protein
Hcol_100026400	hypothetical protein
Hcol_100026500	LysM domain containing protein, putative
Hcol_100026600	hypothetical protein
Hcol_100026700	Transketolase, thiamine diphosphate binding domain containing protein, putative
Hcol_100026800	SNF2 family N-terminal domain/Helicase conserved C-terminal domain containing protein, putative
Hcol_100026900	hypothetical protein
Hcol_100027000	hypothetical protein
Hcol_100027100	WD domain, G-beta repeat, putative
Hcol_100027200	WD domain, G-beta repeat, putative
Hcol_100027300	hypothetical protein
Hcol_100027400	ATPase family associated with various cellular activities (AAA), putative
Hcol_100027500	hypothetical protein
Hcol_100027600	hypothetical protein
Hcol_100027700	hypothetical protein
Hcol_100027800	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100027900	hypothetical protein
Hcol_100028000	hypothetical protein
Hcol_100028100	hypothetical protein
Hcol_100028200	hypothetical protein
Hcol_100028300	hypothetical protein
Hcol_100028400	hypothetical protein
Hcol_100028500	hypothetical protein
Hcol_100028600	hypothetical protein
Hcol_100028700	hypothetical protein
Hcol_100028800	Putative diphthamide synthesis protein, putative
Hcol_100028900	Sexual stage antigen s48/45 domain containing protein, putative
Hcol_100029000	Sexual stage antigen s48/45 domain containing protein, putative
Hcol_100029100	EF-hand domain pair/EF-hand domain containing protein, putative
Hcol_100029200	hypothetical protein
Hcol_100029300	SNF2 family N-terminal domain containing protein, putative
Hcol_100029400	hypothetical protein
Hcol_100029500	hypothetical protein
Hcol_100029600	hypothetical protein
Hcol_100029700	hypothetical protein
Hcol_100029800	hypothetical protein
Hcol_100029900	hypothetical protein
Hcol_100030000	hypothetical protein
Hcol_100030100	hypothetical protein
Hcol_100030200	Protein kinase domain/Protein tyrosine kinase, putative
Hcol_100030300	Protein kinase domain/Protein tyrosine kinase, putative
Hcol_100030400	hypothetical protein
Hcol_100030500	hypothetical protein
Hcol_100030600	hypothetical protein
Hcol_100030700	hypothetical protein
Hcol_100030800	hypothetical protein
Hcol_100030900	hypothetical protein
Hcol_100031000	hypothetical protein
Hcol_100031100	hypothetical protein
Hcol_100031200	hypothetical protein
Hcol_100031300	hypothetical protein
Hcol_100031400	hypothetical protein
Hcol_100031500	hypothetical protein
Hcol_100031600	hypothetical protein
Hcol_100031700	hypothetical protein
Hcol_100031800	hypothetical protein
Hcol_100031900	hypothetical protein
Hcol_100032000	hypothetical protein
Hcol_100032100	hypothetical protein
Hcol_100032200	hypothetical protein
Hcol_100032300	PHD-zinc-finger like domain/PHD-like zinc-binding domain containing protein
Hcol_100032400	hypothetical protein
Hcol_100032500	hypothetical protein
Hcol_100032600	hypothetical protein
Hcol_100032700	hypothetical protein
Hcol_100032800	hypothetical protein
Hcol_100032900	hypothetical protein
Hcol_100033000	hypothetical protein
Hcol_100033100	hypothetical protein
Hcol_100033200	hypothetical protein
Hcol_100033300	hypothetical protein
Hcol_100033400	hypothetical protein
Hcol_100033500	hypothetical protein
Hcol_100033600	hypothetical protein
Hcol_100033700	hypothetical protein
Hcol_100033800	hypothetical protein
Hcol_100033900	hypothetical protein
Hcol_100034000	hypothetical protein
Hcol_100034100	hypothetical protein
Hcol_100034200	hypothetical protein
Hcol_100034300	hypothetical protein
Hcol_100034400	hypothetical protein
Hcol_100034500	hypothetical protein
Hcol_100034600	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain) /RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
Hcol_100034700	hypothetical protein
Hcol_100034800	hypothetical protein
Hcol_100034900	hypothetical protein
Hcol_100035000	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100035100	hypothetical protein
Hcol_100035200	hypothetical protein
Hcol_100035300	AAA domain/Part of AAA domain containing protein, putative
Hcol_100035400	hypothetical protein
Hcol_100035500	hypothetical protein
Hcol_100035600	hypothetical protein
Hcol_100035700	AAA domain/Part of AAA domain/Uncharacterized conserved protein (DUF2075), putative
Hcol_100035800	hypothetical protein
Hcol_100035900	hypothetical protein
Hcol_100036000	hypothetical protein
Hcol_100036100	hypothetical protein
Hcol_100036200	hypothetical protein
Hcol_100036300	hypothetical protein
Hcol_100036400	hypothetical protein
Hcol_100036500	hypothetical protein
Hcol_100036600	hypothetical protein
Hcol_100036700	hypothetical protein
Hcol_100036800	hypothetical protein
Hcol_100036900	hypothetical protein
Hcol_100037000	hypothetical protein
Hcol_100037100	Protein kinase domain containing protein, putative
Hcol_100037200	hypothetical protein
Hcol_100037300	Mitochondrial degradasome RNA helicase subunit C terminal, putative
Hcol_100037400	hypothetical protein
Hcol_100037500	hypothetical protein
Hcol_100037600	hypothetical protein
Hcol_100037700	hypothetical protein
Hcol_100037800	hypothetical protein
Hcol_100037900	hypothetical protein
Hcol_100038000	hypothetical protein
Hcol_100038100	hypothetical protein
Hcol_100038200	HMG (high mobility group) box, putative
Hcol_100038300	hypothetical protein
Hcol_100038400	Protein phosphatase 2C, putative
Hcol_100038500	hypothetical protein
Hcol_100038600	hypothetical protein
Hcol_100038700	hypothetical protein
Hcol_100038800	hypothetical protein
Hcol_100038900	hypothetical protein
Hcol_100039000	hypothetical protein
Hcol_100039100	hypothetical protein
Hcol_100039200	hypothetical protein
Hcol_100039300	hypothetical protein
Hcol_100039400	hypothetical protein
Hcol_100039500	hypothetical protein
Hcol_100039600	SPX domain containing protein, putative
Hcol_100039700	hypothetical protein
Hcol_100039800	hypothetical protein
Hcol_100039900	hypothetical protein
Hcol_100040000	hypothetical protein
Hcol_100040100	hypothetical protein
Hcol_100040200	hypothetical protein
Hcol_100040300	hypothetical protein
Hcol_100040400	hypothetical protein
Hcol_100040500	hypothetical protein
Hcol_100040600	hypothetical protein
Hcol_100040700	hypothetical protein
Hcol_100040800	hypothetical protein
Hcol_100040900	hypothetical protein
Hcol_100041000	hypothetical protein
Hcol_100041100	lipin, N-terminal conserved region containing protein, putative
Hcol_100041200	Sexual stage antigen s48/45 domain containing protein, putative
Hcol_100041300	hypothetical protein
Hcol_100041400	hypothetical protein
Hcol_100041500	hypothetical protein
Hcol_100041600	hypothetical protein
Hcol_100041700	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)/ RNA recognition motif (a.k.a. RRM, RBD, or RNP domain), putative
Hcol_100041800	hypothetical protein
Hcol_100041900	hypothetical protein
Hcol_100042000	hypothetical protein
Hcol_100042100	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100042200	NUC153 domain containing protein, putative
Hcol_100042300	hypothetical protein
Hcol_100042400	GMP synthase C terminal domain containing protein, putative
Hcol_100042500	hypothetical protein
Hcol_100042600	hypothetical protein
Hcol_100042700	hypothetical protein
Hcol_100042800	hypothetical protein
Hcol_100042900	hypothetical protein
Hcol_100043000	hypothetical protein
Hcol_100043100	hypothetical protein
Hcol_100043200	hypothetical protein
Hcol_100043300	hypothetical protein
Hcol_100043400	hypothetical protein
Hcol_100043500	hypothetical protein
Hcol_100043600	hypothetical protein
Hcol_100043700	hypothetical protein
Hcol_100043800	hypothetical protein
Hcol_100043900	hypothetical protein
Hcol_100044000	hypothetical protein
Hcol_100044100	WD domain, G-beta repeat, putative
Hcol_100044200	hypothetical protein
Hcol_100044300	hypothetical protein
Hcol_100044400	Sec7 domain containing protein, putative
Hcol_100044500	hypothetical protein
Hcol_100044600	hypothetical protein
Hcol_100044700	hypothetical protein
Hcol_100044800	hypothetical protein
Hcol_100044900	hypothetical protein
Hcol_100045000	hypothetical protein
Hcol_100045100	hypothetical protein
Hcol_100045200	hypothetical protein
Hcol_100045300	hypothetical protein
Hcol_100045400	hypothetical protein
Hcol_100045500	hypothetical protein
Hcol_100045600	hypothetical protein
Hcol_100045700	hypothetical protein
Hcol_100045800	hypothetical protein
Hcol_100045900	hypothetical protein
Hcol_100046000	hypothetical protein
Hcol_100046100	hypothetical protein
Hcol_100046200	Reticulon, putative
Hcol_100046300	hypothetical protein
Hcol_100046400	hypothetical protein
Hcol_100046500	hypothetical protein
Hcol_100046600	hypothetical protein
Hcol_100046700	hypothetical protein
Hcol_100046800	hypothetical protein
Hcol_100046900	hypothetical protein
Hcol_100047000	hypothetical protein
Hcol_100047100	Cytochrome c oxidase assembly protein CtaG/Cox11, putative
Hcol_100047200	hypothetical protein
Hcol_100047300	hypothetical protein
Hcol_100047400	hypothetical protein
Hcol_100047500	hypothetical protein
Hcol_100047600	hypothetical protein
Hcol_100047700	hypothetical protein
Hcol_100047800	von Willebrand factor type A domain containing protein, putative
Hcol_100047900	hypothetical protein
Hcol_100048000	hypothetical protein
Hcol_100048100	hypothetical protein
Hcol_100048200	hypothetical protein
Hcol_100048300	hypothetical protein
Hcol_100048400	hypothetical protein
Hcol_100048500	hypothetical protein
Hcol_100048600	hypothetical protein
Hcol_100048700	hypothetical protein
Hcol_100048800	hypothetical protein
Hcol_100048900	hypothetical protein
Hcol_100049000	hypothetical protein
Hcol_100049100	hypothetical protein
Hcol_100049200	SWIRM domain containing protein, putative
Hcol_100049300	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100049400	hypothetical protein
Hcol_100049500	hypothetical protein
Hcol_100049600	hypothetical protein
Hcol_100049700	hypothetical protein
Hcol_100049800	hypothetical protein
Hcol_100049900	hypothetical protein
Hcol_100050000	hypothetical protein
Hcol_100050100	hypothetical protein
Hcol_100050200	hypothetical protein
Hcol_100050300	hypothetical protein
Hcol_100050400	hypothetical protein
Hcol_100050500	hypothetical protein
Hcol_100050600	Thioredoxin-like, putative
Hcol_100050700	Ubiquitin carboxyl-terminal hydrolase, family 1, putative
Hcol_100050800	hypothetical protein
Hcol_100050900	hypothetical protein
Hcol_100051000	Myosin head (motor domain), putative
Hcol_100051100	hypothetical protein
Hcol_100051200	hypothetical protein
Hcol_100051300	hypothetical protein
Hcol_100051400	hypothetical protein
Hcol_100051500	hypothetical protein
Hcol_100051600	hypothetical protein
Hcol_100051700	Regulator of chromosome condensation (RCC1) repeat, putative
Hcol_100051800	hypothetical protein
Hcol_100051900	hypothetical protein
Hcol_100052000	hypothetical protein
Hcol_100052100	RNA polymerase Rpb1, domain 5, putative
Hcol_100052200	hypothetical protein
Hcol_100052300	hypothetical protein
Hcol_100052400	hypothetical protein
Hcol_100052500	Uncharacterized protein family UPF0054, putative
Hcol_100052600	Ribosomal L29 protein, putative
Hcol_100052700	hypothetical protein
Hcol_100052800	hypothetical protein
Hcol_100052900	hypothetical protein
Hcol_100053000	LNS2 (Lipin/Ned1/Smp2), putative
Hcol_100053100	hypothetical protein
Hcol_100053200	hypothetical protein
Hcol_100053300	hypothetical protein
Hcol_100053400	hypothetical protein
Hcol_100053500	hypothetical protein
Hcol_100053600	hypothetical protein
Hcol_100053700	hypothetical protein
Hcol_100053800	hypothetical protein
Hcol_100053900	Ubiquitin-2 like Rad60 SUMO-like/Ubiquitin family, putative
Hcol_100054000	Ras family, putative
Hcol_100054100	hypothetical protein
Hcol_100054200	hypothetical protein
Hcol_100054300	hypothetical protein
Hcol_100054400	Tubulin C-terminal domain containing protein, putative
Hcol_100054500	hypothetical protein
Hcol_100054600	hypothetical protein
Hcol_100054700	hypothetical protein
Hcol_100054800	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)/ RNA recognition motif (a.k.a. RRM, RBD, or RNP domain), putative
Hcol_100054900	hypothetical protein
Hcol_100055000	hypothetical protein
Hcol_100055100	hypothetical protein
Hcol_100055200	hypothetical protein
Hcol_100055300	hypothetical protein
Hcol_100055400	Condensin complex subunit 2, putative
Hcol_100055500	hypothetical protein
Hcol_100055600	hypothetical protein
Hcol_100055700	hypothetical protein
Hcol_100055800	hypothetical protein
Hcol_100055900	hypothetical protein
Hcol_100056000	Fumble, putative
Hcol_100056100	AMP-binding enzyme, putative
Hcol_100056200	hypothetical protein
Hcol_100056300	Spb1 C-terminal domain containing protein, putative
Hcol_100056400	hypothetical protein
Hcol_100056500	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100056600	hypothetical protein
Hcol_100056700	hypothetical protein
Hcol_100056800	hypothetical protein
Hcol_100056900	hypothetical protein
Hcol_100057000	hypothetical protein
Hcol_100057100	hypothetical protein
Hcol_100057200	hypothetical protein
Hcol_100057300	hypothetical protein
Hcol_100057400	hypothetical protein
Hcol_100057500	hypothetical protein
Hcol_100057600	hypothetical protein
Hcol_100057700	hypothetical protein
Hcol_100057800	hypothetical protein
Hcol_100057900	hypothetical protein
Hcol_100058000	hypothetical protein
Hcol_100058100	hypothetical protein
Hcol_100058200	Collagen triple helix repeat (20 copies), putative
Hcol_100058300	hypothetical protein
Hcol_100058400	hypothetical protein
Hcol_100058500	hypothetical protein
Hcol_100058600	hypothetical protein
Hcol_100058700	hypothetical protein
Hcol_100058800	hypothetical protein
Hcol_100058900	hypothetical protein
Hcol_100059000	hypothetical protein
Hcol_100059100	hypothetical protein
Hcol_100059200	Patatin-like phospholipase, putative
Hcol_100059300	hypothetical protein
Hcol_100059400	hypothetical protein
Hcol_100059500	hypothetical protein
Hcol_100059600	Glycosyl transferases group 1, putative
Hcol_100059700	hypothetical protein
Hcol_100059800	hypothetical protein
Hcol_100059900	hypothetical protein
Hcol_100060000	hypothetical protein
Hcol_100060100	hypothetical protein
Hcol_100060200	hypothetical protein
Hcol_100060300	hypothetical protein
Hcol_100060400	hypothetical protein
Hcol_100060500	hypothetical protein
Hcol_100060600	hypothetical protein
Hcol_100060700	hypothetical protein
Hcol_100060800	hypothetical protein
Hcol_100060900	hypothetical protein
Hcol_100061000	hypothetical protein
Hcol_100061100	hypothetical protein
Hcol_100061200	hypothetical protein
Hcol_100061300	hypothetical protein
Hcol_100061400	hypothetical protein
Hcol_100061500	hypothetical protein
Hcol_100061600	hypothetical protein
Hcol_100061700	hypothetical protein
Hcol_100061800	hypothetical protein
Hcol_100061900	hypothetical protein
Hcol_100062000	hypothetical protein
Hcol_100062100	hypothetical protein
Hcol_100062200	hypothetical protein
Hcol_100062300	hypothetical protein
Hcol_100062400	hypothetical protein
Hcol_100062500	hypothetical protein
Hcol_100062600	Sec23/Sec24 zinc finger containing protein, putative
Hcol_100062700	SNF2 family N-terminal domain/DEAD/DEAH box helicase/Helicase conserved C-terminal domain containing protein, putative
Hcol_100062800	hypothetical protein
Hcol_100062900	hypothetical protein
Hcol_100063000	hypothetical protein
Hcol_100063100	hypothetical protein
Hcol_100063200	hypothetical protein
Hcol_100063300	hypothetical protein
Hcol_100063400	Ubiquitin-2 like Rad60 SUMO-like/Ubiquitin family, putative
Hcol_100063500	hypothetical protein
Hcol_100063600	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100063700	hypothetical protein
Hcol_100063800	hypothetical protein
Hcol_100063900	hypothetical protein
Hcol_100064000	hypothetical protein
Hcol_100064100	hypothetical protein
Hcol_100064200	hypothetical protein
Hcol_100064300	hypothetical protein
Hcol_100064400	hypothetical protein
Hcol_100064500	hypothetical protein
Hcol_100064600	hypothetical protein
Hcol_100064700	Protein kinase domain/Protein tyrosine kinase, putative
Hcol_100064800	hypothetical protein
Hcol_100064900	hypothetical protein
Hcol_100065000	hypothetical protein
Hcol_100065100	Sodium:neurotransmitter symporter family, putative
Hcol_100065200	hypothetical protein
Hcol_100065300	hypothetical protein
Hcol_100065400	hypothetical protein
Hcol_100065500	hypothetical protein
Hcol_100065600	hypothetical protein
Hcol_100065700	hypothetical protein
Hcol_100065800	hypothetical protein
Hcol_100065900	hypothetical protein
Hcol_100066000	hypothetical protein
Hcol_100066100	hypothetical protein
Hcol_100066200	hypothetical protein
Hcol_100066300	Toprim-like/AAA domain containing protein, putative
Hcol_100066400	hypothetical protein
Hcol_100066500	hypothetical protein
Hcol_100066600	hypothetical protein
Hcol_100066700	hypothetical protein
Hcol_100066800	hypothetical protein
Hcol_100066900	hypothetical protein
Hcol_100067000	hypothetical protein
Hcol_100067100	hypothetical protein
Hcol_100067200	hypothetical protein
Hcol_100067300	hypothetical protein
Hcol_100067400	hypothetical protein
Hcol_100067500	hypothetical protein
Hcol_100067600	hypothetical protein
Hcol_100067700	hypothetical protein
Hcol_100067800	hypothetical protein
Hcol_100067900	hypothetical protein
Hcol_100068000	hypothetical protein
Hcol_100068100	hypothetical protein
Hcol_100068200	hypothetical protein
Hcol_100068300	hypothetical protein
Hcol_100068400	hypothetical protein
Hcol_100068500	hypothetical protein
Hcol_100068600	hypothetical protein
Hcol_100068700	hypothetical protein
Hcol_100068800	hypothetical protein
Hcol_100068900	hypothetical protein
Hcol_100069000	hypothetical protein
Hcol_100069100	hypothetical protein
Hcol_100069200	hypothetical protein
Hcol_100069300	hypothetical protein
Hcol_100069400	Protein phosphatase 2C, putative
Hcol_100069500	von Willebrand factor type A domain containing protein, putative
Hcol_100069600	hypothetical protein
Hcol_100069700	hypothetical protein
Hcol_100069800	hypothetical protein
Hcol_100069900	hypothetical protein
Hcol_100070000	hypothetical protein
Hcol_100070100	hypothetical protein
Hcol_100070200	hypothetical protein
Hcol_100070300	hypothetical protein
Hcol_100070400	Patched family, putative
Hcol_100070500	hypothetical protein
Hcol_100070600	hypothetical protein
Hcol_100070700	hypothetical protein
Hcol_100070800	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100070900	hypothetical protein
Hcol_100071000	hypothetical protein
Hcol_100071100	hypothetical protein
Hcol_100071200	hypothetical protein
Hcol_100071300	hypothetical protein
Hcol_100071400	hypothetical protein
Hcol_100071500	hypothetical protein
Hcol_100071600	Myb-like DNA-binding domain containing protein, putative
Hcol_100071700	hypothetical protein
Hcol_100071800	hypothetical protein
Hcol_100071900	hypothetical protein
Hcol_100072000	hypothetical protein
Hcol_100072100	hypothetical protein
Hcol_100072200	hypothetical protein
Hcol_100072300	hypothetical protein
Hcol_100072400	hypothetical protein
Hcol_100072500	hypothetical protein
Hcol_100072600	hypothetical protein
Hcol_100072700	hypothetical protein
Hcol_100072800	hypothetical protein
Hcol_100072900	hypothetical protein
Hcol_100073000	hypothetical protein
Hcol_100073100	hypothetical protein
Hcol_100073200	hypothetical protein
Hcol_100073300	hypothetical protein
Hcol_100073400	hypothetical protein
Hcol_100073500	hypothetical protein
Hcol_100073600	hypothetical protein
Hcol_100073700	hypothetical protein
Hcol_100073800	hypothetical protein
Hcol_100073900	hypothetical protein
Hcol_100074000	hypothetical protein
Hcol_100074100	hypothetical protein
Hcol_100074200	hypothetical protein
Hcol_100074300	hypothetical protein
Hcol_100074400	hypothetical protein
Hcol_100074500	hypothetical protein
Hcol_100074600	Pre-mRNA 3'-end-processing endonuclease polyadenylation factor C-term, putative
Hcol_100074700	hypothetical protein
Hcol_100074800	hypothetical protein
Hcol_100074900	Putative RNA methyltransferase, putative
Hcol_100075000	hypothetical protein
Hcol_100075100	hypothetical protein
Hcol_100075200	hypothetical protein
Hcol_100075300	hypothetical protein
Hcol_100075400	hypothetical protein
Hcol_100075500	hypothetical protein
Hcol_100075600	hypothetical protein
Hcol_100075700	Major Facilitator Superfamily, putative
Hcol_100075800	hypothetical protein
Hcol_100075900	hypothetical protein
Hcol_100076000	hypothetical protein
Hcol_100076100	hypothetical protein
Hcol_100076200	hypothetical protein
Hcol_100076300	hypothetical protein
Hcol_100076400	hypothetical protein
Hcol_100076500	hypothetical protein
Hcol_100076600	hypothetical protein
Hcol_100076700	hypothetical protein
Hcol_100076800	hypothetical protein
Hcol_100076900	hypothetical protein
Hcol_100077000	hypothetical protein
Hcol_100077100	hypothetical protein
Hcol_100077200	hypothetical protein
Hcol_100077300	hypothetical protein
Hcol_100077400	hypothetical protein
Hcol_100077500	hypothetical protein
Hcol_100077600	hypothetical protein
Hcol_100077700	hypothetical protein
Hcol_100077800	hypothetical protein
Hcol_100077900	hypothetical protein
Hcol_100078000	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100078100	hypothetical protein
Hcol_100078200	hypothetical protein
Hcol_100078300	hypothetical protein
Hcol_100078400	hypothetical protein
Hcol_100078500	hypothetical protein
Hcol_100078600	N-terminal region of Chorein, a TM vesicle-mediated sorter, putative
Hcol_100078700	hypothetical protein
Hcol_100078800	hypothetical protein
Hcol_100078900	hypothetical protein
Hcol_100079000	hypothetical protein
Hcol_100079100	hypothetical protein
Hcol_100079200	hypothetical protein
Hcol_100079300	RNA polymerase II-binding domain., putative
Hcol_100079400	hypothetical protein
Hcol_100079500	hypothetical protein
Hcol_100079600	hypothetical protein
Hcol_100079700	hypothetical protein
Hcol_100079800	hypothetical protein
Hcol_100079900	hypothetical protein
Hcol_100080000	hypothetical protein
Hcol_100080100	hypothetical protein
Hcol_100080200	hypothetical protein
Hcol_100080300	hypothetical protein
Hcol_100080400	hypothetical protein
Hcol_100080500	hypothetical protein
Hcol_100080600	hypothetical protein
Hcol_100080700	hypothetical protein
Hcol_100080800	hypothetical protein
Hcol_100080900	hypothetical protein
Hcol_100081000	hypothetical protein
Hcol_100081100	hypothetical protein
Hcol_100081200	hypothetical protein
Hcol_100081300	hypothetical protein
Hcol_100081400	hypothetical protein
Hcol_100081500	hypothetical protein
Hcol_100081600	hypothetical protein
Hcol_100081700	hypothetical protein
Hcol_100081800	hypothetical protein
Hcol_100081900	hypothetical protein
Hcol_100082000	Protein-tyrosine phosphatase/Dual specificity phosphatase, catalytic domain containing protein, putative
Hcol_100082100	hypothetical protein
Hcol_100082200	hypothetical protein
Hcol_100082300	hypothetical protein
Hcol_100082400	hypothetical protein
Hcol_100082500	hypothetical protein
Hcol_100082600	hypothetical protein
Hcol_100082700	hypothetical protein
Hcol_100082800	hypothetical protein
Hcol_100082900	hypothetical protein
Hcol_100083000	hypothetical protein
Hcol_100083100	hypothetical protein
Hcol_100083200	hypothetical protein
Hcol_100083300	hypothetical protein
Hcol_100083400	hypothetical protein
Hcol_100083500	hypothetical protein
Hcol_100083600	hypothetical protein
Hcol_100083700	Protein kinase domain/Protein tyrosine kinase, putative
Hcol_100083800	hypothetical protein
Hcol_100083900	hypothetical protein
Hcol_100084000	hypothetical protein
Hcol_100084100	hypothetical protein
Hcol_100084200	hypothetical protein
Hcol_100084300	hypothetical protein
Hcol_100084400	hypothetical protein
Hcol_100084500	hypothetical protein
Hcol_100084600	hypothetical protein
Hcol_100084700	hypothetical protein
Hcol_100084800	hypothetical protein
Hcol_100084900	hypothetical protein
Hcol_100085000	hypothetical protein
Hcol_100085100	hypothetical protein
Hcol_100085200	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100085300	hypothetical protein
Hcol_100085400	hypothetical protein
Hcol_100085500	hypothetical protein
Hcol_100085600	hypothetical protein
Hcol_100085700	hypothetical protein
Hcol_100085800	hypothetical protein
Hcol_100085900	hypothetical protein
Hcol_100086000	hypothetical protein
Hcol_100086100	hypothetical protein
Hcol_100086200	hypothetical protein
Hcol_100086300	hypothetical protein
Hcol_100086400	hypothetical protein
Hcol_100086500	hypothetical protein
Hcol_100086600	hypothetical protein
Hcol_100086700	hypothetical protein
Hcol_100086800	Nucleotidyltransferase domain/Cid1 family poly A polymerase, putative
Hcol_100086900	hypothetical protein
Hcol_100087000	hypothetical protein
Hcol_100087100	Thrombospondin type 1 domain containing protein, putative
Hcol_100087200	hypothetical protein
Hcol_100087300	hypothetical protein
Hcol_100087400	hypothetical protein
Hcol_100087500	hypothetical protein
Hcol_100087600	hypothetical protein
Hcol_100087700	hypothetical protein
Hcol_100087800	hypothetical protein
Hcol_100087900	hypothetical protein
Hcol_100088000	WD domain, G-beta repeat, putative
Hcol_100088100	hypothetical protein
Hcol_100088200	hypothetical protein
Hcol_100088300	hypothetical protein
Hcol_100088400	hypothetical protein
Hcol_100088500	hypothetical protein
Hcol_100088600	hypothetical protein
Hcol_100088700	hypothetical protein
Hcol_100088800	PHD-like zinc-binding domain containing protein, putative
Hcol_100088900	hypothetical protein
Hcol_100089000	hypothetical protein
Hcol_100089100	hypothetical protein
Hcol_100089200	hypothetical protein
Hcol_100089300	hypothetical protein
Hcol_100089400	hypothetical protein
Hcol_100089500	hypothetical protein
Hcol_100089600	hypothetical protein
Hcol_100089700	hypothetical protein
Hcol_100089800	hypothetical protein
Hcol_100089900	hypothetical protein
Hcol_100090000	hypothetical protein
Hcol_100090100	hypothetical protein
Hcol_100090200	hypothetical protein
Hcol_100090300	hypothetical protein
Hcol_100090400	hypothetical protein
Hcol_100090500	hypothetical protein
Hcol_100090600	hypothetical protein
Hcol_100090700	hypothetical protein
Hcol_100090800	hypothetical protein
Hcol_100090900	hypothetical protein
Hcol_100091000	hypothetical protein
Hcol_100091100	hypothetical protein
Hcol_100091200	hypothetical protein
Hcol_100091300	hypothetical protein
Hcol_100091400	hypothetical protein
Hcol_100091500	hypothetical protein
Hcol_100091600	hypothetical protein
Hcol_100091700	hypothetical protein
Hcol_100091800	hypothetical protein
Hcol_100091900	hypothetical protein
Hcol_100092000	hypothetical protein
Hcol_100092100	hypothetical protein
Hcol_100092200	hypothetical protein
Hcol_100092300	hypothetical protein
Hcol_100092400	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100092500	Papain family cysteine protease, putative
Hcol_100092600	hypothetical protein
Hcol_100092700	hypothetical protein
Hcol_100092800	SacI homology domain containing protein, putative
Hcol_100092900	hypothetical protein
Hcol_100093000	hypothetical protein
Hcol_100093100	hypothetical protein
Hcol_100093200	hypothetical protein
Hcol_100093300	hypothetical protein
Hcol_100093400	hypothetical protein
Hcol_100093500	hypothetical protein
Hcol_100093600	hypothetical protein
Hcol_100093700	hypothetical protein
Hcol_100093800	hypothetical protein
Hcol_100093900	hypothetical protein
Hcol_100094000	hypothetical protein
Hcol_100094100	hypothetical protein
Hcol_100094200	hypothetical protein
Hcol_100094300	Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I/Phosphoglucomutase/ phosphomannomutase, alpha/beta/alpha domain II/Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III, putative
Hcol_100094400	hypothetical protein
Hcol_100094500	hypothetical protein
Hcol_100094600	hypothetical protein
Hcol_100094700	hypothetical protein
Hcol_100094800	hypothetical protein
Hcol_100094900	hypothetical protein
Hcol_100095000	hypothetical protein
Hcol_100095100	hypothetical protein
Hcol_100095200	hypothetical protein
Hcol_100095300	PHD-like zinc-binding domain/ PHD-zinc-finger like domain containing protein, putative
Hcol_100095400	hypothetical protein
Hcol_100095500	hypothetical protein
Hcol_100095600	hypothetical protein
Hcol_100095700	hypothetical protein
Hcol_100095800	hypothetical protein
Hcol_100095900	Poly(A) polymerase central domain/ Poly(A) polymerase predicted RNA binding domain containing protein, putative
Hcol_100096000	hypothetical protein
Hcol_100096100	hypothetical protein
Hcol_100096200	hypothetical protein
Hcol_100096300	hypothetical protein
Hcol_100096400	hypothetical protein
Hcol_100096500	hypothetical protein
Hcol_100096600	von Willebrand factor type A domain containing protein, putative
Hcol_100096700	hypothetical protein
Hcol_100096800	hypothetical protein
Hcol_100096900	hypothetical protein
Hcol_100097000	hypothetical protein
Hcol_100097100	hypothetical protein
Hcol_100097200	hypothetical protein
Hcol_100097300	hypothetical protein
Hcol_100097400	hypothetical protein
Hcol_100097500	hypothetical protein
Hcol_100097600	hypothetical protein
Hcol_100097700	hypothetical protein
Hcol_100097800	hypothetical protein
Hcol_100097900	hypothetical protein
Hcol_100098000	hypothetical protein
Hcol_100098100	hypothetical protein
Hcol_100098200	hypothetical protein
Hcol_100098300	hypothetical protein
Hcol_100098400	hypothetical protein
Hcol_100098500	hypothetical protein
Hcol_100098600	hypothetical protein
Hcol_100098700	hypothetical protein
Hcol_100098800	hypothetical protein
Hcol_100098900	hypothetical protein

Table 8-2 Genes summary of *H. columbae*. Singletons genes for *H. columbae* genome

Gene ID	Product
Hcol_100099000	hypothetical protein
Hcol_100099100	hypothetical protein
Hcol_100099200	hypothetical protein
Hcol_100099300	hypothetical protein
Hcol_100099400	hypothetical protein

Seqid	Source	Type	Start	End	Score	Strand	Phase	Attributes
seq1584	LTRharvest	repeat_region	1917	4336	.	+	.	ID=repeat_region1
seq1584	LTRharvest	target_site_duplication	1917	1920	.	+	.	Parent=repeat_region1
seq1584	LTRharvest	LTR_retrotransposon	1921	4332	.	+	.	ID=LTR_retrotransposon1; Parent=repeat_region1; ltr_similarity=87.10;seq_number=1584
seq1584	LTRharvest	long_terminal_repeat	1921	2075	.	+	.	Parent=LTR_retrotransposon1
seq1584	LTRharvest	long_terminal_repeat	4184	4332	.	+	.	Parent=LTR_retrotransposon1
seq1584	LTRdigest	RR_tract	4186	4198	.	+	.	Parent=LTR_retrotransposon1
seq1584	LTRharvest	target_site_duplication	4333	4336	.	+	.	Parent=repeat_region1
###								
seq1851	LTRharvest	repeat_region	115	1752	.	+	.	ID=repeat_region2
seq1851	LTRharvest	target_site_duplication	115	118	.	+	.	Parent=repeat_region2
seq1851	LTRharvest	LTR_retrotransposon	119	1748	.	+	.	ID=LTR_retrotransposon2; Parent=repeat_region2; ltr_similarity=90.67;seq_number=1851
seq1851	LTRharvest	long_terminal_repeat	119	311	.	+	.	Parent=LTR_retrotransposon2
seq1851	LTRharvest	long_terminal_repeat	1559	1748	.	+	.	Parent=LTR_retrotransposon2
seq1851	LTRdigest	RR_tract	1579	1586	.	+	.	Parent=LTR_retrotransposon2
seq1851	LTRharvest	target_site_duplication	1749	1752	.	+	.	Parent=repeat_region2
###								
seq1897	LTRharvest	repeat_region	50	2034	.	+	.	ID=repeat_region3
seq1897	LTRharvest	target_site_duplication	50	53	.	+	.	Parent=repeat_region3
seq1897	LTRharvest	LTR_retrotransposon	54	2030	.	+	.	ID=LTR_retrotransposon3; Parent=repeat_region3; ltr_similarity=85.48;seq_number=1897
seq1897	LTRharvest	long_terminal_repeat	54	177	.	+	.	Parent=LTR_retrotransposon3
seq1897	LTRdigest	RR_tract	1898	1905	.	+	.	Parent=LTR_retrotransposon3
seq1897	LTRharvest	long_terminal_repeat	1916	2030	.	+	.	Parent=LTR_retrotransposon3
seq1897	LTRharvest	target_site_duplication	2031	2034	.	+	.	Parent=repeat_region3
###								
seq1926	LTRharvest	repeat_region	2816	5834	.	-	.	ID=repeat_region4
seq1926	LTRharvest	target_site_duplication	2816	2819	.	-	.	Parent=repeat_region4
seq1926	LTRharvest	LTR_retrotransposon	2820	5830	.	-	.	ID=LTR_retrotransposon4; Parent=repeat_region4; ltr_similarity=85.78;seq_number=1926
seq1926	LTRharvest	long_terminal_repeat	2820	3031	.	-	.	Parent=LTR_retrotransposon4
seq1926	LTRdigest	RR_tract	3019	3036	.	-	.	Parent=LTR_retrotransposon4
seq1926	LTRharvest	long_terminal_repeat	5613	5830	.	-	.	Parent=LTR_retrotransposon4
seq1926	LTRharvest	target_site_duplication	5831	5834	.	-	.	Parent=repeat_region4
###								
seq2039	LTRharvest	repeat_region	2413	6089	.	-	.	ID=repeat_region5
seq2039	LTRharvest	target_site_duplication	2413	2416	.	-	.	Parent=repeat_region5
seq2039	LTRharvest	LTR_retrotransposon	2417	6085	.	-	.	ID=LTR_retrotransposon5; Parent=repeat_region5; ltr_similarity=88.46;seq_number=2039
seq2039	LTRharvest	long_terminal_repeat	2417	2517	.	-	.	Parent=LTR_retrotransposon5
seq2039	LTRdigest	RR_tract	2490	2514	.	-	.	Parent=LTR_retrotransposon5
seq2039	LTRharvest	long_terminal_repeat	5982	6085	.	-	.	Parent=LTR_retrotransposon5
seq2039	LTRharvest	target_site_duplication	6086	6089	.	-	.	Parent=repeat_region5
###								
seq233	LTRharvest	repeat_region	15	2550	.	?	.	ID=repeat_region6
seq233	LTRharvest	target_site_duplication	15	18	.	?	.	Parent=repeat_region6
seq233	LTRharvest	LTR_retrotransposon	19	2546	.	?	.	ID=LTR_retrotransposon6; Parent=repeat_region6; ltr_similarity=89.93;seq_number=233
seq233	LTRharvest	long_terminal_repeat	19	157	.	?	.	Parent=LTR_retrotransposon6
seq233	LTRharvest	long_terminal_repeat	2411	2546	.	?	.	Parent=LTR_retrotransposon6
seq233	LTRharvest	target_site_duplication	2547	2550	.	?	.	Parent=repeat_region6
###								
seq2341	LTRharvest	repeat_region	7464	8676	.	?	.	ID=repeat_region7
seq2341	LTRharvest	target_site_duplication	7464	7467	.	?	.	Parent=repeat_region7

Table 8-3 Evidence of LTR-retrotransposon presents in *H. columbae genome*

Seqid	Source	Type	Start	End	Score	Strand	Phase	Attributes
seq2341	LTRharvest	LTR_retrotransposon	7468	8672	.	?	.	ID=LTR_retrotransposon7; Parent=repeat_region7; ltr_similarity=85.93;seq_number=2341
seq2341	LTRharvest	long_terminal_repeat	7468	7598	.	?	.	Parent=LTR_retrotransposon7
seq2341	LTRharvest	long_terminal_repeat	8538	8672	.	?	.	Parent=LTR_retrotransposon7
seq2341	LTRharvest	target_site_duplication	8673	8676	.	?	.	Parent=repeat_region7
###								
seq2495	LTRharvest	repeat_region	1165	3430	.	?	.	ID=repeat_region8
seq2495	LTRharvest	target_site_duplication	1165	1168	.	?	.	Parent=repeat_region8
seq2495	LTRharvest	LTR_retrotransposon	1169	3426	.	?	.	ID=LTR_retrotransposon8; Parent=repeat_region8; ltr_similarity=87.04;seq_number=2495
seq2495	LTRharvest	long_terminal_repeat	1169	1276	.	?	.	Parent=LTR_retrotransposon8
seq2495	LTRharvest	long_terminal_repeat	3321	3426	.	?	.	Parent=LTR_retrotransposon8
seq2495	LTRharvest	target_site_duplication	3427	3430	.	?	.	Parent=repeat_region8
###								
seq2626	LTRharvest	repeat_region	5475	7008	.	?	.	ID=repeat_region9
seq2626	LTRharvest	target_site_duplication	5475	5478	.	?	.	Parent=repeat_region9
seq2626	LTRharvest	LTR_retrotransposon	5479	7004	.	?	.	ID=LTR_retrotransposon9; Parent=repeat_region9; ltr_similarity=88.12;seq_number=2626
seq2626	LTRharvest	long_terminal_repeat	5479	5578	.	?	.	Parent=LTR_retrotransposon9
seq2626	LTRharvest	long_terminal_repeat	6904	7004	.	?	.	Parent=LTR_retrotransposon9
seq2626	LTRharvest	target_site_duplication	7005	7008	.	?	.	Parent=repeat_region9
###								
seq654	LTRharvest	repeat_region	425	1761	.	?	.	ID=repeat_region10
seq654	LTRharvest	target_site_duplication	425	428	.	?	.	Parent=repeat_region10
seq654	LTRharvest	LTR_retrotransposon	429	1757	.	?	.	ID=LTR_retrotransposon10; Parent=repeat_region10; ltr_similarity=93.14;seq_number=654
seq654	LTRharvest	long_terminal_repeat	429	627	.	?	.	Parent=LTR_retrotransposon10
seq654	LTRharvest	long_terminal_repeat	1554	1757	.	?	.	Parent=LTR_retrotransposon10
seq654	LTRharvest	target_site_duplication	1758	1761	.	?	.	Parent=repeat_region10

Table 8-3.: Evidence of LTR-retrotransposon presents in *H. columbae genome*

References

Bensch, S., Canbäck, B., DeBarry, J. D., Johansson, T., Hellgren, O., Kissinger, J. C., Palinauskas, V., Videvall, E., and Valkiūnas, G. (2016). The Genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. *Genome biology and evolution*, 8(5):1361–1373.

Böhme, U., Otto, T. D., Cotton, J. A., Steinbiss, S., Sanders, M., Oyola, S. O., Nicot, A., Gandon, S., Patra, K. P., Herd, C., et al. (2018). Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome research*, 28(4):547–560.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). Ltrharvest, an efficient and flexible software for de novo detection of ltr retrotransposons. *BMC bioinformatics*, 9(1):18.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906):498.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, 26(5):589–595.

Magoč, T. and Salzberg, S. L. (2011). Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957–2963.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling genomes and mini-metagenomes from highly chimeric reads. In *Annual International Conference on Research in Computational Molecular Biology*, pages 158–170. Springer.

Steinbiss, S., Silva-Franco, F., Brunk, B., Foth, B., Hertz-Fowler, C., Berriman, M., and Otto, T. D. (2016). Companion: a web server for annotation and analysis of parasite genomes. *Nucleic acids research*, 44(W1):W29–W34.

Steinbiss, S., Willhoeft, U., Gremme, G., Kurtz, S., Steinbiss, S., Willhoeft, U., Gremme, G., and Fine-grained, S. K. (2010). Ltrdigest user's manual.

Valkiūnas, G. (2005). Avian malaria parasites and other haemosporidia CRC press. *Florida, Boca Raton*.

General Conclusions and Perspectives

- The importance of standardization of animal models allow us to characterize parasitic life cycles, to evaluate the response to the host and the vector when they are infected or not, to use a single parasite linage or strain, to extend our knowledge of the parasite-host-vector relationship that can be correlated with wild infections, among many others.
- The methodology used for the enrichment of parasites in blood samples allowed to obtain enough genomic sequences to assembled, annotated and characterized the mitochondrial and apicoplast genomes.
- The phylogenetic, phylogenetic and evolutionary analyzes on the ApiGenome provide evidence about the evolutionary dynamics of these genomes and which seems to be modeled by the GC content bias. In addition, these analyzes highlighted the importance of including molecular markers from this genome for future phylogenetic approaches.
- The draft nuclear genome was sufficient for identification of a large number of genes, however, *H. columbae* must be resequencing in order to complete the genome, to increase coverage, and to achieve comparative genomics analysis.
- This first genomic study of *H. columbae* will complement the previously published transcriptome of this same species. However, it should be continue to improve these omic data in order to implement this species as a standard model for the study of avian malaria.

Anexos

Participación en Congresos y Workshops

Figure A-1.: Participación Congreso **"First International Congress of Science, Technology and Innovation of the Americas"**; Modalidad Póster.

Figure A-2.: Participación Workshop **"Genomic epidemiology and Evolutionary Concepts in Infectious Diseases"**.

Figure A-3.: Participación Workshop y Congreso **"Fifth International and Interdisciplinary Workshop on Mathematical Modeling of Environment and Evolution on Social and Life Process"**; Modalidad Ponente.

北京师范大学 生命科学学院
COLLEGE OF LIFE SCIENCES, BEIJING NORMAL UNIVERSITY

CERTIFICATE OF ATTENDANCE

This Certificate is Awarded to

Axl Stivel Giraldo Cepeda

in recognition of participation in

The 4th International Conference on Malaria and Related Haemosporidian Parasites of Wildlife, at Beijing Normal University in Beijing, China, from November 1-6, 2018. In this conference Mrs. Axl Stivel Giraldo Cepeda has presented the oral contribution entitled "Experimental infection models: overcoming the genomic challanges of avian haemosporidians".

Chairman of the Organising Committee

Title Dr. Lu Dong *Date* 12 November, 2018

Figure A-4.: Participación Congreso **"4th International Conference on Malaria and Related Haemosporidian Parasites of Wildlife"**; Modalidad Ponente.

Figure A-5.: Participación Workshop **2019 EuPathDB Workshop**

UNIVERSIDAD
NACIONAL
DE COLOMBIA
Consejo de Facultad de Ciencias
Sede Bogotá

RESOLUCIÓN N°0173 DEL 23 DE ABRIL DE 2019

(Acta 007 de 2019)

"Por la cual se AVALA movilidad académica a unos estudiantes de Posgrado de la Facultad de Ciencias"

EL CONSEJO DE FACULTAD DE CIENCIAS
En uso de sus funciones delegadas en el Parágrafo IV del Artículo 4 de la Resolución 105 de 2017 de Vicerrectoría Académica y,

CONSIDERANDO:

QUE el Artículo 4 de la Resolución 105 de 2017 de la Vicerrectoría Académica reglamenta los requisitos para los intercambios académicos de estudiantes.

QUE 1. AXL STIVEL GIRALDO CEPEDA, con DNI 1032455849, de la Maestría en Ciencias-Biología, a través del Comité Asesor de posgrado de Biología, solicitó al Consejo de Facultad aval para realizar pasantía de investigación en el Laboratorio del Institute for Genomics and Evolutionary Medicine de la Universidad de Temple, de la ciudad de Filadelfia, Estados Unidos, en el periodo comprendido del 30 de abril al 24 de mayo de 2019.

QUE el Comité Asesor de posgrado de Biología en su sesión del 13 de marzo de 2019, Acta No.07, recomendó avalar la movilidad saliente en la modalidad de pasantía al laboratorio del Institute for Genomics and Evolutionary Medicine de la Universidad de Temple, bajo la tutoría de la profesora Nubia Matta. El trabajo en el laboratorio será supervisado por el Dr. Anannias Escalante. La pasantía se desarrollaría del 30 de abril al 24 de mayo.

QUE 2. AURA MARIA MORENO ECHEVERRI, con DNI 1072668329, de la Maestría en Ciencias Farmacéuticas, a través del Comité Asesor de Posgrados de Farmacia, solicitó al Consejo de Facultad aval para realizar pasantía de investigación al grupo de investigación BioNanoiTechnology de la Universidad de Wageningen, en el periodo comprendido del 01 de septiembre al 13 de diciembre de 2019.

QUE el Comité Asesor de Posgrados del Área Curricular de Farmacia según acta No. 5 de marzo 22 de 2019, recomendó avalar la pasantía de investigación en el grupo de investigación BioNanoiTechnology de la Universidad de Wageningen.

QUE 3. ANDREA VERONICA RODRIGUEZ MAYOR, con DNI 1013661924, de la Maestría en Ciencias Química, a través del Comité Asesor de Posgrado en Ciencias Química, solicitó al Consejo de Facultad aval para realizar pasantía de investigación al Departamento de Química del Centro de Investigación y de Estudios Avanzados del IPN - CINVESTAV, México D.F., en el periodo comprendido entre 1° de mayo hasta 26 de julio de 2019.

QUE el Comité Asesor de Posgrado en la sesión del 14 de marzo de 2019 (ACTA DE COMITÉ No. CACQ-08) recomendó dar el aval académico para realizar la pasantía de investigación a la estudiante ANDREA VERONICA RODRIGUEZ MAYOR a partir del 1 de mayo del 2019 y hasta el 26 de julio de 2019 en ciudad de México D.F. en el Departamento de Química del Centro de Investigación y de Estudios Avanzados del IPN - CINVESTAV1.

QUE el Consejo de Facultad de Ciencias en sesión del 04 de abril de 2019, Acta 007, luego de analizar las solicitudes presentadas y la recomendación de los Comités Asesores de Posgrado,

R E S U E L V E

ARTÍCULO 1. AVALAR pasantía de investigación en el Laboratorio del Institute for Genomics and Evolutionary Medicine de la Universidad de Temple, de la ciudad de Filadelfia, Estados Unidos, en el periodo comprendido del 30 de abril al 24 de mayo de 2019, bajo la tutoría de la profesora Nubia Matta y la supervisión por el Dr. Anannias Escalante, al estudiante AXL STIVEL GIRALDO CEPEDA, con DNI 1032455849, de la Maestría en Ciencias-Biología.

ARTÍCULO 2. AVALAR pasantía de investigación al grupo de investigación BioNanoiTechnology de la Universidad de Wageningen, en el periodo comprendido del 01 de septiembre al 13 de diciembre de 2019, bajo la tutoría de la profesora Yolima Baena Aristizábal., a la estudiante AURA MARIA MORENO ECHEVERRI, con DNI 1072668329, de la Maestría en Ciencias Farmacéuticas.

ARTÍCULO 3. AVALAR pasantía de investigación al Departamento de Química del Centro de Investigación y de Estudios Avanzados del IPN - CINVESTAV, México D.F., en el periodo comprendido entre 1° de mayo hasta 26 de julio de 2019, a la estudiante ANDREA VERONICA RODRIGUEZ MAYOR, con DNI 1013661924, de la Maestría en Ciencias Química.

ARTÍCULO 4. NOTIFICAR la presente decisión a los estudiantes haciéndoles saber que contra la misma procede recurso de reposición ante el Consejo de Facultad y en subsidio de apelación ante el Consejo de Sede, recursos que deben ser presentados con el lleno de los requisitos exigidos por el Artículo 77 del Código de Procedimiento Administrativo y de lo Contencioso Administrativo (Ley 1437 de 2011), dentro de los diez días siguientes a la notificación de la presente resolución en la Secretaría de Facultad de Ciencias.

ARTÍCULO 5. ENVIAR copia a las Direcciones de Área Curricular de Biología, Farmacia y Química.

N O T I F I Q U E S E Y C Ú M P L A S E
Dada en Bogotá D.C., a los veintitrés (23) días del mes de abril de dos mil diecinueve (2019).

JAIRO ALEXIS RODRÍGUEZ LÓPEZ
EL PRESIDENTE

CESAR AUGUSTO GÓMEZ SIERRA
EL SECRETARIO

Elaborado por Leonor Díaz

International Journal for Parasitology 48 (2018) 657–670

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Primers targeting mitochondrial genes of avian haemosporidians: PCR detection and differential DNA amplification of parasites belonging to different genera

M. Andreína Pacheco [a,*], Axl S. Cepeda [b,1], Rasa Bernotienė [c,1], Ingrid A. Lotta [b], Nubia E. Matta [b], Gediminas Valkiūnas [c], Ananias A. Escalante [a,*]

[a] Department of Biology/Institute for Genomics and Evolutionary Medicine (iGEM), Temple University, Philadelphia, PA 19122, USA
[b] Universidad Nacional de Colombia, Sede Bogotá-Facultad de Ciencias, Departamento de Biología, Grupo de Investigación Caracterización genética e inmunología, Carrera 30 No. 45-03, Bogotá 111321, Colombia
[c] Nature Research Centre, Akademijos 2, LT-08412 Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 20 November 2017
Received in revised form 9 February 2018
Accepted 15 February 2018
Available online 3 April 2018

Keywords:
Avian malaria
Cytochrome b
Haemoproteus
Mitochondrial genome
Nested-multiplex PCR
Plasmodium
Primers

ABSTRACT

Haemosporida is a diverse group of vector-borne parasitic protozoa, ubiquitous in terrestrial vertebrates worldwide. The renewed interest in their diversity has been driven by the extensive use of molecular methods targeting mitochondrial genes. Unfortunately, most studies target a 478 bp fragment of the cytochrome b (*cytb*) gene, which often cannot be used to separate lineages from different genera found in mixed infections that are common in wildlife. In this investigation, an alignment constructed with 114 mitochondrial genome sequences belonging to four genera (*Leucocytozoon*, *Haemoproteus*, *Plasmodium* and *Hepatocystis*) was used to design two different sets of primers targeting the *cytb* gene as well as the other two mitochondrial DNA genes: cytochrome c oxidase subunit 1 and cytochrome c oxidase subunit 3. The design of each pair of primers required consideration of different criteria, including a set for detection and another for differential amplification of DNA from parasites belonging to different avian haemosporidians. All pairs of primers were tested in three laboratories to assess their sensitivity and specificity under diverse practices and across isolates from different genera including single and natural mixed infections as well as experimental mixed infections. Overall, these primers exhibited high sensitivity regardless of the differences in laboratory practices, parasite species, and parasitemias. Furthermore, those primers designed to separate parasite genera showed high specificity, as confirmed by sequencing. In the case of *cytb*, a nested multiplex (single tube PCR) test was designed and successfully tested to differentially detect lineages of *Plasmodium* and *Haemoproteus* parasites by yielding amplicons with different sizes detectable in a standard agarose gel. To our knowledge, the designed assay is the first test for detection and differentiation of species belonging to these two genera in a single PCR. The experiments across laboratories provided recommendations that can be of use to those researchers seeking to standardise these or other primers to the specific needs of their field investigations.

1. Introduction

Haemosporidians (Phylum Apicomplexa, Order Haemosporida) are a diverse group of unicellular blood parasites that infect a variety of vertebrate hosts including amphibians, reptiles, birds and mammals (Garnham, 1966; Valkiūnas, 2005; Telford, 2009). These vector-borne parasitic protozoa are classified into four families, the Plasmodiidae, Garniidae, Haemoproteidae and Leucocytozoidae, with most of the known species belonging to three genera, *Plasmodium*, *Haemoproteus* and *Leucocytozoon* (Valkiūnas, 2005; Telford, 2009). In the case of avian haemosporidians, these cosmopolitan parasites are found in birds belonging to different families with remarkably complex patterns in terms of host-parasite speciation and evolution (Ricklefs and Fallon, 2002; Bensch et al., 2004; Ricklefs et al., 2004; Valkiūnas, 2005; Nilsson et al., 2016; Pacheco et al., 2018). Many infections appear to be subpatent and submicroscopic, making it difficult to ascertain the prevalence

* Corresponding authors.
E-mail addresses: Maria.Pacheco@temple.edu (M.A. Pacheco), Ananias.Escalante@temple.edu (A.A. Escalante).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ijpara.2018.02.003
0020-7519/

of these parasites in specific host populations. Furthermore, those parasite species linked to severe avian diseases are often particularly difficult to diagnose using microscopy during the exoerythrocytic stage of development (Valkiūnas and Iezhova, 2017). This calls for improvement of molecular tools for detection of haemosporidian infections.

Investigations into the diversity of these parasites have been possible due to the extensive use of mitochondrial genes, particularly cytochrome b (*cytb*), in ecological and evolutionary studies (Escalante et al., 1998; Perkins and Schall, 2002; Ricklefs and Fallon, 2002; Bensch et al., 2004; Hellgren et al., 2004). The use of this locus is facilitated by its relative conservation across species and its high copy number which together allow its amplification by PCR from a broad range of host species. Indeed, *cytb* has become a de facto DNA barcoding gene for avian malaria (Bensch et al., 2009; Outlaw and Ricklefs, 2014; Valkiūnas et al., 2017) even though it has not been properly standardised (Bergsten et al., 2012; Pacheco et al., 2018). Despite the higher sensitivity of molecular methods to detect parasites than that of microscopy (e.g., Cheng et al., 2015), a limitation on the use of molecular approaches in studying avian parasites is how to link sequences/molecular markers to species when our understanding of their taxonomy and biodiversity remains limited. The situation is further complicated by the fact that haemosporidians, similar to other parasitic organisms (Poulin, 2007), are usually part of multiple species infections. In fact, there is a high prevalence of haemosporidian infections of different genetic lineages and/or species belonging to the same and different genera in wild birds (Valkiūnas et al., 2003, 2006; Beadell et al., 2004; Pérez-Tris and Bensch, 2005; Loiseau et al., 2010; Silva-Iturriza et al., 2012; Dimitrov et al., 2014; Clark et al., 2016), reptiles (Falk et al., 2011; Telford, 2009), and primates (Pacheco et al., 2012, 2013; Muehlenbein et al., 2015). These mixed infections often predominate in some bird populations and are a challenge to researchers attempting to identify avian parasites because most described species remain poorly characterised in terms of molecular markers; the situation worsens in the case of potential new species.

Current PCR methods target conserved regions of the *cytb* gene across haemosporidian species of different genera, so those often overlook mixed infections simply because the same or similar amplicon sizes are expected (Pérez-Tris and Bensch, 2005; Valkiūnas et al., 2006; Martínez et al., 2009; Zehtindjiev et al., 2012). Furthermore, direct sequencing of PCR products may create chimeras or "consensus" sequences that are impossible to assign to species. Although the combination of PCR and microscopy can mitigate the problem of detecting mixed infections, parasite species identification in multiple infections using PCR assays remains a problem (Pérez-Tris and Bensch, 2005; Ishtiaq et al., 2017). To illustrate the situation, a recent study on avian haemosporidians using experimental mixed infections from different genera showed that a single PCR assay markedly underestimated the number of species and/or lineages found in mixed infections. Furthermore, most of the lineages presented in a mixed infection were detected only when at least three PCR assays were done in parallel (Bernotienė et al., 2016).

In this investigation, taking advantage of an alignment constructed with a total of 114 mitochondrial (mt)DNA genome sequences (Pacheco et al., 2018) belonging to four genera (*Leucocytozoon*, *Haemoproteus*, *Plasmodium* and *Hepatocystis*), different sets of primers targeting mtDNA genes (specifically *cytb*) were designed for both PCR detection and differential DNA amplification of parasites belonging to different genera of avian haemosporidians. To evaluate the sensitivity and specificity of these sets of primers, we compare the results from three laboratories that used different practices and reagents to test them. These different settings allowed testing of the primers on different laboratory strains, natural single infections, and natural and experimental mixed infections.

2. Materials and methods

2.1. Primer design

Two groups of primers are reported in this study: (i) four pairs for PCR detection of parasites or group 1 and (ii) five for differential DNA amplification of parasite DNA belonging to different genera of avian haemosporidians or group 2 (Fig. 1). As for other primers currently available, the primers of group 1 (PCR detection) can be used for lineage identification and direct sequencing of PCR amplicons for phylogenetic purposes if the samples are single infections. The advantage of the primers included in group 1 over others (currently used) is that amplicons are larger than 900 bp, providing more informative sites for phylogenetic analyses, and those overlap with the *cytb* fragment commonly used in parasitological and ecological investigations. Thus, group 1 primers can support more demanding phylogenetic inferences while allow comparison of the sampled lineages with those previously reported. For differential DNA amplification of avian haemosporidians belonging to different genera (group 2), genus-specific primers for *Plasmodium* or *Haemoproteus* (*Parahaemoproteus*) spp. were tested.

All primers were designed using conserved gene regions of cytochrome c oxidase subunit 1 (*cox1*), cytochrome c oxidase subunit 3 (*cox3*), and *cytb* in an alignment of 114 mtDNA genome sequences (5,125 kbp excluding gaps, DOI: https://doi.org/10.17632/jtz23sgttf.1), that include all the sequences available at GenBank to date from *Leucocytozoon* ($n = 13$), *Haemoproteus* ($n = 27$), *Plasmodium* ($n = 71$ including bird, reptile and human parasites), and *Hepatocystis* ($n = 3$) spp. In addition, different lineages for some parasite species, which were identified using morphological characters (e.g., *Plasmodium lutzi*, *Haemoproteus columbae*, etc.), were included in the alignment (see Supplementary Table S1 for sample codes in Pacheco et al. (2018)). In order to achieve specific PCR amplification with high yields, standard primer design considerations were followed (Jennings, 2017), in particular: (i) optimal length varies between 18 and 30 bp; this length allows for adequate specificity and remains short enough for primers to bind easily to the template at the annealing temperature; (ii) optimal melting temperatures are in the range of 52–58 °C that generally produces the best results; (iii) optimal GC content (the number of Gs and Cs in the primer as a percentage of the total bases) of the primer is between 30 to 60%; (iv) due to the stronger bonding of G and C bases, primers have one or two G and/or C bases within the last five bases from the 3′ end of primers (GC clamp) to promote specific binding at the 3′ end; (v) primer secondary structures such as hairpins, self-and cross-dimer were avoided (this is necessary because secondary structures produced by intermolecular or intramolecular interactions can affect primer-template annealing, generating poor or no yield of PCR product); (vi) repeats with a di-nucleotide occurring many times consecutively were avoided because those can misprime; (vii) primers with long runs of a single base were avoided as they can misprime so runs of a maximum of 4 bp were accepted; (viii) to improve specificity of the primers, template secondary structure and cross homology were also avoided (designed primers do not amplify other genes in the mixture); and (ix) in the case of degenerate primers, only a maximum of four positions in the oligonucleotide containing a mixture of base pairs was allowed. All these requirements were checked using the online tool Oligo Calc: Oligonucleotide Properties Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html).

In order to test the reproducibility of the amplification yield with these primers across different laboratory practices, experiments were performed in three laboratories using diverse types of samples, PCR conditions and equipment (Fig. 2); laboratory A,

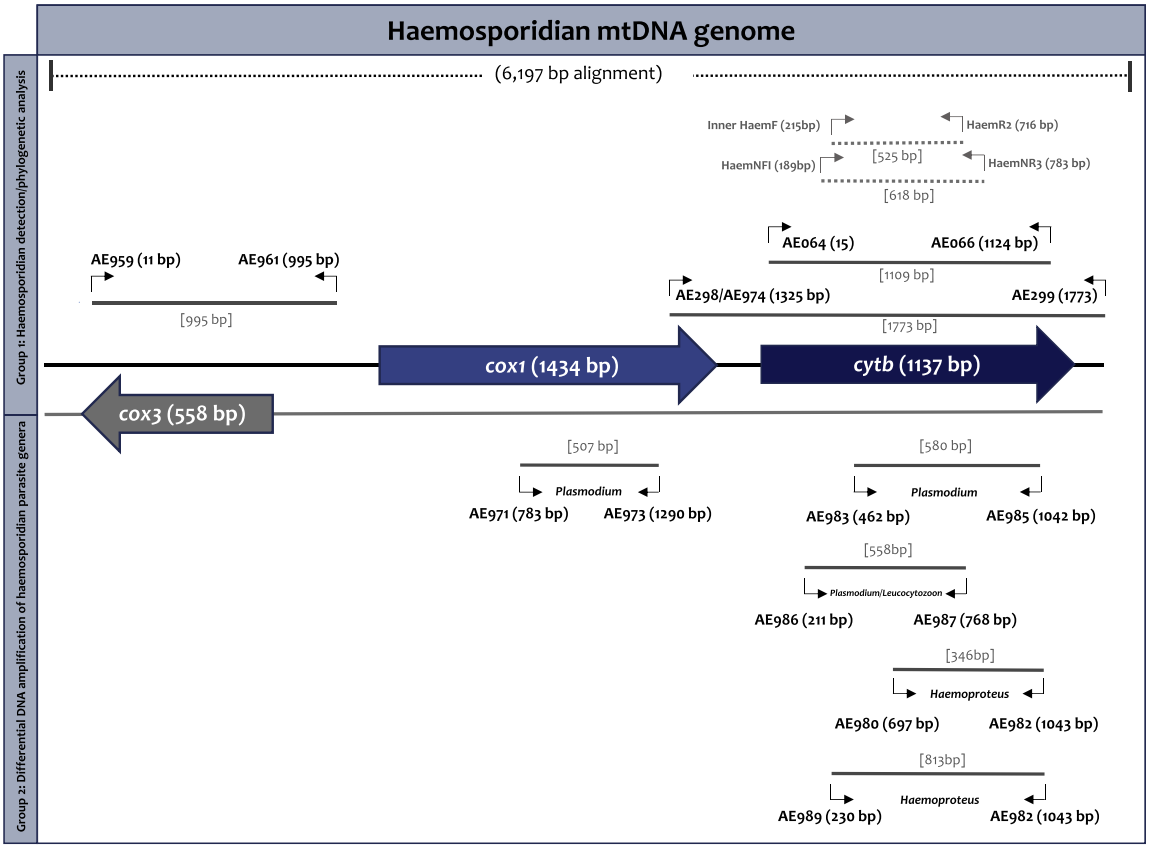


Fig. 1. Diagram of mitochondrial genes (mtDNA) with primer locations for PCR detection (longer fragments, >995 bp) and differential DNA amplification of haemosporidian parasite genera. Primer specificity for differential DNA amplification of haemosporidian parasite genera is shown. Primer locations are indicated by parentheses and are relative to the genes in the alignment, and the amplicon size for each set of primers are in square brackets. The regions that could be amplified using previous published primers are shown for comparation (primers HaemF/HaemR2 and HaemNF1/HaemNR3, Bensch et al. 2000; Hellgren et al. 2004). *Cox1*, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; *cytb*: cytochrome b.

Evolutionary Dynamics of Infectious Diseases Laboratory at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University, USA; laboratory B, P.B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania; and laboratory C, Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia. First, the primers were tested and a general PCR protocol was optimised by laboratory A, and then the other two laboratories adapted the protocol to test the primers on their samples. Given that each laboratory has its own practices and reagents, some conditions were differed between them. Nevertheless, we tested how robust the amplifications were when these primers were used for a diverse set of samples. Primer sequences, properties of each primer (including the targeted genera) and the PCR conditions used in each laboratory are shown in Table 1.

2.2. Samples and PCR protocols

Supplementary Table S1 describes the parasite species (*cytb* lineage code/sample ID), host and intensity of parasitemia (%) of the samples used in each laboratory. The intensity of parasitemia was estimated by each laboratory as a percentage by actual counting of the number of parasites per 10,000 erythrocytes.

2.2.1. Evolutionary Dynamics of Infectious Diseases Laboratory, USA (laboratory A)

Archived field samples and laboratory strains of haemosporidian parasites were used to test all the primers described in Table 1 (Supplementary Table S1). Genomic DNA was extracted from whole blood or tissues (liver) using a DNeasy® Blood & Tissue Kit (Qiagen, GmbH, Hilden, Germany). DNA from parasite lineages which have been previously published, from lizards and birds as well as well-known human parasites, were amplified as follows: PCRs were carried out in 50 µl with 2 µl of total genomic DNA, 2.5 mM $MgCl_2$, 1X PCR buffer, 0.2 mM of each deoxynucleoside triphosphate, 0.4 mM of each primer, and 0.03 U/µl of AmpliTaq polymerase (Applied Biosystems, Roche, USA). Amplification conditions for all PCRs were: a partial denaturation at 94 °C for 4 min, 36 cycles of 1 min at 94 °C, 1 min at primer melting temperature (°C, Table 1) and 2 min extension at 72 °C, with a final extension of 10 min added to the last cycle. A negative control (dH_2O), and positive control (*Plasmodium vivax*) were included. All amplifications were evaluated by running the total PCR products (50 µl) on a 1% agarose gel. Bands with the expected molecular size were excised from the gel (the size depended of the pair of primer, Table 1), purified using an QIAquick® Gel extraction kit (Qiagen,

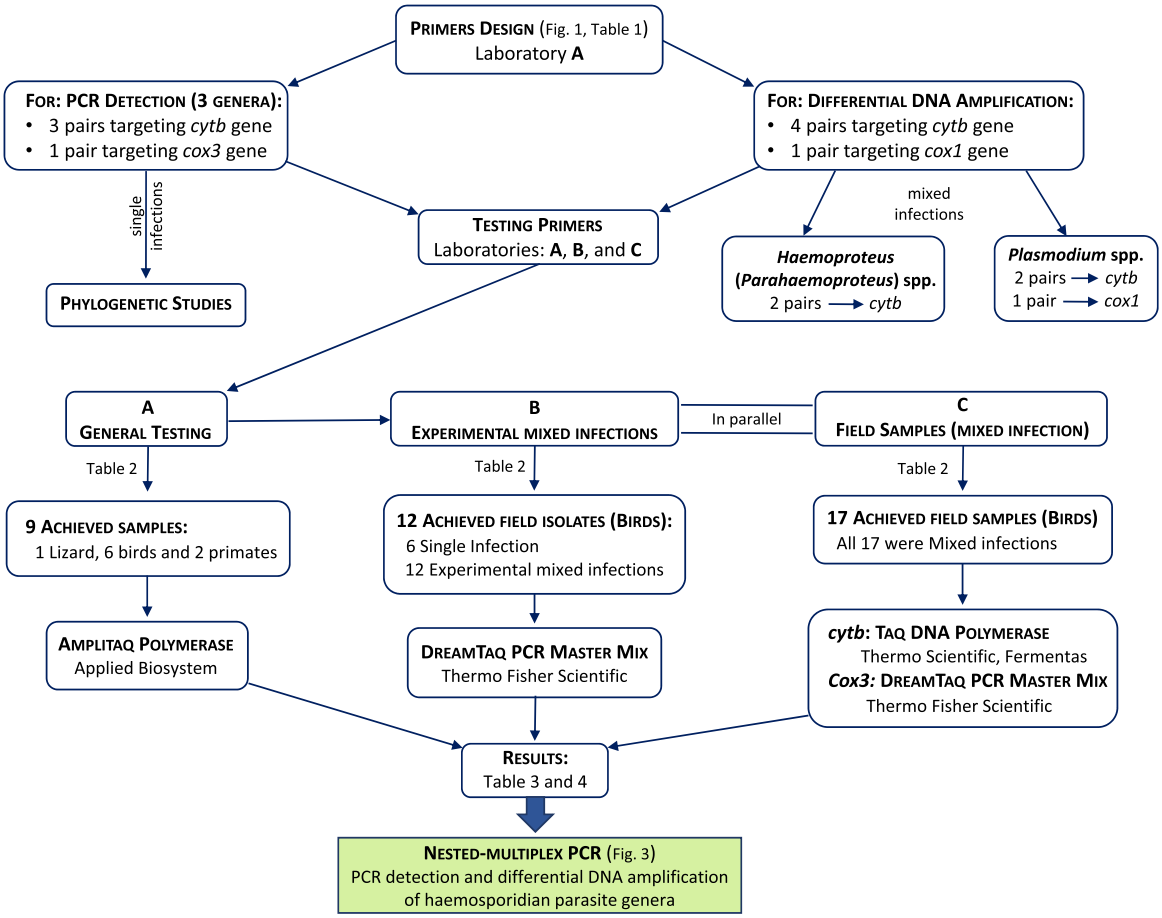


Fig. 2. Diagram of the experimental design of the current study. Laboratories (Labs): A, Evolutionary Dynamics of Infectious Diseases Laboratory at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University, USA; B, P. B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania; and C, Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia. *Cox1*, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; and *cytb*, cytochrome b.

GmbH, Hilden, Germany), and both strands for each PCR product were sequenced with the corresponding pairs of primers using an Applied Biosystems 3730 capillary sequencer. The obtained sequences were confirmed using Basic Local Alignment Search Tool (BLAST, at the National Center for Biotechnology Information: https://blast.ncbi.nlm.nih.gov/Blast.cgi) and those were aligned by using ClustalX v2.0.12 and Muscle as implemented in SeaView v4.3.5 (Gouy et al., 2010). The primers designed and tested in laboratory A were validated at the P.B. Šivickis Laboratory of Parasitology, Lithuania (laboratory B) and Host-Parasite Relationship Laboratory, Colombia (laboratory C), each using their own samples. Primer annealing temperature (°C) and $MgCl_2$ concentration for each pair of primers used by each laboratory are also shown in Table 1 for comparison.

2.2.2. P.B. Šivickis Laboratory of Parasitology, Lithuania (laboratory B)

Archived *Haemoproteus*, *Leucocytozoon* and *Plasmodium* parasite samples were obtained from naturally infected birds captured at the Biological Station of the Zoological Institute of the Russian Academy of Sciences on the Curonian Spit in the Baltic Sea, following the current laws of Lithuania and Russia (Supplementary Table S1). Whole blood samples were stored in SET buffer (0.05 M Tris, 0.15 M NaCl, 0.5 M EDTA, pH 8.0). Parasite species and their lineages were identified by microscopic examination of blood films and PCR targeting the *cytb* gene, respectively. Samples with single infections, as determined both by microscopic examination and PCR-based testing, were used in this study (Supplementary Table S1). Total DNA was extracted from samples using an ammonium acetate extraction method (Richardson et al., 2001). Quantification of DNA was performed by using a Nanodrop spectrophotometer (IMPLEN Nanophotometer P330). The samples, in which the total DNA concentration exceeded 100 ng/µl, were resolved with TE buffer (10 mM Tris, 1 mM EDTA) to a final DNA concentration of 41 ± 5.3 ng/µl.

Using the genomic DNA from single infections, 12 experimental mixes of different haemosporidian combinations were prepared for primer testing (Supplementary Table S1). Such mixed infections often occur naturally in wildlife in Europe (Valkiūnas et al., 2006; Dimitrov et al., 2014, 2015). In all cases, 15 µl of total DNA of a similar concentration of each parasite lineage was used for the preparation of mixes. Bernotienė et al. (2016) provided a detailed description of preparation of the experimental parasite mixes.

All PCRs were performed in 25 µl reactions with 2 µl of total genomic DNA, 0.2 mM of each deoxynucleoside triphosphate, 0.4 mM of each primer, and 12.5 µL of DreamTag Master Mix (it includes DreamTaq DNA Polymerase, 2X DreamTaq buffer, 0.4

Table 1
Primers, their properties, and summary of PCR conditions tested by the three laboratories (A–C).

							PCR protocols								
							A		B		C		Genus Target		
		Code (pairs)	Primers sequences (5′–3′)	Size (bp)	GC (%)	FS[a] (bp)	Ta (°C)	$MgCl_2$ (μM)	Ta (°C)	$MgCl_2$ (μM)	Ta (°C)	$MgCl_2$ (μM)	L	H	P
Group 1: Haemosporidian detection/phylogenetic analysis	*cytb*	AE298-EF	TGTAATGCCTAGACGTATTCC	21	43	1773-1741	53	2.5	53	4	53	2.5	X	X	X
		AE299-ER	GTCAAWCAAACATGAATATAGAC	23	30										
		AE064-IF[b]	TCTATTAATTTAGYWAAAGCAC	22	41	1109	56	2.5	56	4	56	2.5	X	X	X
		AE066-IR[b]	GCTTGGGAGCTGTAATCATAAT	22	41										
		AE974-EF	TGTAATGCCTAGAMGWATWCC	21	38-43	1741	56	2.5	56	4	56	2.5	X	X	X
		AE299-ER	GTCAAWCAAACATGAATATAGAC	23	30										
	cox3	AE959-F	CCATACAATYTCNACRAAATGCC	23	35-48	995	55	2.5	55	4	55	2.5	X	X	X
		AE961-R	CTGTTATCCCCGGCGAACC	19	63										
Group 2: Differential DNA amplification of haemosporidian parasite genera	*cytb*	AE980-F	AAAGTTTATTTGGWATWYTRCCWTTAG	27	22-30	346	57	2.5	57	4	54.3	2.5		X	
		AE982-R	AAACGACCATATAAAATRWARATAG	25	20-28										
		AE989-F	TATGCAYGCTACHGGWGCTAC	21	48-57	813	57	2.5	57	4	54.3	2.5		X	
		AE982-R	AAACGACCATATAAAATRWARATAG	25	20-28										
		HaemF-EF[b,c]	ATGGTGCTTTCGATATATGCATG	23	39	828	56	2.5	50	4	53.5	2.5		X	
		AE982-ER[b]	AAACGACCATATAAAATRWARATAG	25	20-28										
		AE983-F	TGGATHTGTGGWGGATATYTWG	22	36-45	580	57	2.5	57	4	56	2.5			X
		AE985-R	AACGACCATATAWAATGWADATATC	25	24-28										
		AE986-F	AGTGGATGGTGYTTYAGATAYTTAC	25	32-44	558	57	2.5	57	4	58.4	2.5	X		X
		AE987-R	AGGTGTTGCATATNTATYWACTGG	24	33-42										
	cox1	AE971-F	AGTCATGTAATMTCWACTAAYTAYTC	26	23-35	507	56	2.5	56	4					X
		AE973-R	AACTACTCCTAYRAARAATAACATTG	26	23-35										

(A) Evolutionary dynamics of infectious diseases Laboratory at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University-USA, (B) P. B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania, and (C) Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia. (*L*) *Leucocytozoon* spp., (*H*) *Haemoproteus* (*Parahaemoproteus*) spp., and (*P*) *Plasmodium* spp. X corresponds to positive amplifications. *Cox1*, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; and *cytb*, cytochrome b.

[a] FS is the amplicon fragment size including primer region, and Ta is the annealing temperature.

[b] These primers can be used for a nested PCR.

[c] Primer published by Bensch et al. 2000.

mM each dNTP, and 4 mM $MgCl_2$; Thermo Fisher Scientific, Lithuania). One negative control (nuclease-free dH_2O) and one positive control (an infected sample, which was positive by microscopic examination of blood films) were used. Amplification conditions for all PCRs were the same as used by laboratory A. All amplifications were evaluated by running 3 μl of the final PCR products on 2% agarose gels. PCR products from all positive amplifications were precipitated with ammonium acetate and 95% ethanol, and were sequenced with corresponding primers twice for both strands. We used dye terminator cycle sequencing (Big Dye) and loaded samples onto an ABI 201 PRISM TM 3100 capillary sequencing robot (Applied Biosystems, USA). The obtained sequences were aligned and analyzed using the Bioedit program (Hall, 1999). Mixed infections were determined by visualising double-base calling in sequence electropherograms, and sequences obtained from experimental mixes were compared with corresponding sequences from initial single parasite infections.

2.2.3. Host-Parasite Relationship Laboratory, Colombia (laboratory C)

Archived samples with naturally mixed infections were used to test only primer sensitivity (Supplementary Table S1). Parasites species and their lineages were previously identified by examination of blood films using microscopy, PCR targeting the *cytb* gene, and sequencing. Total DNA was extracted from whole blood samples preserved in SET buffer or EDTA using a phenol–chloroform method (Sambrook et al., 1989), Zymo DNA Purification kit (Zymo Research Inc., Orange, USA) or DNeasy Blood & Tissue kit (Qiagen, GmbH, Hilden, Germany). Quantification of the total DNA was performed by using a Nanodrop spectrophotometer (Thermo Fisher Scientific, California, USA). Quality of the extracted DNA was verified by PCR with primers targeting the host *cytb* gene (Sawabe et al., 2010).

All PCRs targeting the *cytb* gene were carried out in 50 μl reactions with 2 μl of total genomic DNA, 2.5 mM $MgCl_2$, 1X PCR buffer, 0.2 mM of each deoxynucleoside triphosphate (Promega, Madison, Wisconsin), 0.4 mM of each primer, and 0.03 U/μl of Taq DNA polymerase (Fermentas, Thermo Fisher Scientific, USA). Amplification conditions for all PCRs were the same as used by laboratory A. However, PCRs for *cox3* were also carried out in 50 μl following the *cytb* protocol used by Bernotienė et al. (2016). In the case of *cox1*, different PCR protocols, which have been used for *cytb* and *cox1*, were also tested (Bensch et al., 2000; Hellgren et al., 2004; Beadell et al., 2004; Martinsen et al., 2008; Pacheco et al., 2011; Bernotienė et al., 2016). In all cases, a negative control (dH_2O), and a positive control (infected bird sample) were included. Amplifications were visualised by running 3 μl of PCR products on a 1.5% agarose gel stained with SYBR™ Safe DNA Gel Stain (ThermoFisher Scientific, California, USA). Total PCR products were precipitated with ammonium acetate and 95% ethanol (Bensch et al., 2000), and bi-directional sequencing of amplification products was conducted using a 3730xl DNA Analyzer (Applied Biosystems, Foster City, USA). Then all sequences were aligned using MEGA v7.0 (Kumar et al., 2016) and the lineages were confirmed using BLAST.

2.3. Primer sensitivity and specificity

The primers are compared in terms of their sensitivity (or probability of detecting the parasite DNA if the host is infected) and specificity (true negative rate) as those were independently estimated for the results from each laboratory.Sensitivity was calculated as $= \frac{\#\ of\ true\ positives}{\#\ of\ positives} = = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ positives}$, and specificity as $= \frac{\#\ of\ true\ negatives}{\#\ of\ negatives} = = \frac{\#\ of\ true\ negatives}{\#\ of\ true\ negatives + \#\ of\ false\ positives}$.

2.4. Nested multiplex PCR for differential DNA amplification of Haemoproteus and Plasmodium spp. from field isolates

Using the genomic DNA from single infections of four bird parasite lineages that were previously characterised by laboratory B, four experimental mixes from a combination of 10 μl of total DNA from each species were prepared (see Fig. 3 for lineage combinations and initial parasitemias). The mixed DNAs were amplified by a nested multiplex PCR as follows: first, a primary amplification was performed using primers AE298/AE299, and the reaction was carried out in 50 μl with 4 μl of total mixed genomic DNA, 2.5 mM $MgCl_2$, 1X PCR buffer, 1.25 mM of each deoxynucleoside triphosphate, 0.4 mM of each primer, and 0.03 U/μl of AmpliTaq polymerase (Applied Biosystems, Roche-USA). Amplification conditions for the PCR were: a partial denaturation at 94 °C for 4 min, 25 cycles with 1 min at 94 °C, 1 min at 55 °C and 2 min extension at 72 °C, with a final extension of 10 min added to the last cycle. Second, a nested-multiplex PCR was done using both pairs of primers, AE980/982 and AE983/985. PCRs were also carried out in 50 μl with 1 μl of the primary PCR, 2.5 mM $MgCl_2$, 1X PCR buffer, 1.25 mM of each deoxynucleoside triphosphate, 0.4 mM of each primer, and 0.03 U/μl of AmpliTaq polymerase (Applied Biosystems, Roche-USA). Amplification conditions for the PCR were: a partial denaturation at 94 °C for 4 min, 30 cycles with 1 min at 94 °C, 1 min at 59 °C and 2 min extension at 72 °C, with a final extension of 10 min added to the last cycle. A negative control (dH_2O) was included in both PCRs. Then, all amplifications were evaluated by running the total PCR products (50 μl) on a 1.5% agarose gel. Other primer combinations were tested as part of the nested multiplex PCR protocols, but those showed low specificity (e.g., AE989/982 and AE983/985 or AE989/982 and AE986/987) so only the combination of the pair of primers that gave positive amplification are reported.

Furthermore, this nested multiplex protocol was tested at the laboratories B and C. Although the same conditions for the PCR were used, laboratories B and C used different Taq DNA polymerases (DreamTaq Master Mix, Thermo Fisher Scientific-Lithuania, and Taq DNA polymerase, Thermo Scientific, Fermentas, USA, respectively) and samples with experimental (laboratory B) or naturally mixed infections (laboratory C) (see Fig. 3 for lineage combinations and initial parasitemias). In addition to primers AE298/AE299 for the primary PCR, the combination of primers AE974/299 was also tested.

2.5. Phylogenetic signal of cytb and cox3 fragments

To show the differences in phylogenetic signal (whether there are informative sites to solve phylogenetic relationships between a given set of taxa) of the *cytb* (479 bp from the commonly used primers or 1065 bp from the primers reported here) and *cox3* (761 bp from the primers reported here) fragments, three alignments were constructed using 70 mtDNA genome sequences available for *Leucocytozoon*, *Haemoproteus* and *Plasmodium* spp. (Pacheco et al., 2018); two alignments for *cytb* fragments (479 bp and 1065 bp), and one using concatenated *cytb* + *cox3* fragments (1065 bp + 761 bp = 1826 bp) (DOI: https://doi.org/10.17632/jtz23sgttf.1). In all cases, the primer regions were not included in these alignments. Then, phylogenetic relationships were estimated by Bayesian methods implemented in MrBayes v3.2.6 with the default priors (Ronquist and Huelsenbeck, 2003). A general time reversible model with gamma-distributed substitution rates and a proportion of invariant sites (GTR+ Γ +I) was used for each alignment; it was the model with the lowest Bayesian Information Criterion (BIC) scores as estimated by MEGA v7.0.14 (Kumar et al., 2016). Bayesian support for the nodes was inferred in MrBayes

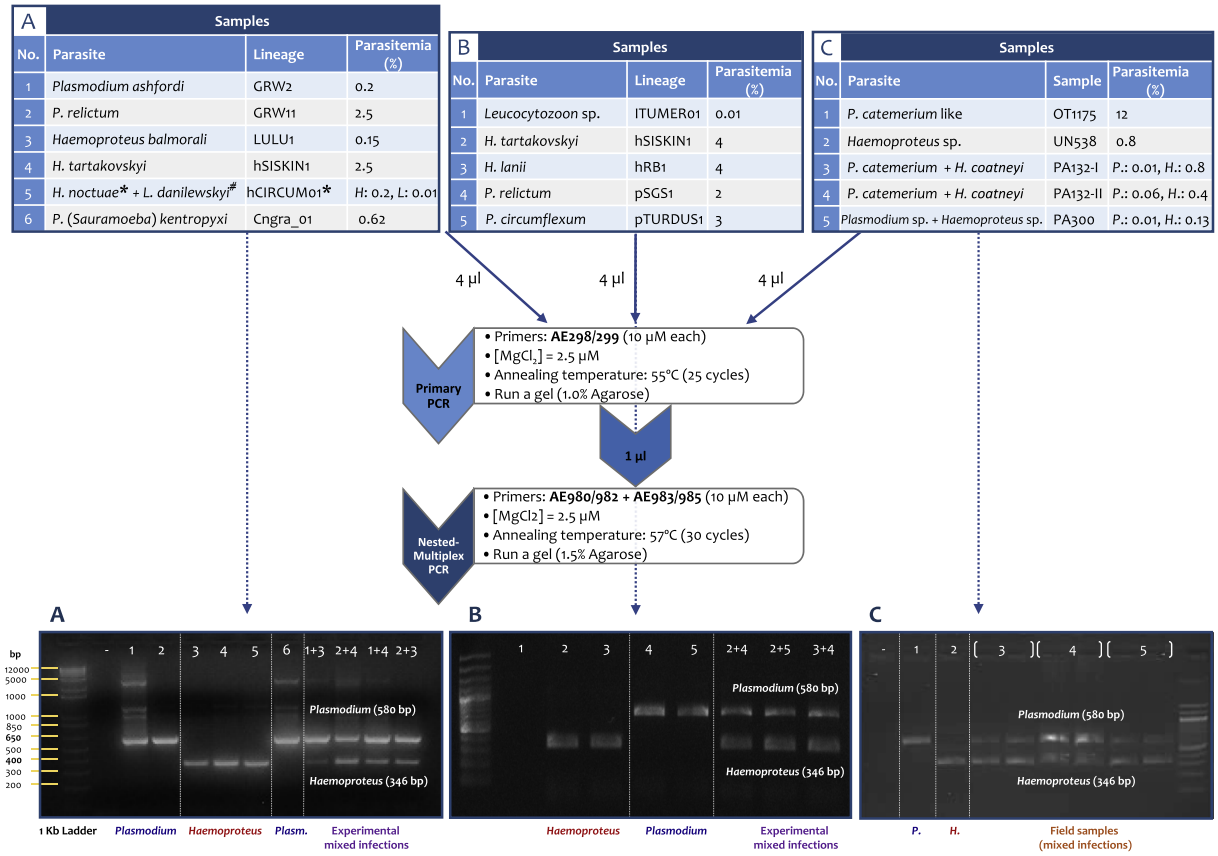


Fig. 3. Experimental design and results of a nested multiplex PCR for PCR detection and differential DNA amplification of haemosporidian parasite genera of parasites belonging to *Plasmodium* (*P.*) and *Haemoproteus* (*H.*) genera. Parasites, linages and parasitemias are shown. The numbers correspond to the parasite species and the combination of experimental mixed infection used to test this protocol. (A) Evolutionary dynamics of infectious diseases Laboratory. at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University-USA, (B) P. B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania, and (C) Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia (see Fig. 2). *Correspond to the *Haemoproteus* lineage and # no lineage has been identified for *Leucocytozoon danilewskyi*. *Plasm.*, *Plasmodium*.

by sampling every 500 generations from two independent chains lasting 2×10^6 Markov Chain Monte Carlo (MCMC) steps. The chains were assumed to have converged once the average standard deviation of the posterior probability was below 0.01 and the value of the potential scale reduction factor (PSRF) was between 1.00 and 1.02 (Ronquist and Huelsenbeck, 2003). As a "burn-in," 50% of the sample was then discarded once convergence was reached.

3. Results

New sets of primers for detection/phylogenetic analysis and differential DNA amplification of avian haemosporidian mitochondrion genes were successfully tested at three laboratories (Table 2) following their own practices and protocols (Table 1). Sensitivity and specificity results are shown as a proportion (instead of %) due to the differences in the numbers of well-characterised samples that each laboratory was able to use. However, no major inconsistencies in the results were found between laboratories, so an overall percentage was estimated, combining all the results for each pair of primers (Table 3).

3.1. Haemosporidian parasite detection and phylogenetic analysis

In the case of PCR detection of haemosporidian parasites, *cytb* primers AE298/299 and AE974/299 were used for outer PCRs and then, if amplicons were not detected in an agarose gel, AE064/066 were used in a nested PCR. It is worth noting that laboratory C (Colombia) tested these two pairs of primers on a set of 32 field samples with single or mixed natural infections (Supplementary Table S2). Although there were differences in the success of DNA amplifications between laboratories, the overall sensitivity of these three *cytb* sets of primers was greater than 80% (Table 3). Similar results were also found for *cox3* primers AE959/961 (Table 3). The Bayesian phylogenies estimated from the *cytb* and *cox3* fragments are shown in Fig. 4 and Supplementary Fig. S1. Overall, the *cytb* fragment of 1065 bp (Fig. 4) and concatenated *cytb*/*cox3* fragments (1826 bp) (Supplementary Fig. S1) gave phylogenies with well-solved clades.

3.2. Differential DNA amplification of haemosporidian parasite genera

With regard to the five pairs of *cytb* primers designed for differential DNA amplification of haemosporidian parasite genera, the primer combinations AE980/982, AE989/982 and HaemF/AE982 amplified only DNA of *Haemoproteus* (*Parahaemoproteus*) spp. (Table 2). The overall sensitivity of these primers was higher than 71%, and the specificity was 100% (Table 3), thus these can be used for diagnosis and selective amplification of *Haemoproteus* during mixed infections with *Plasmodium* and/or *Leucocytozoon* spp. The *cytb* pair of primers AE983/985 amplified only DNA of *Plasmodium*

Table 2
Results for PCR assays using the new sets of primers for detection/phylogenetic analysis and differential DNA amplification of parasites belonging to three avian haemosporidian genera.

		Gene	Detection/phylogenetic analysis				Differential DNA amplification					
			cytb			*cox3*	*cytb* (for *H.* spp.)			*cytb* (for *P.* spp.)		*cox1* (for *P.* spp.)
Lab.		Primer codes (pairs)/parasites species	298/299	064/066	974/299	959/961	980/982	989/982	HaemF/982	983/985	986/987	971/973
A	Field and strain isolates	*Leucocytozoon danilewskyi* + *Haemoproteus noctuae* (hCIRCUM01)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	*L.* (+)	(−)
		Leucocytozoon sp. (PA262)	*L.* (+)	*L.* (+)	*L.* (+)	*L.* (+)	(−)	(−)	(−)	(−)	*L.* (+)	(−)
		Haemoproteus macrovacuolatus (CA1017)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	(−)
		Haemoproteus lanii (hRB1)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	(−)
		Plasmodium kentropyxi (Cngra_01)	*P.* (+)	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	*P.* (+)	*P.* (+)
		Plasmodium ashfordi (GRW02)	*P.* (+)	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	(−)	*P.* (+)
		Plasmodium unalis (TFUS06)	*P.* (+)	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	*P.* (+)	*P.* (+)
		Plasmodum vivax (Salvador I)	*P.* (+)	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	(−)	*P.* (+)
		Plasmodium falciparum (3D7)	*P.* (+)	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	(−)	*P.* (+)
B	Single infections	*Leucocytozoon* sp. (ITUMER01)	*	*	*L.* (+)	*L.* (+)	(−)	*	*	(−)	*L.* (+)	(−)
		Haemoproteus tartakovskyi (hSISKIN1)	*	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	(−)
		H. lanii (hRB1)	*	*H.* (+)	*	(−)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	(−)
		Plasmodium relictum (pSGS1)	*	*P.* (+)	*	*P.* (+)	(−)	(−)	(−)	*P.* (+)	*P.* (+)	*P.* (+)
		P. relictum (pGRW4)	*	*P.* (+)	*	(−)	(−)	(−)	(−)	*P.* (+)	(−)	*P.* (+)
		Plasmodium circumflexum (pTURDUS1)	*	*P.* (+)	*P.* (+)	*P.* (+)	(−)	(−)	(−)	*P.* (+)	*P.* (+)	*P.* (+)
	Experimentally mixed infections	*P. relictum* (pSGS1 (1)) + *P. relictum* (pGRW4)	*	1 (+)	(−)	1 (+)	(−)	(−)	(−)	(−)	1 (+)	1 (+)
		P. circumflextum (pTURDUS1 (2)) + *P. relictum* (pSGS1)	*	2 (+)	(−)	2 (+)	(−)	(−)	(−)	2 (+)	2 (+)	*
		Plasmodium elongatum (pERIRUB1 (3)) + *P. relictum* (pSGS1)	*	1 (+)	*	1 (+)	(−)	(−)	(−)	3 (+)	3 (+)	3 (+)
		Leucocytozoon sp. (ITUMER01) + *P. relictum* (pSGS1)	*	*	*	*L.* (+)	(−)	*	*	(−)	*L.* (+)	*
		Leucocytozoon sp. (ITUMER01) + *P. relictum* (pGRW4)	*	*	*	*L.* (+)	(−)	*	*	(−)	*L.* (+)	*
		Haemoproteus. parabelopolskyi (hSYBOR1) + *Haemoproteus belopolskyi* (hHIICT1)	*	*	*	*	*	*	*	(−)	*	(−)
		H. tartakovskyi (hHAWF1) + *P. relictum* (pSGS1)	*	*H.* (+)	(−)	*P.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	*P.* (+)	*P.* (+)
		H. lanii (hRB1) + *P. relictum* (pSGS1)	*	*P.* (+)	(−)	*P.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	*P.* (+)	*P.* (+)
		Haemoproteus tartakovskyi (hSISKIN1) + *P. relictum* (pSGS1)	*	*P.* (+)	*	*H.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	*P.* (+)
		Haemoproteus minutus (hTURDUS2) + *P. relictum* (pSGS1)	*	*H.* (+)	*	*H.* (+)	*H.* (+)	(−)	*H.* (+)	(−)	(−)	*P.* (+)
		Haemoproteus. motacillae (hYWT1) + *P. relictum* (pSGS1)	*	*P.* (+)	*	(−)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	(−)
		H. parabeloposkyi (hSYBOR1) + *P. relictum* (pSGS1)	*	*P.* (+)	*	*P.* (+)	*H.* (+)	*H.* (+)	*H.* (+)	(−)	(−)	*P.* (+)
C	Field samples (mixed infections)	*Leucocytozoon* sp. + *Haemoproteus* sp. (OT412)	*	*	(−)	(−)	(−)	(−)	(−)	*	(−)	*
		Leucocytozoon sp. + *Haemoproteus* sp. (PA215)	*	*	(+)	(−)	(+)	*H.* (+)	*	*	*L.* (+)	(−)
		Leucocytozoon sp. + *Haemoproteus* sp. (OT611)	*	*	(+)	(+)	(+)	*H.* (+)	*H.* (+)	*	(−)	*
		Leucocytozoon sp. + *Haemoproteus* sp. (PA220)	*	*	(+)	*H.* (+)	(+)	*H.* (+)	*H.* (+)	*	*L.* (+)	(−)
		Leucocytozoon sp. + *Haemoproteus* sp. (PA287)	*	*	(+)	(−)	(+)	(−)	(−)	*	(−)	*
		Leucocytozoon sp. + *Haemoproteus* sp. (PA046)	*	*	(+)	*H.* (+)	*H.* (+)	(−)	*	*	*L.* (+)	(−)
		Leucocytozoon sp. + *Haemoproteus coatneyi* (OT 490)	*	*	(−)	(−)	(−)	(−)	(−)	*	(−)	*
		Leucocytozoon sp. + *H. coatneyi* (OT632)	*L.* (+)	*L.* (+)	(+)	*L.* (+)	*H.* (+)	*	*	*	*L.* (+)	*
		Leucocytozoon sp. + *H. coatneyi* (PA005)	*	*	(+)	*L.* (+)	(+)	*H.* (+)	*	*	*L.* (+)	(−)
		Leucocytozoon majoris + *H. coatneyi* (PA022)	*	*	(+)	*H.* (+)	(+)	*H.* (+)	*H.* (+)	*	(−)	*
		Leucocytozoon sp. + *Plasmodium homopolare* (OT597)	*	*	(+)	(+)	*	*	*	*P.* (+)	*L.* (+)	(−)
		Leucocytozoon sp. + *P. homopolare* (OT601)	*	*	(+)	(+)	*	*	*	*P.* (+)	*L.* (+)	*
		Leucocytozoon fringillinarum + *Plasmodium* sp. (OT615)	*	*	(+)	*L.* (+)	*	*	*	*P.* (+)	*L.* (+)	(−)
		Leucocytozoon sp. + *Plasmodium unalis* (OT761)	*	*	(+)	(−)	*	*	*	*P.* (+)	(−)	(−)
		L. majoris + *Plasmodium lutzi* (PA047)	*	*	(+)	*H.* (+)	*	*	*	*P.* (+)	*L.* (+)	(−)
		Leucocytozoon sp. + *Haemoproteus* sp. + *Plasmodium* sp. (PA300)	*	*	(+)	*H.* (+)	*H.* (+)	*	*	*P.* (+)	*L.* (+)	*
		Haemoproteus sp. + *Plasmodium* sp. (PA206)	*H.* + *P.* (+)	*H.* + *P.* (+)	(+)	*H.* (+)	*H.* (+)	(−)	*H.* (+)	*P.* (+)	(−)	(−)

(A) Evolutionary dynamics of infectious diseases Laboratory at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University-USA, (B) P. B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania (mixes were produced by mixing extracted DNA, see the Methods), and (C) Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia. (*L*) *Leucocytozoon* spp., (*H*) *Haemoproteus* (*Parahaemoproteus*) spp., and (*P*) *Plasmodium* spp.. (+) PCR with positive and (−) negatives amplifications. *Cox1*, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; and *cytb*, cytochrome b.
* Sample not tested.

Table 3
Sensitivity and specificity of the different sets of primers targeting mtDNA genes designed for both PCR detection and differential DNA amplification of parasites belonging to three genera of avian haemosporidian parasites.

		Haemosporidian detection/phylogenetic analysis				Differential DNA amplification of mixed infection with *Plasmodium* and *Haemoproteus* spp.					
		cytb			*cox3*	*cytb* (H.)			*cytb* (P.) *cytb* (P. + L.)		*cox1* (P.)
Lab.	Measure	298/299	064/066	974/299	959/961	980/982	989/982	HaemF/982	983/985	986/987	971/973
A	Sensitivity	9/9	9/9	9/9	9/9	3/3	3/3	5/5	5/5	2*P.*/5, 2*L.*/2	5/5
	Specificity	NA	NA	NA	NA	6/6	6/6	4/4	4/4	3*H.*/3	4/4
B (SI)	Sensitivity	*	5/5	3/3	4/6	2/2	2/2	3/3	3/3	2*P.*/3, 1*L.*/1	3/3
	Specificity	NA	NA	NA	NA	4/4	3/3	3/3	3/3	2*H.*/2	3/3
B (EMI)	Sensitivity	*	9/9	0/4	10/11	6/6	5/6	2/3	2/3	5*P.*/11, 2*L.*/2	7/8
	Specificity	NA	NA	NA	NA	5/5	3/3	9/9	9/9	7*H.*/7	1/1
C	Sensitivity	23/32	23/32	15/17	12/17	10/12	5/10	7/7	7/7	0*P.*/7, 10*L.*/16	0/5
Overall	Specificity%	32/41 = **78**	46/55 = **83.6**	27/33 = **81.8**	35/43 = **81.4**	21/23 = **91.3**	15/21 = **71.4**	17/18 = **94.4**	17/18 = **94.4**	21*P.*/26 = **80.8**, 15*L.*/21 = **100**	15/21 = **71**
	Sensitivity%	NA	NA	NA	NA	15/15 = **100**	12/12 = **100**	16/16 = **100**	16/16 = **100**	23/23 = **100**	8/8 = **100**

(A) Evolutionary dynamics of infectious diseases Laboratory at the Institute for Genomics and Evolutionary Medicine (iGEM), Temple University-USA, (B) P. B. Šivickis Laboratory of Parasitology, Institute of Ecology, Nature Research Centre, Lithuania (mixes were produced by mixing extracted DNA, see the Methods), and (C) Host-Parasite Relationship Laboratory at Universidad Nacional de Colombia, Colombia. (SI) Single infection and (EMI) experimentally mixed infections. (*L*) *Leucocytozoon* spp., (*H*) *Haemoproteus* (*Parahaemoproteus*) spp., and (*P*) *Plasmodium* spp. NA, do not apply. SI, single infections, EMI, Experimentally mixed infections.
The overall specificity and the sensitivity are shown as a percentage (%) in bold. *Cox1*, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; and *cytb*, cytochrome b.
* Sample not tested.

spp. during mixed infections with *Haemoproteus* (*Parahaemoproteus*) spp. and/or *Leucocytozoon* spp. (Table 2). Their overall sensitivity was 94.4% with 100% specificity. However, the pair of primers AE986/987 amplified DNA of both *Leucocytozoon* and *Plasmodium* spp. In cases of *Leucocytozoon* and *Plasmodium* spp. mixes, these primers were biased towards amplifying *Leucocytozoon* spp. DNA, and in case of *Plasmodium* and *Haemoproteus* spp. mixes, they detected *Plasmodium* spp. (Tables 2 and 3).

The pair of primers AE971/973 only amplified DNA of the *cox1* gene from *Plasmodium* spp. during mixed infections with *Haemoproteus* (*Parahaemoproteus*) spp. and/or *Leucocytozoon* spp. (Table 2) with 100% specificity (Table 3). In addition to this set of primers, a combination of primers, forward AE965 (5′ AAAGTTTTAGGWTTATAYTAYYTATGG 3′) and reverse AE966 (5′ AAGAGARCATAHCATATTCCAWCC ′3), that amplified a 1,296 bp fragment of the *cox1* gene was only tested in laboratory A, using the same samples and PCR protocol but with 54 °C for the annealing temperature. The specificity of this pair of primers was 100% (4/4) amplifying only DNA of *Plasmodium* spp. from bird ($n = 2$) and lizard ($n = 1$) samples, and human samples ($n = 2$) with a sensitivity of 100% (5/5). Therefore, primers AE971/973 and AE965/966 (both targeting *cox1*) can be used for diagnosis and selective amplification of *Plasmodium* spp.. It is worth noting that these two sets of primers also can be used to amplify *Plasmodium* spp. from lizards.

3.3. Nested multiplex PCR for differential DNA amplification of *Haemoproteus* and *Plasmodium* spp. from field isolates

Results from the three laboratories for the nested multiplex PCR for differential DNA amplification of *Haemoproteus*/*Plasmodium* spp. are shown in Fig. 3. All experimental or naturally mixed infections with a similar or different parasitemia of *Plasmodium* and *Haemoproteus* were successfully amplified. Although some unspecific products/background can be observed in the gels as a result of the nested protocol, the products (strong bands) at the proper sizes corresponded to *Haemoproteus*/*Plasmodium* spp. Indeed, given the sizes of the amplicons obtained from multiplexing both pairs of primers AE980/982 and AE983/985, the bands corresponding to each genus could be perfectly separated and excised from an agarose gel (1.5%) when a sample has a mixed infection with an equal or dissimilar parasitemia of *Plasmodium* and *Haemoproteus* spp. (see Table 4 and Fig. 3). Similarly, using the pair of primers AE974/299 for the primary PCR, people from laboratory C were also able to amplify their samples (data not shown). Thus, the three laboratories could successfully amplify and differentiate mixed *Plasmodium* and *Haemoproteus* infections using this single tube PCR multiplex assay.

4. Discussion

The challenges faced by those describing the biodiversity of haemosporidian parasites cannot be equated to a well-defined diagnostic problem such as the differential detection of human malaria parasites. Yet, even when there is a handful of *Plasmodium* species causing human malaria, sensitivity is still an issue in the case of low parasitemias (sub-microscopic infections) or when there are mixed infections (e.g., Demas et al., 2011; Cheng et al., 2015). Thus, the situation is understandably more complex in wildlife biodiversity studies targeting a pool of parasite species with limited taxonomic information in a broad range of host species. Considering this context, a combination of microscopy and PCR assays (including restriction enzyme-based PCR, nested PCR, and quantitative PCR) has been used to detect avian haemosporidian infections and mixed infections, and to characterise parasite prevalence and genetic diversity across different host species and geo-

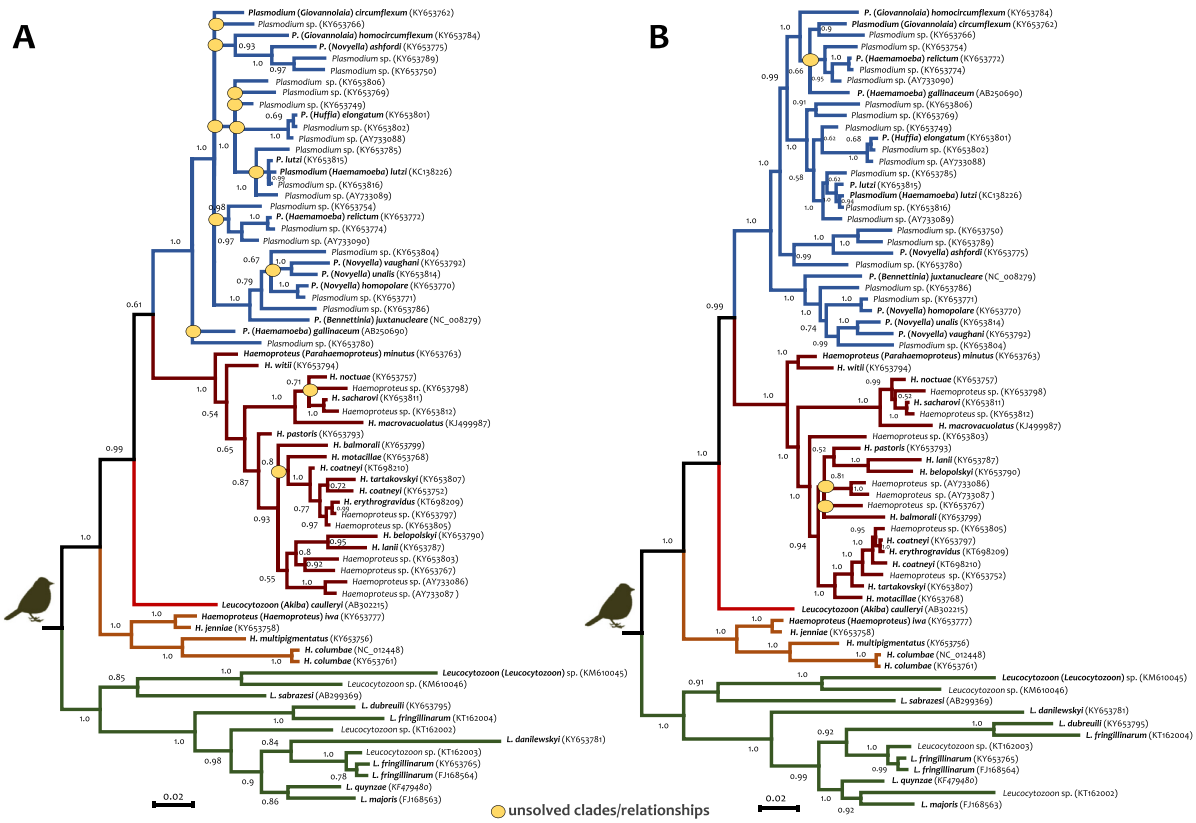


Fig. 4. Bayesian phylogenetic hypotheses of avian haemosporidian parasites based on the partial cytochrome b gene (*cytb*) sequence (70 sequences, 479 bp (A) and 1109 bp (B)). The values at the nodes are posterior probabilities. Branches show the haemosporidian genera. Species names are shown in bold.

graphic locations. However, not surprisingly, the specificity and sensitivity across methods depend on several factors such as the intensity of infection (parasitemia), the combination of parasite lineages present in a sample, and target (parasite) DNA quantity and quality. As a result, failure in detection of a clearly visible and even predominant parasite in blood samples, lineages or mixed infections of *Plasmodium* and *Haemoproteus* spp. has often been reported in avian malaria research (Pérez-Tris and Bensch, 2005; Valkiūnas et al., 2006, 2016; Martínez et al., 2009; Zehtindjiev et al., 2012; Schaer et al., 2015; Bernotienė et al., 2016). The PCR protocols currently in use underestimate haemosporidian mixed infections of different species and genetic lineages of haemosporidian parasites that are predominant in wildlife (Pérez-Tris and Bensch, 2005; Zehtindjiev et al., 2012; Bernotienė et al., 2016). That is expected because PCR assays currently used on avian parasites target conserved regions of mitochondrial genes (e.g., *cytb*) of haemosporidians (Bensch et al., 2000; Richard et al., 2002; Beadell et al., 2004; Hellgren et al., 2004). Here, new sets of haemosporidian mitochondrial primers (especially the *cytb* gene), designed for both PCR detection (longer fragments, >995 bp) and differential DNA amplification of haemosporidian parasite genera are reported (Table 4). It is important to highlight that, in order to fully characterise mixed infections with different species from the same or different genera, methods such as cloning or next-generation sequencing (NGS) are required (e.g., Pacheco et al., 2011; Barbosa et al., 2017). However, the genus-specific primers reported here can mitigate, at least in part, the problem of mixed infections by allowing the detection of two species from different genera. Although these primers have high sensitivity and specificity (in the case of differential DNA amplification), there are important factors that need to be considered when standardising this PCR protocol (or any other) in the context of a specific investigation.

The implementation of a PCR detection protocol should consider the intensity of infections (parasitemias) since it determines the quantity of target DNA (parasite DNA in the whole extraction that is expected to be mostly DNA from the host). Parasitemia patterns can change not only between parasite species and hosts but also geographically. The amount of total DNA has a noticeable effect on the outcome of a PCR procedure since both an excess or insufficient amount of template are the most common causes of failure. Although counterintuitive, the use of too much total DNA template results in false priming and even poor DNA synthesis during the elongation phase of the PCR. On the other hand, when the total amount of the DNA template is extremely low, there is a greater probability of loss due to a number of possible causes such as clotting, adsorption, and/or chemical or enzymatic degradation. Furthermore, a small amount of target DNA increases the risk of contamination from impurities that can get into the PCR mix (Altshuler, 2006).

PCR failure can also occur when the ratio of target DNA (e.g., parasite *cytb*) to non-target DNA (e.g., host DNA) is very low. In this case, the concentration of the target DNA should be considered in the number of cycles used in the reaction. In our hands, using an elevated concentration of the target DNA combined with the normal or higher than normal number of cycles can cause the accelerated accumulation of non-specific products. To avoid this, reducing the number of cycles is highly recommended. Indeed,

Table 4
Summary of the results of PCR assays using different sets of primers for detection/phylogenetic analysis and differential DNA amplification of avian haemosporidian parasites belonging to three genera.

	Gene	Code (pairs)		Notes
Haemosporidian detection/phylogenetic analysis	***cytb***	AE298-F AE299-R AE974-F AE299-R AE064-F AE066-R	All three genera	All 3 primer combinations amplified DNA of parasites belonging to *Leucocytozoon*, *Haemoproteus* and *Plasmodium* spp. Primers AE064/066 can be used as inner primes for a nested PCR. Due to the amplicon size, these new sets of primers can be used for detection and, only in the case of single infection, for phylogenetic analysis.
	cox3	AE959-F AE961-R		These primers amplified DNA of parasites belonging to *Leucocytozoon*, *Haemoproteus* and *Plasmodium* species. This fragment (995 bp) can be used as a new molecular marker.
AE961-R Differential DNA amplification of three avian haemosporidian genera	***cytb***	AE980-F AE982-R	*Haemoproteus*	These primers amplified only DNA of *Haemoproteus* (*Parahaemoproteus*) spp. thus can be used for diagnostics and selective amplification of *Haemoproteus* during mixed infections with *Plasmodium* and *Leucocytozoon* species. This set of primers can be also used for a nested-multiplex PCR with primers AE983/985.
		AE989-F AE982-R		These primers amplified only DNA of *Haemoproteus* (*Parahaemoproteus*) spp. thus can be used for diagnostics and selective amplification of *Haemoproteus* during mixed infections with *Plasmodium* and *Leucocytozoon* species.
		HaemF AE982-R		These primers also amplify only DNA of *Haemoproteus* (*Parahaemoproteus*) spp. HaemF can be used as an external primer for a semi-nested PCR, using primers AE989/AE982 as inner primers.
		AE983-F AE985-R	*Plasmodium*	These primers amplified only DNA of *Plasmodium* spp. thus can be used for diagnostics and selective amplification of Plasmodium during mixed infections with *Haemoproteus* and *Leucocytozoon* species. This set of primers can be used for a nested-multiplex PCR with primers AE980/982.
		AE986-F AE987-R		These primers amplified DNA of *Leucocytozoon* and *Plasmodium* spp. In cases of *Leucocytozoon*/*Plasmodium* spp. mixes, these primers amplified DNA of *Leucocytozoon* spp., and in case of *Plasmodium*/*Haemoproteus* spp. mixes, they detect *Plasmodium* spp.
	cox1	AE971-F AE973-R	*Plasmodium*	This primer pair amplified only DNA of *Plasmodium* spp., thus can be used for diagnostics and selective amplification of *Plasmodium* spp. DNA during mixed infections with *Haemoproteus* and *Leucocytozoon* species. This fragment (507 bp) can be used as a new molecular marker.
		AE965-F AE966-R		This primer pair amplified DNA of parasites belonging to *Plasmodium* species. This fragment might be used as a new molecular marker.

Cox1, cytochrome c oxidase subunit 1; *cox3*, cytochrome c oxidase subunit 3; and *cytb*, cytochrome b.

low concentrations of primer, target, Taq, magnesium, and nucleotides are recommended as these generally ensure cleaner amplification products and lower background (Altshuler, 2006). For blood parasites such as haemosporidians, estimating the right amount of DNA to be used in a PCR assay is problematic. The total amount of the DNA extracted contains DNA from the host (non-target DNA) as well as the parasite (target DNA). Thus, measurements of total DNA mostly correspond to the vertebrate host. Given that the concentration of target DNA is determined by the intensity of the infection or parasitemia, whenever the parasitemia is considered high (by blood-smear microscopy) it is recommended to make serial dilutions, including a non-diluted sample, of the original DNA extraction to do the PCRs. In this study, the dilution of the original extraction improved the PCR sensitivity when parasitemia exceeded 2%. On the contrary, when the parasitemia is very low (not or hardly detectable by blood-smear microscopy), increasing the amount of total DNA is suggested. In the case of a mixed infection, the amount of total target DNA is the result of an unknown ratio of the lineages present in the mix (different parasitemias), so serial dilutions of the total DNA extraction are also recommended. Finally, it is worth mentioning that the yield of the PCR evidenced by the strength of the band observed in an agarose gel is also affected by the parasitemia. Many protocols standardise the amount of PCR product used to observe a band in an agarose gel (e.g., 2 µl); however, low parasitemias will produce an almost imperceptible band with such a small amount of product. A simple way to mitigate this problem is to load the entire PCR product (e.g., 25 or 50 µl) in the gel and then excise the band with the expected size for sequencing if it is needed.

Since DNA extraction is an important stage in molecular detection, it requires a sensitive and cost-effective method. With regard to quantity/quality, the DNA concentration and total yield of extracted DNA vary between the methods. For example, Phenol–Chloroform protocols can extract significantly more concentrated DNA compared with other protocols such as commercial kits without any protocol modifications (Psifidi et al., 2015). Although this protocol yielded highly concentrated DNA, in some cases DNA pellet could be lost and PCR inhibitors can be present. Thus, it is important to consider all these factors before processing field samples, in order to obtain enough target DNA and achieve a successful PCR amplification.

Primers for detection of haemosporidian parsites reported in this investigation (group 1, see Table 1) successfully amplify larger fragments of *cytb* (1109–1741 bp) and *cox3* (761 bp) genes of three genera included in this study. These fragments have more informative sites that can be used for phylogenetic reconstruction methods. Indeed, when a comparison of phylogenies was made using Bayesian approaches with *cytb* fragments (479 bp and 1065 bp, without primer regions), and concatenated *cytb*/*cox3* genes (1826 bp), better results were obtained for both *cytb* (1065 bp) and *cytb*/*cox3* (1826 bp) gene phylogenies with more well-supported clades (Fig. 4, Supplementary Fig. S1). It is worth noting that phylogenetic studies using single genes or the concatenation of *cytb* and *cox3* genes/fragments only can be done when a single infection in the host samples has been confirmed by microscopic, nested PCR, and careful visual inspection of the electropherograms.

Likely because it was the first mitochondrial gene used in a haemosporidian parasite phylogeny (Escalante et al., 1998), a *cytb* fragment (479 bp) has been the marker of choice in ecological, tax-

onomic and phylogenetic investigations of avian malaria parasites. Although usually insufficient for accurate phylogenetic reconstructions due to its limited number of sites (Fig. 4), this fragment has allowed the correct identification of morphologically distinct species currently available (Bensch et al., 2009; Outlaw and Ricklefs, 2014; Lotta et al., 2016; Pacheco et al., 2018). Given that there is no evidence indicating that the haemosporidian *cytb* gene is saturated, it has approximately the same A/T content across species, and exhibits a relatively high substitution rate of evolution (Escalante et al., 1998; Perkins, 2008; Pacheco et al., 2018), this gene (not the small fragment) is a suitable molecular marker to be used in phylogenetic studies (Fig. 4). However, the *cox3* gene could also be a good candidate to be considered for barcoding studies. Indeed, these mitochondrial genes evolved at distinct rates, with *cox3* having the highest substitution rate (0.00474 substitutions/site/million years), followed by *cytb* (0.00419) and *cox1* (0.00371) (Pacheco et al., 2018). In all cases, the use of DNA barcoding approaches requires development of criteria for species delimitation that can link taxa, usually described by using morphology, to molecular data, including understanding the geographic variation of species at the molecular level (Bergsten et al., 2012).

This study showed a sensitive methodology that can be used to estimate the parasite diversity and prevalence in single or mixed infections by two species of different genera. Although mixed infections of haemosporidian parasites are common in wild bird species worldwide, (Valkiūnas et al., 2003, 2006; Beadell et al., 2004; Pérez-Tris and Bensch, 2005; Loiseau et al., 2010; Silva-Iturriza et al., 2012; Dimitrov et al., 2014; González et al., 2014; Lotta et al., 2016; Mantilla et al., 2016), their detection involves methodologies that are difficult to apply in biodiversity research due to parasite lineage/species combinations which are difficult to predict and heterogeneity in their parasitemias. These characteristics are often unique to each field study site and bird population. The situation is particularly difficult in tropical areas where the diversity of haemosporidian parasites is often high and remains insufficiently described (Loiseau et al. 2010; González et al., 2014; Lotta et al., 2016; Mantilla et al., 2016). The genus-specific primers proposed here can mitigate this problem by improving the differential detection of parasites belonging to different genera by PCR assays. Primers successfully tested in this study are promising, although more assays need to be done by the avian malaria community to standardise those. A summary of the results obtained for each genus-specific pair of primers is given in Table 4 and Supplementary Table S3. The tables also indicate whether or not the obtained amplicons overlap with the data available in the Malavi databases (http://mbio-serv2.mbioekol.lu.se/Malavi/, Bensch et al., 2009).

Taking advantage of genus-specific primers reported here, a novel nested multiplex PCR protocol is proposed for differential DNA amplification of *Haemoproteus*/*Plasmodium* spp. Importantly, this protocol provides an opportunity to determine the presence of mixed infections of parasites belonging to these genera in blood samples, avoiding the sequencing stage. Whereas there are apparently more combinations, it is important to consider the primer specificity when preparing a multiplex assay, especially since competition exists when multiple target sequences are in a single reaction vessel. Here, out of all possible pairs of primer combinations tested for the nested multiplex PCR, only the combinations of primers AE980/982 and AE983/985 gave excellent results. These primers (AE980/982 and AE983/985) could be used in a multiplex without nesting the PCR (no primary amplification with AE298/299) if lineages of both species have high parasitemias.

Importantly, the combination of other genus-specific primers gave cross-annealing between primer pairs. As an example, two bands with different molecular sizes were observed in a multiplex experiment with different primers, but both were the same *Plasmodium* spp. as corroborated by sequencing. Nevertheless, the multiplex PCR assay proposed here using primers AE980/982 and AE983/985 performed well across laboratories, providing an inexpensive, fast and easy method that not only can be used to detect *Haemoproteus* and *Plasmodium* parasites but also mixed infection with these species at the same time. The expenses of reagents and preparation time is less in multiplex PCR than in systems where several tubes of single PCRs are used. A multiplex reaction is ideal for conserving templates in short supply. Another advantage of a nested multiplex PCR is that false negatives are often revealed in multiplex assays because each amplicon provides an internal control for the other amplified fragments. However, it is worth noting that nested PCRs (single or multiplex) are prone to contamination simply because aerosol DNAs, which otherwise are not detected, could be accidentally amplified. The development of rigorous cleaning protocols and good practices such as including multiple negative controls are especially important whenever nested PCR protocols are implemented.

The results from a nested multiplex PCR such as the one proposed here can help to identify whether all haemosporidian blood stage parasites observed in a blood-smear correspond to single or different genera, avoiding parasite misidentification, particularly at the stage of young gametocytes and/or trophozoites of *Plasmodium* and *Haemoproteus* spp. This is an important diagnostic issue in wildlife when only young blood stages are available in blood smears (Valkiūnas, 2005) and/or new species are present. It is difficult and often even impossible to distinguish young blood stages of parasites belonging to these two genera under a light microscope during mixed infections. It is worth noting that detection of mixed infection with different lineages of the same genus is still challenging and it only can be done by methods which are more expensive and laborious such as cloning or NGS (e.g., Pacheco et al., 2011; Barbosa et al., 2017). If cloning is the method of choice, it is recommended to clone two or three independent PCR products using reagents of high quality and sequencing more than three clones for PCR products to reduce the problems associated with this technique (see Pérez-Tris and Bensch, 2005). NGS is likely the future preference for parasite biodiversity studies. However, it requires development of suitable protocols and bioinformatic expertise (e.g., Barbosa et al., 2017). We anticipate that the proposed primers could be adapted for NGS target deep-sequencing approaches.

In summary, new sets of haemosporidian mitochondrial primers (especially the *cytb* gene), designed for both PCR detection (longer fragments, >995 bp) and differential DNA amplification of haemosporidian parasite genera were tested under different laboratory working conditions (equipment, reagents) with promising results (Table 4, Supplementary Table S3). These primers have high sensitivity and specificity, and the novel nested-multiplex PCR protocol could be an excellent tool for detection and characterisation of haemosporidian infections. More importantly, the amplicons obtained using these primers also overlap with the data that is already available in the different databases, allowing the comparison of new data with those sequences already available.

Acknowledgements

This work was supported in part by the US National Institutes of Health (grant R01 GM080586 to AAE, Temple University, USA), Research Council of Lithuania (grant MIP-045/2015 to GV), and by the División de Investigación y Extensión of the Universidad Nacional de Colombia (grant 37416 to AGC and NEM). Tatjana A. Iezhova, Vaidas Palinauskas, Dovilė Bukauskaitė and Mikas Ilgūnas are acknowledged for participation in field work and assistance in the laboratory in Vilnius, Lithuania. We thank all the students of the Host-Parasite Relationship Research Group at Universidad

Nacional de Colombia, and the people from the DNA Laboratory at the School of Life Sciences (Arizona State University, USA) for their technical support during the sequencing process.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijpara.2018.02.003.

References

Altshuler, M.L., 2006. PCR Troubleshooting: The Essential Guide. Caister Academic Press, United Kingdom.

Barbosa, A.D., Gofton, A.W., Paparini, A., Codello, A., Greay, T., Gillett, A., Warren, K., Irwin, P., Ryan, U., 2017. Increased genetic diversity and prevalence of co-infection with Trypanosoma spp. in koalas (*Phascolarctos cinereus*) and their ticks identified using next-generation sequencing (NGS). PLoS One 12, e0181279.

Beadell, J.S., Gering, E., Austin, J., Dumbacher, J.P., Peirce, M.A., Pratt, T.K., Atkinson, C.T., Fleischer, R.C., 2004. Prevalence and differential host-specificity of two avian blood parasite genera in the Australo-Papuan region. Mol. Ecol. 13, 3829–3844.

Bensch, S., Hellgren, O., Pérez-Tris, J., 2009. MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. Mol. Ecol. Resour. 9, 1353–1358.

Bensch, S., Pérez-Tris, J., Waldenström, J., Hellgren, O., 2004. Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? Evolution 58, 1617–1621.

Bensch, S., Stjernman, M., Hasselquist, D., Ostman, O., Hansson, B., Westerdahl, H., Pinheiro, R.T., 2000. Host specificity in avian blood parasites: a study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. Proc. Biol. Sci. 267, 1583–1589.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The effect of geographical scale of sampling on DNA barcoding. Syst. Biol. 61, 851–869.

Bernotienė, R., Palinauskas, V., Iezhova, T., Murauskaitė, D., Valkiūnas, G., 2016. Avian haemosporidian parasites (Haemosporida): A comparative analysis of different polymerase chain reaction assays in detection of mixed infections. Exp. Parasitol. 163, 31–37.

Cheng, Q., Cunningham, J., Gatton, M.L., 2015. Systematic review of sub-microscopic *P. vivax* infections: prevalence and determining factors. PLoS Negl. Trop. Dis. 9, e3413.

Clark, N.J., Wells, K., Dimitrov, D., Clegg, S.M., 2016. Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. J. Anim. Ecol. 85, 1461–1470.

Demas, A., Oberstaller, J., DeBarry, J., Lucchi, N.W., Srinivasamoorthy, G., Sumari, D., Kabanywanyi, A.M., Villegas, L., Escalante, A.A., Kachur, S.P., Barnwell, J.W., Peterson, D.S., Udhayakumar, V., Kissinger, J.C., 2011. Applied genomics: data mining reveals species-specific malaria diagnostic targets more sensitive than 18S rRNA. J. Clin. Microbiol. 49, 2411–2418.

Dimitrov, D., Palinauskas, V., Iezhova, T.A., Bernotienė, R., Ilgūnas, M., Bukauskaitė, D., Zehtindjiev, P., Ilieva, M., Shapoval, A.P., Bolshakov, C.V., Markovets, M.Y., Bensch, S., Valkiūnas, G., 2015. *Plasmodium* spp.: an experimental study on vertebrate host susceptibility to avian malaria. Exp. Parasitol. 148, 1–16.

Dimitrov, D., Zehtindjiev, P., Bensch, S., Ilieva, M., Iezhova, T., Valkiūnas, G., 2014. Two new species of *Haemoproteus* Kruse, 1890 (Haemosporida, Haemoproteidae) from European birds, with emphasis on DNA barcoding for detection of haemosporidians in wildlife. Syst. Parasitol. 87, 135–151.

Escalante, A.A., Freeland, D.E., Collins, W.E., Lal, A.A., 1998. The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. Proc. Natl. Acad. Sci. U.S.A. 95, 8124–8129.

Falk, B.G., Mahler, D.L., Perkins, S.L., 2011. Tree-based delimitation of morphologically ambiguous taxa: a study of the lizard malaria parasites on the Caribbean island of Hispaniola. Int. J. Parasitol. 41, 967–980.

Garnham, P.C.C., 1966. Malaria Parasites and Other Haemosporidia. Blackwell Scientific Publications, Oxford.

González, A.D., Matta, N.E., Ellis, V.A., Miller, E.T., Ricklefs, R.E., Gutiérrez, H.R., 2014. Mixed species flock, nest height, and elevation partially explain avian haemoparasite prevalence in Colombia. PLoS One 9, e100695.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hellgren, O., Waldenström, J., Bensch, S., 2004. A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. J. Parasitol. 90, 797–802.

Ishtiaq, F., Rao, M., Huang, X., Bensch, S., 2017. Estimating prevalence of avian haemosporidians in natural populations: a comparative study on screening protocols. Parasit. Vectors. 10, 127.

Jennings, W.B., 2017. Phylogenomic Data Acquisition: Principles and Practice. CRC Press, Taylor & Francis Group, Boca Raton, Florida.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Loiseau, C., Iezhova, T., Valkiūnas, G., Chasar, A., Hutchinson, A., Buermann, W., Smith, T.B., Sehgal, R.N., 2010. Spatial variation of haemosporidian parasite infection in African rainforest bird species. J. Parasitol. 96, 21–29.

Lotta, I.A., Pacheco, M.A., Escalante, A.A., González, A.D., Mantilla, J.S., Moncada, L.I., Adler, P.H., Matta, N.E., 2016. *Leucocytozoon* diversity and possible vectors in the Neotropical highlands of Colombia. Protist 167, 185–204.

Mantilla, J.S., González, A.D., Lotta, I.A., Moens, M., Pacheco, M.A., Escalante, A.A., Valkiūnas, G., Moncada, L.I., Pérez-Tris, J., Matta, N.E., 2016. *Haemoproteus erythrogravidus* n. sp. (Haemosporida, Haemoproteidae): Description and molecular characterization of a widespread blood parasite of birds in South America. Acta Trop. 159, 83–94.

Martínez, J., Martínez-De La Puente, J., Herrero, J., Del Cerro, S., Lobato, E., Rivero-De Aguilar, J., Vásquez, R.A., Merino, S., 2009. A restriction site to differentiate *Plasmodium* and *Haemoproteus* infections in birds: on the inefficiency of general primers for detection of mixed infections. Parasitology 136, 713–722.

Martinsen, E.S., Perkins, S.L., Schall, J.J., 2008. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. Mol. Phylogenet. Evol. 47, 261–273.

Muehlenbein, M.P., Pacheco, M.A., Taylor, J.E., Prall, S.P., Ambu, L., Nathan, S., Alsisto, S., Ramirez, D., Escalante, A.A., 2015. Accelerated diversification of nonhuman primate malarias in Southeast Asia: adaptive radiation or geographic speciation? Mol. Biol. Evol. 32, 422–439.

Nilsson, E., Taubert, H., Hellgren, O., Huang, X., Palinauskas, V., Markovets, M.Y., Valkiūnas, G., Bensch, S., 2016. Multiple cryptic species of sympatric generalists within the avian blood parasite *Haemoproteus majoris*. J. Evol. Biol. 29, 1812–1826.

Outlaw, D.C., Ricklefs, R.E., 2014. Species limits in avian malaria parasites (Haemosporida): how to move forward in the molecular era. Parasitology 141, 1223–1232.

Pacheco, M.A., Cranfield, M., Cameron, K., Escalante, A.A., 2013. Malarial parasite diversity in chimpanzees: the value of comparative approaches to ascertain the evolution of *Plasmodium falciparum* antigens. Malar. J. 12, 328.

Pacheco, M.A., Escalante, A.A., Garner, M.M., Bradley, G.A., Aguilar, R.F., 2011. Haemosporidian infection in captive masked bobwhite quail (*Colinus virginianus ridgwayi*), an endangered subspecies of the northern bobwhite quail. Vet. Parasitol. 182, 113–120.

Pacheco, M.A., Matta, N.E., Valkiūnas, G., Parker, P.G., Mello, B., Stanley Jr., C.E., Lentino, M., Garcia-Amado, M.A., Cranfield, M., Kosakovsky Pond, S.L., Escalante, A.A., 2018. Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. Mol. Biol. Evol. 35, 383–403.

Pacheco, M.A., Reid, M.J., Schillaci, M.A., Lowenberger, C.A., Galdikas, B.M., Jones-Engel, L., Escalante, A.A., 2012. The origin of malarial parasites in orangutans. PLoS One 7, e34990.

Pérez-Tris, J., Bensch, S., 2005. Diagnosing genetically diverse avian malarial infections using mixed-sequence analysis and TA-cloning. Parasitology 131, 15–23.

Perkins, S.L., 2008. Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: corroborative and new evidence for the origins of human malaria. Mitochondrial DNA 19, 471–478.

Perkins, S.L., Schall, J.J., 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. J. Parasitol. 88, 972–978.

Poulin, R., 2007. Evolutionary Ecology of Parasites. Princeton University Press, Princeton.

Psifidi, A., Dovas, C.I., Bramis, G., Lazou, T., Russel, C.L., Arsenos, G., Banos, G., 2015. Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. PLoS One 10, e0115960.

Richard, F.A., Sehgal, R.N., Jones, H.I., Smith, T.B., 2002. A comparative analysis of PCR-based detection methods for avian malaria. J. Parasitol. 88, 819–822.

Richardson, D.S., Jury, F.L., Blaakmeer, K., Komdeur, J., Burke, T., 2001. Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). Mol. Ecol. 10, 2263–2273.

Ricklefs, R.E., Fallon, S.M., 2002. Diversification and host switching in avian malaria parasites. Proc. Biol. Sci. 269, 885–892.

Ricklefs, R.E., Fallon, S.M., Bermingham, E., 2004. Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. Syst. Biol. 53, 111–119.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Sawabe, K., Isawa, H., Hoshino, K., Sasaki, T., Roychoudhury, S., Higa, Y., Kasai, S., Tsuda, Y., Nishiumi, I., Hisai, N., Hamao, S., Kobayashi, M., 2010. Host-feeding habits of *Culex pipiens* and *Aedes albopictus* (Diptera: Culicidae) collected at the urban and suburban residential areas of Japan. J. Med. Entomol. 47, 442–450.

Schaer, J., Reeder, D.M., Vodzak, M.E., Olival, K.J., Weber, N., Mayer, F., Matuschewski, K., Perkins, S.L., 2015. *Nycteria* parasites of Afrotropical insectivorous bats. Int. J. Parasitol. 45, 375–384.

Silva-Iturriza, A., Ketmaier, V., Tiedemann, R., 2012. Prevalence of avian haemosporidian parasites and their host fidelity in the central Philippine islands. Parasitol. Int. 61, 650–657.

Telford Jr., S.R., 2009. Hemoparasites of the Reptilia: Color Atlas and Text. CRC Press, Taylor & Francis Group, Boca Raton, Florida.

Valkiūnas, G., 2005. Avian Malaria Parasites and Other Haemosporidia. CRC Press, Boca Raton, Florida.
Valkiunas, G., Bensch, S., Iezhova, T.A., Krizanauskiené, A., Hellgren, O., Bolshakov, C. V., 2006. Nested cytochrome b polymerase chain reaction diagnostics underestimate mixed infections of avian blood haemosporidian parasites: microscopy is still essential. J. Parasitol. 92, 418–422.
Valkiūnas, G., Iezhova, T.A., Shapoval, A.P., 2003. High prevalence of blood parasites in hawfinch *Coccothraustes coccothraustes*. J. Nat. Hist. 37, 2647–2652.
Valkiūnas, G., Iezhova, T.A., 2017. Exo-erythrocytic development of avian malaria and related haemosporidian parasites. Malar. J. 16, 101.
Valkiūnas, G., Ilgūnas, M., Bukauskaitė, D., Iezhova, T.A., 2016. Description of *Haemoproteus ciconiae* sp. nov. (Haemoproteidae, Haemosporida) from the white stork *Ciconia ciconia*, with remarks on insensitivity of established polymerase chain reaction assays to detect this infection. Parasitol. Res. 115, 2609–2616.
Valkiūnas, G., Ilgūnas, M., Bukauskaitė, D., Palinauskas, V., Bernotienė, R., Iezhova, T. A., 2017. Molecular characterization and distribution of *Plasmodium matutinum*, a common avian malaria parasite. Parasitology 144, 1726–1735.
Zehtindjiev, P., Križanauskienė, A., Bensch, S., Palinauskas, V., Asghar, M., Dimitrov, D., Scebba, S., Valkiūnas, G., 2012. A new morphologically distinct avian malaria parasite that fails detection by established polymerase chain reaction-based protocols for amplification of the cytochrome B gene. J. Parasitol. 98, 657–665.