Comparative analysis of mitochondrial DNA in haemogregarine (Apicomplexa: Adeleorina) blood parasites

T Haarhoff

Dissertation accepted in fulfilment of the requirements for the degree *Master of Environmental Sciences* at the North-West University

Supervisor: Dr EC Netherlands
Co-Supervisor: Prof L du Preez
Co-Supervisor: Dr M Pretorius

Graduation May 2024
30766095

ACKNOWLEDGEMENTS

I dedicate this dissertation to my God, Jesus, who has helped and guided me, and opened doors for me to complete this study that would otherwise have never happened.

I would like to thank my supervisor, Dr Ed Netherlands, for laying the foundation of my laboratory skills and shaping me into a scientist. Your guidance and wit were extraordinary. I would like to thank Prof Louis for all the effort you put in for me to make my studies a success and memories to look back on, there is no study leader like yourself. I would also like to thank my co-supervisor, Dr Marrianne Pretorious, for your help and support and for allowing me to spend all the project funds.I would like to thank my dad for funding my years of studies and always believing in me. What you've sacrificed to get me where I am will never be forgotten. Thank you to my mom who is an inspiration for me to NEVER give up.

I would like to thank Willie Landman for all the support and care extended to me throughout my studies. Your advice at critical points have helped steer this study towards its success.

Lastly, I would like to thank my sweetheart, Kayla Viviers. Your support, encouragement and love are intertwined throughout the words and pages of this paper. Thank you for listening to me when I rant about the difficulties of phylogenies and stubborn experiments. Your unfailing attempt to understand my work and conversations about the difficulties have helped me come to most of the breakthroughs I needed.

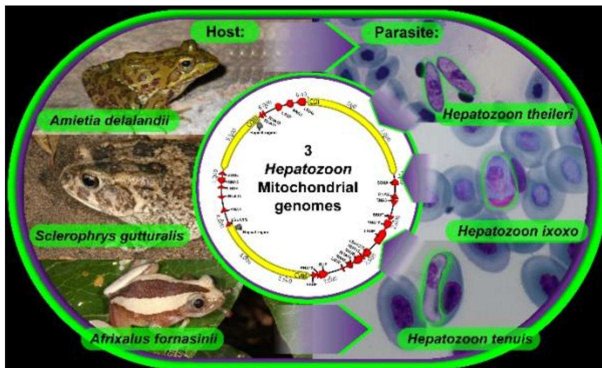
The North-West University and Unit of Environmental Sciences are thanked for this opportunity and the use of their facilities.

Ethics declarations. This study received the relevant ethical approval from the North-West University's AnimCare ethics committee (ethics number: NWU-01399-23-A9).

Financial support. The study was financially supported by the South African National Biodiversity Institute (SANBI) and the National Research Foundation (NRF), Grant No: 129114. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the funding bodies. The NWU is thanked for providing financial support in 2022 and 2023 under the NWU Masters bursary.

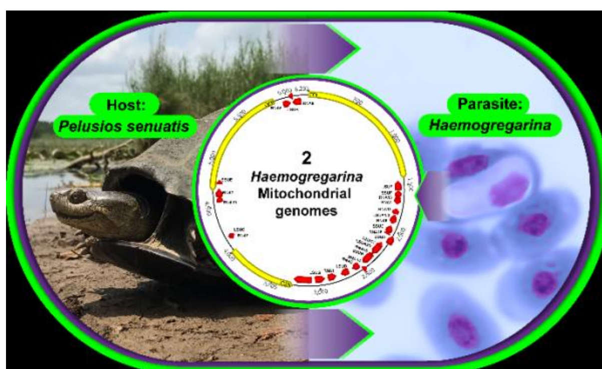
Abstract

Chapter 2.1:



Southern Africa is recognised as one of the most biodiverse regions globally, with a continual expansion of biodiversity marked by the continuous discovery of new parasites parasitising the country's rich fauna. Species of *Hepatozoon* (Adeleorina: Hepatozoidae) comprises a substantial group of parasites, with well over 300 species. Among Southern African frogs, these haemogregarines stand as the most common blood parasites. The considerable diversity within the genus presents a notable phylogenetic challenge due to the limited resolution of the universal 18S rDNA marker in classifying such large groups. The protein-coding genes of the mitochondrial genome have been proposed as markers with higher phylogenetic resolution. In Chapter 2.1 the mitochondrial genomes of three *Hepatozoon* spp. infecting South African anurans was sequenced through Sanger-sequencing by means of primer walking. The present study yielded three complete mitochondrial genomes, each approximately 6500 base pairs in length, encoding the three protein-coding genes commonly observed in apicomplexan mitochondrial genomes. Supplementary genetic markers, including the ITS-1 region and the 23S pl-rDNA were tested however, the primers used for these markers seem ineffective for the haemogregarines of this study. In conclusion, this study confirms the effectiveness of Sanger sequencing through primer walking for mitochondrial genome sequencing. Furthermore, it underscores the need for improved primers designed specifically for the ITS-1 and 23S markers.

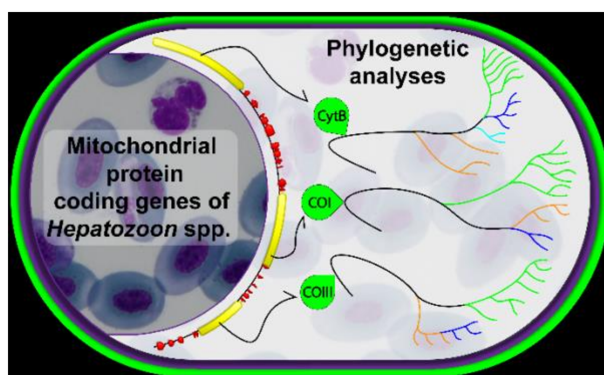
Chapter 2.2:



The mitochondrial protein-coding genes in species of *Haemogregarina* (Adeleorina: Haemogregarinidae) have been examined, revealing remarkable variation even among closely related members. Recent research suggests that these protein-coding genes could be valuable for understanding the evolutionary relationships of haemogregarines. However, the unexpectedly high diversity observed among *Haemogregarina* spp. might have broader implications. In Chapter 2.2, the full mitochondrial genomes of two *Haemogregarina* spp. were

sequenced with Sanger-sequencing by means of primer walking. The two mitochondrial genomes generated resemble that of other *Haemogregarina* spp. The protein-coding genes' high sequence diversity is consistent with that found in previous studies. Additionally, like other haemogregarine protein-coding genes the COIII gene has the highest sequence diversity compared to the COI and CytB genes. However, it remains unclear whether the sequence diversity of *Haemogregarina* mitochondrial protein-coding genes reduce their resolution or usefulness in phylogenetic analyses. More mtDNA data will need to be generated for the COI and CytB genes of other *Haemogregarina* spp. for comprehensive phylogenetic studies to be conducted.

Chapter 2.3:



Of the Apicomplexans, the genus *Hepatozoon* (Adeleorina: Haemogregarinidae) has mostly been placed on the parasite back burner in favour of other more medically important parasites. However, *Hepatozoon* have been recorded in all vertebrate classes and merit more attention. Nevertheless, *Hepatozoon* spp. are major herpetological parasites, accounting for the majority of reported infections in Southern Africa. Currently, the genus comprises of more than 300 species worldwide of which the phylogenetic placement of certain species have been somewhat of a debate. Although it is evident that the genus needs to be subdivided into multiple genera, the methodology for precisely delineating these divisions remains uncertain. The cytochrome *c* oxidase subunit I (COI) mitochondrial protein-coding gene sequence has been proposed as an alternative genetic marker to the traditional nuclear 18S rRNA gene sequence. The gene has proven useful in delineating recent speciation events in other apicomplexan parasites. In this study the mitochondrial protein-coding genes COI, cytochrome *c* oxidase subunit III (COIII), and cytochrome B (CytB) were amplified and sequenced to analyse their ability to resolve recent speciation events within *Hepatozoon*. A total of 34 protein-coding gene sequences were generated and analysed. The investigation revealed that combining the COI and COIII gene sequences proves most effective in resolving the deeper nodes of the phylogenetic tree. Combining these two gene sequences with the 18S and CytB sequences further aid in resolving these deeper nodes. More data will be needed to determine whether the same conclusions is reached with a larger pool of samples. Considering the challenges posed by the high sequence variability of the COIII gene, it may prove beneficial to undertake whole mitochondrial genome sequencing of *Hepatozoon* spp., particularly given the difficulties encountered in designing primers for the COIII region.

Key Words: *Hepatozoon, Haemogregarina,* haemogregarines, mitochondrial genome, mitochondrial protein-coding genes, phylogenetic analysis

TABLE OF CONTENTS

ABSTRACT **1**

PREFACE **8**

CHAPTER 1: GENERAL INTRODUCTION **9**

Introduction. 10

CHAPTER 2: ARTICLES **17**

Chapter 2.1:

Generating the full mitochondrial genomes through primer walking of *Hepatozoon ixoxo, Hepatozoon tenuis, and Hepatozoon theileri* (Adeleorina: Hepatozoidae), parasitising South African Anurans. 18

Chapter 2.2:

The complete mitochondrial genomes of two *Haemogregarina* spp. (Adeleorina: Haemogregarinidae), parasitising the Serrated Hinged Terrapin *Pelusios sinuatis* of Southern Africa. 35

Chapter 2.3

A comparative phylogenetic analysis of mitochondrial protein-coding genes from *Hepatozoon* spp. (Adeleorina: Hepatozoidae), parasitising South African vertebrates 51

CHAPTER 3: GENERAL DISCUSSION **82**

Discussion 83

REFERENCE LIST **87**

APPENDIX A: SUPPLEMENTARY DATA **97**

TABLES & FIGURES

CHAPTER 1

Figure 1.1: The taxonomic classification of Apicomplexa according to The Taxonomicon (Brands, 2023). 10

CHAPTER 2.1

Table 1: Primer information and conditions of PCR reactions performed in this study. 25

Table 2: Percentage nucleotide identity between some adeleorinid COI genes (displayed in bold) and CytB gene (displayed as normal text) sequences. 29

Figure 2.1.1: Complete circular mitochondrial genomes of *Hepatozoon theileri* ex *Amietia delalandii*, *Hepatozoon ixoxo* ex *Schlerophrys gutturalis* and *Hepatozoon tenuis* ex *Afrixalus fornasini*. 27

Figure 2.1.2: The mitochondrial genome structure of *H. ixoxo, H. theileri*, and *H. tenuis* compared against those of other *Hepatozoon* spp. and *Klossia razorbacki*. 30

Figure 2.1.3: The primer-walking technique used to sequence the whole mt-genomes of *Hepatozoon* spp. of this study is displayed in the above diagram. 33

CHAPTER 2.2

Table 3: Primers used to generate the sequences of the 18S and mitochondrial genome fragments. 41

Figure 2.2.1: The mitochondrial genome of two *Haemogregarina* spp. parasitising the serrated hinged terrapin *Pelusios sinuatus*. 43

Figure 2.2.2: This illustration compares the organisation of available mitochondrial genomes of haemogregarines in adeleorina. 44

Figure 2.2.3: The CDS genes of the two mt-genomes generated in this study is aligned individually with 93% similarity cost matrix. 44

Figure 2.2.4: The 18S rDNA sequence of multiple adeleorinid haemoparasites was used to construct the phylogenetic tree above. 45

Figure 2.2.5: Top: The COI protein-coding genes of *Hepatozoon clamantae* and *Hepatozoon catesbianae* is compared. 48

CHAPTER 2.3

Table 4: 34 Mitochondrial protein-coding gene sequences were generated from 8 *Hepatozoon* spp. and are listed. Isolate 27A1 has two COIII fragments generated and are listed within one cell, one atop the other. 61

Table 5: Phylogenetic analysis of genetic markers 18S, COI, COIII, and CytB. 60

Table 6: Results of the Partition homogeneity (ILD) test for both conditions A and B of each genetic markers compared. 63

Table 7: Primers designed. 73

Figure 2.3.1: Patristic distance graphs. 64

Figure 2.3.2: The 18S rDNA phylogenetic tree includes species of *Hepatozoon* available on Genbank. 65

Figure 2.3.3: The 18S COI phylogenetic tree of species of *Hepatozoon* represented in this study. 67

Figure 2.3.4: The 18S COIII phylogenetic tree of species of *Hepatozoon* represented in this study. 68

Figure 2.3.5: The 18S CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 69

Figure 2.3.6: The 18S-COI-COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 70

CHAPTER 3

Figure 3.1: An illustration of taxa within adeleorina, the mitochondrila DNA that has been gathered and those that are not yet available. 85

APPENDIX A: SUPPLEMENTARY DATA

CHAPTER 2.1

Supp. Table 1: COI and RNA fragments of the mitochondrial genomes generated for *Hepatozoon ixoxo, H. theileri* and *H. tenuis*. 97

CHAPTER 2.2

Supp. Table 2: Annotation of fragmented ribosomal RNA (rRNA) genes encoded in the mitochondrial genomes of Haemogregarina sp. A and B based on pairwise comparisons with conserved annotated genes from Klossiella equi (MH203050), and *Hepatozoon catesbianae* (KF894962). 98

CHAPTER 2.3

Supp. Table 3: Some of the sequence data used to construct the phylogenetic trees of this study was extracted from Genbank and are listed. 99

Supp. Figure 1: The COI phylogenetic tree of species of *Hepatozoon* represented in this study. 101

Supp. Figure 2: The COIII phylogenetic tree of species of *Hepatozoon* represented in this study. 102

Supp. Figure 3: The CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 103

Supp. Figure 4: The COI-COIII phylogenetic tree of species of *Hepatozoon* represented in this study. 104

Supp. Figure 5: The COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 105

Supp. Figure 6: The COI-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 106

Supp. Figure 7: The COI-COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. 107

PREFACE

This dissertation is written to be submitted in the article format of the journal *Parasitology* (Cambridge).

Articles within this dissertation:

- **Chapter 2.1: (Submission in progress)** - Generating the full Mitochondrial Genomes through Primer walking of *Hepatozoon ixoxo, Hepatozoon tenuis, and Hepatozoon theileri* (Adeleorina: Hepatozoidae), Parasitising South African Anurans.
- **Chapter 2.2: (Submission in progress)** - The Complete Mitochondrial Genomes of two *Haemogregarina* spp. (Adeleorina: Haemogregarinidae), Parasitising the serrated hinged terrapins *Pelusios senuatis* of Southern Africa.
- **Chapter 2.3:** A comparative phylogenetic analysis of mitochondrial protein-coding genes from *Hepatozoon* spp. (Adeleorina: Hepatozoidae), parasitising South African vertebrates.

The student was first author, wrote, designed, conceived, and analysed the data for all the above articles. The co-authors are aware of and have given permission for the articles to be submitted for degree purposes.

CHAPTER 1

GENERAL INTRODUCTION

1.1. Introduction

The phylum Apicomplexa is home to a large group of successful unicellular protists that have adeptly formed complex host-parasite interactions across a broad spectrum of hosts or environments. As obligate intracellular parasites, these protists possess an apical complex, a distinctive organelle characteristic, although not restricted to Apicomplexans (hence the name) (Okamoto and Keeling, 2014). Consensus on the taxonomy of Apicomplexa is somewhat disordered in both the literature and online sources, and attempts to find sources for accurate classification proved difficult. Although Levine (1988) stands as the most frequently cited reference for Apicomplexan taxonomy, it is more than 30 years old. For the present study, the taxonomy of Apicomplexa aligns with the classification outlined in a recent review of the phylum (infraphylum) by Cavalier-Smith (2014).

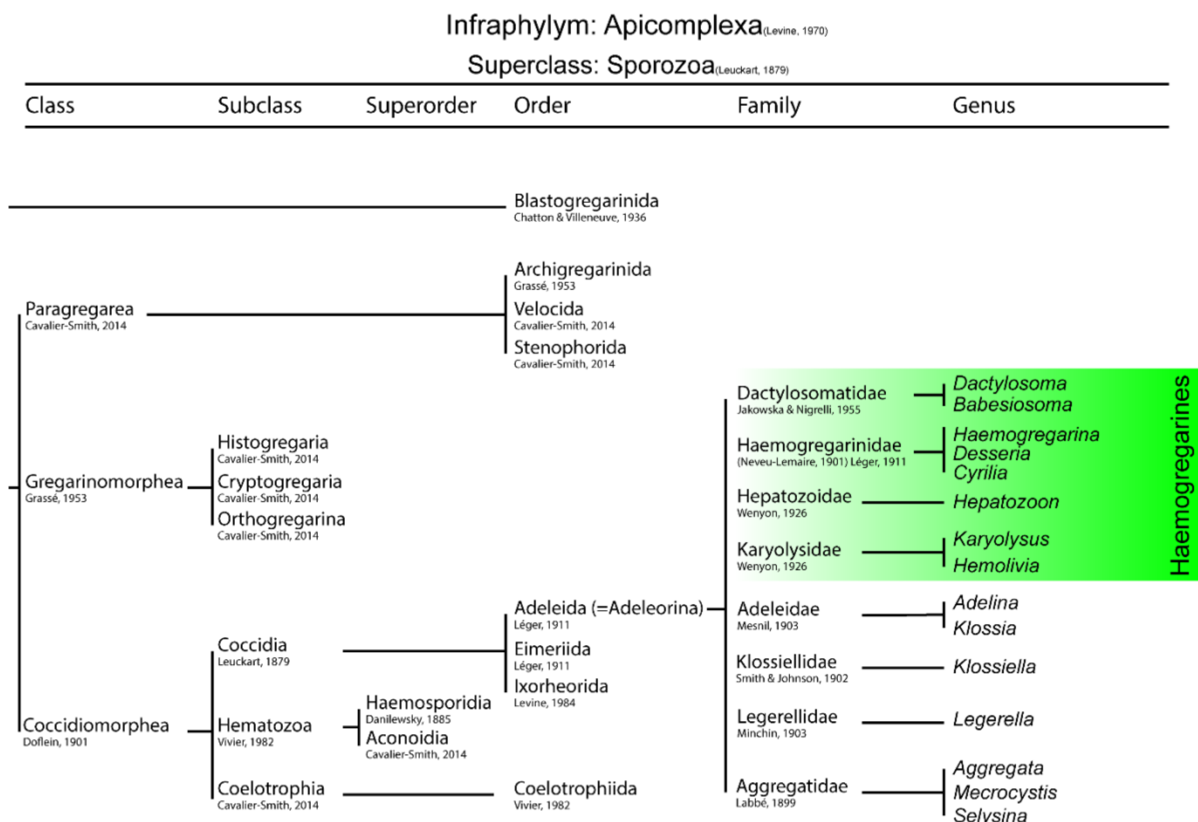


Figure 1.1.1.1: The taxonomic classification of Apicomplexa according to The Taxonomicon (Brands, 2023). Parasites being studied are from the order Adeleorina and undergo a heteroxenous life cycle, known as haemogregarines, highlighted in green.

In this review of the taxonomy of gregarines, Cavalier-Smith moves the suborder Adeleorina to the rank of order and its name to Adeleida, Léger, 191. Cavalier-Smith's interpretation of the taxonomy of Apicomplexa is pragmatic due to his use of molecular phylogenetics and morphological characteristics for the classification and identification of groups classify. The taxonomy of Apicomplexa according to Cavalier-Smith (2014) and supported by

Ruggiero *et al*. (2015) is presented in Figure 1.1. With this taxonomic tree the Apicomplexans focused on in this study are introduced.

Within the class Coccidiomorphea Doflein, 1901, subclass Coccidia Leukart, 1879, order Adeleorina Léger, 1911, lies a group of heteroxenous haemoparasites known as the haemogregarines. These haemogregarines are classified into four families: Dactylosomatidae Jakowska and Nigrelli, 1955, Haemogregarinidae Léger, 1911, Hepatozoidae Miller, 1908, and Karyolysidae Labbé, 1894. This study will mainly focus on species of *Haemogregarina* and *Hepatozoon* from Haemogregarinidae and Hepatozoidae, respectively. The most specious of these four families is Hepatozoidae (over 310 species). However, this has not always been the case. Nevertheless, historical records reveal that, *Haemogregarina* was a reservoir for poorly described taxa (Barta, 2000) with over 300 species represented; however, only species isolated from turtles had full life cycle descriptions of which *Haemogregarina stepanowi* was consequently the type species of the genus (Siddall, 1995). This was problematic since the host range of *Haemogregarina* spp. at that time covered all vertebrate groups. A systematic review by Siddall in 1995 saw the first attempt to resolve some of these issues. As a result, Siddall erected the genus *Desseria* Siddall, 1955, to accommodate marine fish haemogregarines that lack erythrocytic merogony. Thirteen species parasitising marine fish were retained as *Haemogregarina* (*sensu lato*) due to uncertainty regarding their genus placement, attributed to their possession of erythrocytic merogony—a feature absent in *Desseria* spp. Subsequently, all remaining *Haemogregarina* spp. that infect birds, crocodilians, lizards, and snakes were transferred to *Hepatozoon*. Additionally, Desser *et al*. (1995) proposed the reassignment of all anuran *Haemogregarina* to the genus *Hepatozoon*. Subsequently, Smith (1996) reassigned additional unresolved *Haemogregarina* spp. along with species from other genera to the *Hepatozoon*. Presently, species of *Hepatozoon* infect a wide range of vertebrate hosts, ranging from mammals, amphibians, and reptiles, to birds and fish (Smith, 1996; Barta *et al.,* 2012; Cardoso *et al.,* 2022). On the other hand the genus *Haemogregarina* now comprises species that infect turtles and fish (Davies *et al.,* 2004; Telford *et al.,* 2009: Esteves-Siva *et al.,* 2019).

Scientific consensus suggest that the genus *Hepatozoon* should undergo subdivision into multiple genera (Siddall, 1995; Smith, 1996; Smith & Desser, 1997; Barta *et al*., 2012; Maia *et al.,* 2016a; Hrazdilová *et al.,* 2021).Traditionally, morphological characteristics were used to determine the phylogenetic relationships of organisms and to subsequently resolve the taxonomy of groups (Siddall, 1995; Smith, 1996; Smith & Desser, 1997; Barta *et al.,* 2012). These characteristics include the morphology of different life stages within the host, effect of the parasite on host cells, and host associations. Limitations with this phylogenetic method include the prerequisite for scientists to possess substantial knowledge of the biology and morphology of the organisms being studied as well as an objective determination of the phylogenetic weight assigned to specific morphological features. In combination with these limitations, additional

factors compound the problem. Morphology among haemogregarine taxa is filled with homoplasy and the weight of morphological features is often subjective rather than objective (Sloboda *et al.,* 2007; Barta *et al.,* 2012). While reflecting on the implications of these limitations and factors, it is not surprising that *Haemogregarina* and *Hepatozoon* have had such a disorderly history.

In the late 1990s, when the phylogenetic relationships based on the morphological characteristics of haemogregarines were used to classify species into the respective groups, DNA sequences were already being used in phylogenetic analyses (Brown, 2002). During this period phylogenetic studies on these parasites faced a delay, probably due to the lack of genetic data available for these organisms and due to limitations to isolate parasite DNA from the host. Around 1999, there was a notable increase in genetic data for Coccidia, initiating the first phylogenetic analyses of coccidian apicomplexans with 18S rDNA (also known as the SSU rRNA) in the years that followed (Carreno & Barta, 1999; Criado-Fornelio *et al.,* 2003a, b; Criado-Fornelio *et al.,* 2006; Kopecna *et al.,* 2006; Rubini *et al.,* 2006; Barta *et al.,* 2012). Barta *et al.,* (2012) was able to do an in-depth phylogenetic analysis of the adeleorinid in relation to other Apicomplexan parasites. Their study revealed that the molecular data aligned with the suggestions by Siddall (1995) and Smith *et al.* (2000) that the genus *Hepatozoon* is paraphyletic and should instead be divided into multiple genera. With these findings, DNA sequence data from conserved regions in genomic DNA proved to be useful in resolving phytogenic relationships at the species and genus level.

Molecular data indicates that the genus *Hepatozoon* is divided by members of Karyolysidae (*Hemolivia* and *Karyolysus*) into two clades (Barta *et al.,* 2012; Karadjian *et al.,* 2015; Maia *et al.,* 2016a). In their study, Karadjian *et al.* (2015) recognised clade A as the clade containing two subclades with *Hepatozoon* spp. parasitising carnivores, sister to the subclade containing *Hepatozoon* spp. and *Karyolysus* spp. parasitising Lacertids, Scincids, and Colubrids (lizards and snakes). Clade B also contained two subclades comprising *Hepatozoon* spp. parasitising anurans, birds, marsupials, rodents, snakes, turtles, and lizards, sister to the subclade of *Hemolivia* spp. Based on their findings, Karadjian *et al.* (2015) recommended the renaming of all *Hepatozoon* spp. within clade B as *Bartazoon.*, defined by their infection of various vertebrate hosts, using biting insects as vectors, and utilising syzygy as mode of fertilisation. In contrast, Maia *et al.* (2016a), argues that the renaming of this clade as *Bartazoon* is premature. Additionally, unidentified haemogregarines of geckos could potentially be placed within *Karyolysus*, further reducing the support for monophyly (Maia *et al.,* 2016a). Furthermore, the type species of *Hepatozoon* is *H. perniciosum,* which undergoes syngamy during fertilisation and infects rodents. This is problematic since the lifecycle of this haemogregarine would place it outside the carnivore *Hepatozoon* spp., but the molecular data places it within the *Bartazoon* (Karadjian *et al.,* 2015; Maia *et al.,* 2016a). For these reasons, the genus *Bartazoon* has not been widely accepted and the species that infect anurans, marsupials, rodents, various reptiles, and

birds remain as *Hepatozoon* and the paraphyly continues to be unresolved (Cook *et al.*, 2018; Netherlands *et al.,* 2018). Gutierrez-Liberato *et al.* (2021) later reinforced the rejection of this genus with robust molecular evidence from the rat *Hepatozoon*, confirming the issues raised by Maia *et al.* (2016a).

It is evident that more data surrounding these haemogregarines is needed to resolve these phylogenetic problems. While biology and morphology remain an important pillar for supporting the phylogeny and classification of organisms, it has limitations in distinguishing between species when used alone (Telford *et al.,* 2009). Additionally, while the 18S rDNA genetic marker has proven to be useful in resolving phylogenetic relatedness, it has shown to be unable to resolve the relationships of closely related species (Martynova-VanKlei *et al.,* 2008; Morrison *et al.,* 2004; Ogedengbe *et al.,* 2011). Alternative, faster-evolving genetic markers such as the gene encoding the mitochondrial cytochrome *c* oxidase subunit I (COI) was proposed to be more useful in resolving recent speciation events (Boore, 1999; Le *et al.,* 2000; Lane, 2009; Ogedengbe *et al.,* 2011; Ogedengbe *et al.,* 2018). In haemogregarines, three genomes can be found of which two are extrachromosomal genomes, one from the mitochondria and the other from the apicoplast (Feagin, 1994; Léveillé *et al.,* 2021). Apicomplexan mitochondrial genomes are small compared to other eukaryotic mitochondrial genomes (some as small as 6kbp in *Plasmodium*), and the structure of the mitochondrial genome is conserved, consisting of three protein-coding genes, the COI, cytochrome *c* oxidase subunit III (COIII) and cytochrome b (CytB), and various RNA fragments. The first mitochondrial genome of a *Hepatozoon* sp. to be sequenced was by Léveillé *et al.* (2015). This genome belongs to *Hepatozoon catesbianae* which infects the green frog *Lithobates clamitans*, has three protein-coding genes, and is approximately 6397 bps in length. Following the first sequenced mitochondrial genome of a *Hepatozoon* sp., subsequent efforts have led to the sequencing of mitochondrial genomes of *H. griseisciuri* (two genotypes) from the Ontario eastern grey squirrel *Sciurus carolinensis* (Léveillé *et al.,* 2020), *H. clamatae* and an unidentified *Hepatozoon* sp. from the green frog *Lithobates clamitans* (Léveillé *et al.,* 2021), two genomes from two unidentified *Hepatozoon* spp. parasitising rats *R. rattus* and *R. norvegicus* (Hrazdilová *et al.,* 2021), and some mitochondrial genome fragments have been sequenced for *H. canis* (Léveillé *et al.,* 2019b). In total, six complete mitochondrial genomes have been sequenced for *Hepatozoon* spp. with limited protein-coding gene sequences. *Haemogregarina* have not received as much attention with regard to the sequencing of the mitochondrial genomes or mitochondrial protein-coding genes as *Hepatozoon*. Currently, no complete mitochondrial genome sequences exist publicly for *Haemogregarina* on the NCBI database; however, the full mitochondrial genomes of four genotypes have been sequenced for *Haemogregarina balli* but are unpublished (Léveillé *et al.*, 2019a) and sequences of the COI, COIII and CytB genes are available from *Haemogregarina* spp. parasitising turtles in Columbia (Gutierrez-Liberato *et al.,* 2021).

The viability of the mitochondrial protein-coding genes as genetic markers in *Hepatozoon* has yet to be fully tested due to limited data available; however, these genes have been tested for *Haemogregarina*. The pairwise differences of the COI, COIII and CytB genes of anuran *Hepatozoon* (*H. catesbianae, H. clamatae,* and *Hepatozoon* sp. (MN245142)) COI, COIII and CytB are 0.4-0.70%, 0.80-1.80%, and 0.30-0.40% (Léveillé *et al.,* 2021). The pairwise differences of the rodent *Hepatozoon* (*Hepatozoon* spp. of rats and *H. griseisciuri* genotypes A and B) COI and CytB genes are 0.07-13.33%, and 0.09-11.21% (Gutierrez-Liberato *et al.,* 2021). Léveillé *et al.* (2021) suggested that the COIII gene may be a good type of target for barcoding among closely related *Hepatozoon* spp.They noted that the primers used in amplification of the COIII gene in ranid *Hepatozoon* may fail to bind in more distantly related *Hepatozoon* mitochondrial genomes due to the high variability of the gene and recommend the use of conserved regions upstream and downstream of the target region. While high variability has been detected in one of the mitochondrial protein-coding genes of *Hepatozoon*, the same (if not more) variability has been detected in all the mitochondrial protein-coding genes of *Haemogregarina* spp. Léveillé *et al.* (2019a) discovered that the mitochondrial genomes of four *Haemogregarina balli* genotypes had a pairwise difference of 24-35%. The protein-coding genes of these *Haemogregarina balli* genotypes had a pairwise difference of 23-33% for COI, 31-43% for COIII and 25-38% for CytB. Furthermore, Gutierrez-Liberato *et al.* (2021) have performed phylogenetic analyses on *Haemogregarina* spp. parasitising turtles in Columbia using the three mitochondrial protein-coding genes. In their analyses, sixteen turtle species (458 individuals) were screened and 119 nucleotide sequences of the COI, COIII, and CytB genes from *Haemogregarina* spp. were obtained. Their results indicate 16-25% pairwise differences among the genes that encode COI and COIII proteins. As a result, Gutierrez-Liberato *et al.* (2021) speculates that pairwise differences above 30% in mitochondrial markers could be associated with different *Haemogregarina* species. Although mitochondrial genomes and its protein-coding genes of *Haemogregarina* are highly variable, they still retain phylogenetic signal (Gutierrez-Liberato *et al.,* 2021).

Problem statement

The genus *Hepatozoon* replaces *Haemogregarina* as a reservoir genus for poorly described haemogregarines, containing more than 300 species. The taxonomy of the genus is somewhat obscure due to a lack of sufficient genetic markers capable of resolving recent speciation events. The mitochondrial protein-coding genes have been proposed and have proven in other organisms to be useful in delimitating nodes deep within a phylogenetic analysis. Mitochondrial protein-coding gene data is lacking for *Hepatozoon* and consequently no phylogenetic studies have been conducted with these genetic markers on this genus. The mitochondrial protein-coding genes of

the genus *Haemogregarina* shows high sequence variability among close species. It is currently unclear whether this variability is consistent throughout *Haemogregarina* and whether it has phylogenetic implications due to a lack of mitochondrial DNA data.

Aims of this study

In this study the aim was to broaden our understanding and provide more data on the mitochondrial DNA of Adeleorinid coccidia from the genera *Hepatozoon* and *Haemogregarina*.

CHAPTER 1

Objectives of this study

The objectives of this study were to:

1. Sequence the full mitochondrial genomes of *Hepatozoon* spp. parasitising anurans and *Haemogregarina* spp. parasitising terrapins in South Africa.

2. Design and test primers that would target the mitochondrial protein-coding genes of members of *Hepatozoon*.

3. Sequence the mitochondrial protein-coding genes of various *Hepatozoon* spp. and analyse their usefulness for phylogenetic studies:

    a. Analyse phylogenetic characteristics of each protein-coding gene compared to the traditional 18S rRNA gene, including informative sites, variable sites and conserved sites.

    b. Perform congruency tests on all genetic markers with different configurations.

    c. Construct and analyse phylogenetic trees with different genetic marker combinations.

CHAPTER 2

ARTICLES

CHAPTER 2.1

Generating the full mitochondrial genomes through primer walking of *Hepatozoon ixoxo, Hepatozoon tenuis,* and *Hepatozoon theileri* (Adeleorina: Hepatozoidae), parasitising South African anurans

Graphical abstract for Chapter 2.1

CHAPTER 2.1

Generating the full mitochondrial genomes through primer walking of *Hepatozoon ixoxo, Hepatozoon tenuis,* and *Hepatozoon theileri* (Adeleorina: Hepatozoidae), parasitising South African anurans.

Tiaan Haarhoff, Marianne Venter[2], Louis H. du Preez[1,3], and Edward C. Netherlands[1,4]*

[1] African Amphibian Conservation Research Group, Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

[2] Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom, South Africa.

[3] South African Institute for Aquatic Biodiversity, Somerset Street, Makhanda 6140, South Africa

[4] Department of Zoology and Entomology, University of the Free State, PO Box 339, Bloemfontein, 9300, South Africa.

***Author for correspondence:** EC Netherlands, E-mail: NetherlandsEC@ufs.ac.za

Word count: 4784

2.1.1. Abstract

Southern Africa is recognised as one of the most biodiverse regions globally, with a continual expansion of biodiversity marked by the continuous discovery of new parasites parasitising the country's rich fauna. Species of *Hepatozoon* (Adeleorina: Hepatozoidae) comprises a substantial group of parasites, with well over 300 species. Among Southern African frogs, these haemogregarines stand as the most common blood parasites. The considerable diversity within the genus presents a notable phylogenetic challenge due to the limited resolution of the universal 18S rDNA marker in classifying such large groups. The protein-coding genes of the mitochondrial genome have been proposed as markers with higher phylogenetic resolution. In Chapter 2.1 the mitochondrial genomes of three *Hepatozoon* spp. parasitising South African anurans was sequenced through Sanger-sequencing by means of primer walking. The present study yielded three complete mitochondrial genomes, each approximately 6500 base pairs in length, encoding the three protein-coding genes commonly observed in apicomplexan mitochondrial genomes. The ITS-1 region and the 23S pl-rDNA were tested as supplementary genetic markers however, the primers used for these markers seem ineffective for the haemogregarines of this study, underscoring the need for improved primer design. In conclusion, this study confirms the effectiveness of Sanger sequencing through primer walking for mitochondrial genome sequencing.

Key words: *Hepatozoon*, haemogregarines, Adeleorina, mitochondrial genome DNA, protein-coding genes, primer walking, Sanger-sequencing

Key findings:

- Full mitochondrial genome sequences of ~6400bp each for three *Hepatozoon* spp. parasitising South African anurans.
- The mitochondrial genome organisation is conserved among closely related species.
- Sanger-sequencing through primer walking is sufficient for generating high quality mitochondrial DNA sequences for adeleorinids.
- Primers targeting the ITS-1 region and 23S pl-rDNA too specific for haemogregarines.

2.1.2. Introduction

The genus *Hepatozoon* is the largest of Apicomplexan haemogregarine (Coccidiomorphea, Coccidia, Adeleornia) which follow a heteroxenous life cycle (Smith, 1996). Species of *Hepatozoon* have been reported to form intermediate host associations with all terrestrial vertebrate groups and are transmitted by a wide range of invertebrate definite hosts such as ticks, mites, mosquitoes, and leeches (Smith, 1996; Davies & Johnston, 2000; Netherlands *et al.,* 2020). This diverse host association has led to the emergence of more than 300 species. Due to this large diversity, it has become increasingly more difficult to differentiate between closely related species based off morphology and biology alone (Barta *et al.,* 2012; Léveillé *et al.,* 2015; Netherlands *et al.,* 2020).

The introduction of molecular markers in phylogenetic analyses, has led to re-descriptions, revisions, and attempted reclassification of some *Hepatozoon* spp. into new genera (Siddall, 1995; Maia *et al.,* 2016a; Barta *et al.,* 2012). However, this addition to the taxonomy toolset does not come without its own challenges. The most common used genetic marker, 18S rDNA is conserved among eukaryotes and has been useful in species identification (Ogedengbe *et al.,* 2011; Léveillé *et al.,* 2019b). The level of conservation of the 18S rRNA gene makes it useful in genus and species identification without prior knowledge of parasite morphology or sporogony (Barta *et al.,* 2012; Léveillé *et al.,* 2019b). On the other hand, the 18S rRNA gene is too conserved among closely related species, reducing its resolution in molecular phylogenetic analyses (Zhao *et al.,* 2001; Morrison *et al.,* 2004; Léveillé *et al.,* 2019b). This limitation has been demonstrated in previous studies. Morrison *et al.* (2004) found that the 18S rDNA sequence could not elucidate the monophyly of *Eimeria* spp., a group of apicomplexan parasites that has veterinary importance in the poultry industry. The genus *Eimeria* is speciose, comprising over 1000 recognised species. It is uncertain whether the genus *Hepatozoon* is as specious, but concerns have been raised (Barta *et al.,* 2012; Maia *et al.,* 2012; Haklová-Kočíková *et al.,* 2014; Cook *et al.,* 2016). The need for better genetic markers is clear.

To address this issue, additional genes have been suggested as better molecular markers for use in phylogenetic analysis of closely related species, such as the cytochrome *c* oxidase subunit I (COI) mitochondrial protein-coding gene (Lane, 2009; Ogedengbe *et al.,* 2011). The mitochondrial genome size of Apicomplexans is reduced in most cases, ranging between 6-8kbp (Léveillé, 2015). Along with the COI, two other protein-coding genes of the mitochondrial genome are also conserved in apicomplexans, which are the cytochrome *c* oxidase subunit III (COIII) and the cytochrome b (CytB). These CoDing Sequence (CDS) genes are fast-evolving and are useful in resolving taxonomic and phylogenetic classification challenges faced by large groups such as *Hepatozoon* (Léveillé *et al.,* 2015; Ogedengbe *et al.,* 2011; Pawlowski *et al.,* 2012). The ability of COI to delineate monophyletic groups has shown promising results in previous studies (Ogedengbe *et al.,* 2011; Teletchea, 2010). Due to its small size, it is appealing to sequence the whole mitochondrial genome instead of only the CDS genes. Sequencing the whole genome of haemogregarines may have some benefits, such as knowing how mitochondrial genome organisation and gene orientation differs between genera or species, and having full CDS gene sequences is better than partial sequences.

Along with the COI protein-coding gene, the nuclear ribosomal internal transcribed spacer 1 (ITS-1), and the large subunit (23S) ribosomal RNA gene sequence from the apicoplast are proposed as faster evolving regions that could potentially elucidate recent evolutionary events among closely related species (Léveillé *et al.,* 2019b; 2019c; 2021). The ITS-1 region is a variable non-coding region in the three-part ITS region of the nuclear genome in eukaryotes (Wang *et al.,* 2014). The apicoplast genome and its associated genes are found to be conserved among apicomplexans (Mathur *et al.,* 2020). The use of apicoplast ribosomal subunits as genetic barcodes to supplement molecular phylogeny in Apicomplexa could be of benefit. Unfortunately, the database required to use these markers for phylogenetic analyses is limited (Wang *et al.,* 2014). Due to this the effectiveness of these markers is ambiguous, requiring more sequences to determine their reliability.

The haemogregarines from the present study are *Hepatozoon theileri, Hepatozoon tenuis, and Hepatozoon ixoxo*, all of which infect South-African anurans. *Hepatozoon theileri* infect the common river frog *Amietia delalandii* (Netherlands *et al.,* 2014b), *H. tenuis* infects the spiny reed frog *Afrixalus fornasinii* (Netherlands *et al.,* 2017), and *H. ixoxo* infects three species of toad in the genus *Sclerophrys,* namely *Sclerophrys pusilla, Sclerophrys garmani,* and *Sclerophrys gutturalis* (Netherlands *et al.,* 2014a). All three species have been described morphologically and 18S rDNA sequences are available (Netherlands *et al.,* 2014a; 2014b; 2017).

The aim of this study was amplify and sequence mitochondrial DNA of species of *Hepatozoon* parasitising South African anurans. The objectives of this study were (1) to sequence the full mitogenomes (mitochondrial genomes) of three *Hepatozoon* spp. through Sanger-sequencing by means of primer design and primer walking; (2) explain the methodology of

sequencing the full mitogenomes; and (3) to generate additional genetic markers that could be used for phylogenetic analyses in future related studies.

2.1.3. Materials and methods

2.1.3.1. Specimen collection and blood smear preparation

Amietia delalandii (n = 15) was caught by hand from the North-West University botanical garden ponds in Potchefstroom (26°40'56'' S; 27°05'43'' E). *Sclerophrys gutturalis* toads were caught during a separate study in the Vhembe biosphere (23°05'29.9" S; 30°05'17.1" E) where blood was extracted and stored in absolute alcohol. Whole blood samples were also available for *Afrixalus fornasinii* from a separate study, caught at St. Lucia on Monzi Farm, KwaZulu-Natal, South Africa (28°26′56″S 32°17′18″E). This study received the relevant ethical approval: (North-West University ethics approval no NWU-00380-16-A5).

Some blood was prepared as a blood smear on a microscope slide, air-dried, fixed, and stained with Giemsa-stain (FLUKA, Sigma-Aldrich, Steinheim, Germany). The blood smears were then screened with a Nikon Eclipse E800 compound microscope (Nikon, Amsterdam, Netherlands) using the 100x immersion oil objective for *Hepatozoon* sp. infection. Samples with the highest observed infection were chosen for molecular analysis. All *Amietia* specimens caught in this study were released at the site of capture after blood samples were taken.

2.1.3.2. DNA extraction and PCR

DNA was extracted from whole blood samples using the NucleoSpin®Tissue Genomic DNA Tissue Kit (Macherey-Nagel, Düren, Gemrany) following the kit instructions.

The 18S nu-rDNA gene, ITS-1 region, and 23S pl-rDNA were amplified using the DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific, Johannesburg, SA) at a reaction volume of 25µl, with 1µl (100-500ng/µl) DNA template and 10µM of each primer (Table 1). The three samples were labelled Sample A (from *Schlerophrys gutturalis*), B (from *Amietia delalandii*), and C (from *Afrixalus* fornasinii) respectively.

Table 1: Primer information and conditions of PCR reactions performed in this study.

Barcode	Primer name	Primer sequence	Annealing temp	Extension time	Fragment size	Reference
18S	Adel_18S_13F	5′-CCTGCCAGTAGTCATATGCT-3′	56	2:00	~1300bp	Léveillé *et al*., 2021
	Adel_18S_1522R	5'-AYCCTATTTAGCAGGTTAAGGT-3'				Léveillé *et al*., 2020
ITS-1	Api_18S_1000F	5'-AGATACCGTCGTAGTCTTAACT-3'	62	2:00	~1000	Léveillé *et al*., 2019c
	Api_5.8S_R	5′-GCTRCGKYCTTCATCG-3′				Léveillé *et al*., 2021
23S	Api_PL_23S_2414F	5′-TAACGGTCCDAAGGTAGCG-3′	61	1:00	~700	Léveillé *et al*., 2021
	Api_PL_23S_3135R	5′-TTYTGAACCCAGCTCACGT-3′				Léveillé *et al*., 2021
Mitochondrial genome	Api_LSUG_UNI_F	5′-TTGAGRCAGTTTGTTCCCTATCT-3′	62	2:30	~1000-3000	Léveillé *et al*., 2019c
	Api_LSUF_UNI_R	5′-ACCTGTTATCCCCGGCGWA-3′				Léveillé *et al*., 2019c
	Api_LSUG_R	5′-AGATAGGGAACAAACTGYCTCAA-3′				Léveillé *et al*., 2019c
	Api_LSUF_F	5′-GTWCGCCGGGGATAACAGGT-3′				Léveillé *et al*., 2019c
	Api_LSUE_UNI_F	5'-GGTAAGACCCTGAGCACCT-3'				Léveillé *et al*., 2019c
	BarHep_COI_274R	5′-CCACTTCRGACGAGCCACA-3′				Léveillé *et al*., 2020
	Adel_RNA13_R	5′-GCTWACTTCCCGGTCAAAC-3′				Léveillé *et al*., 2020
	Api_SSUE_Uni_R	5'-CGTGACGAGCGGTGTGT-3'				Léveillé *et al*., 2019c
	Api_RNA18_Uni_F	5'-TYGGTATTGCATGCCTGGT-3'				
	Api_RNA10_Uni_F	5'-GTAAGGAAWAGGWAAGRTTAACCG-3'				Léveillé *et al*., 2019c

Description: The table lists the primers used for the different genetic markers tested, as well as the primers used to generate the full mitochondrial genomes of this study. The primer name, sequence, conditions, and references are named in the columns from left to right.

2.1.3.3. *Primer-walking*

The sequencing technique to generate the full mitogenomes in this study is primer-walking. The following section explains the steps used to perform this technique. Initially the mitogenome is amplified as two large fragments (~3000bp). According to the mitogenome of *Hepatozoon catesbianae* generated by Léivellé *et al.* (2015) the rRNA fragments LSUG and LSUF divides the mitogenome of this closely related species into two equal parts. A PCR reaction was performed with the primers LSUG_R and LSUF_F to generate the first ~3000bp fragment using Q5® High-Fidelity 2X Master Mix (New England Biolabs) in a final volume of 25µl, with 1µl (100-500ng/µl) DNA template and 10µM of each primer. The same PCR reaction was performed for the second ~3000bp fragment but with the primers LSUG_F and LSUF_R. The first 1000bp of each fragment was sequenced separately. Near the middle of these sequenced fragments a primer was designed to sequence the next 1000bp fragments. This process was repeated until both 3000bp fragments were sequenced. A small region of un-sequenced DNA would remain where the primers LSUG and LSUF bound to the template DNA, which was covered by additional PCR reactions. These additional PCR reactions contained DreamTaq Green PCR Master Mix (2X) in a final volume of 25µl, with 1µl (100-500ng/µl) DNA template and two 10µM primers located at opposite ends of the gap. PCR conditions and primer sequences are listed in Table 1. The PCR products were visualized through gel electrophoresis on 1% agarose gel.

2.1.3.4. *Mitochondrial genome construction and annotation*

All the sequences obtained were edited, aligned, and annotated using Geneious Prime 2021.2.2 (https://www.geneious.com). The sequenced mtDNA (mitochondrial DNA) fragments obtained from primer walking and additional PCR reactions for each mitogenome was imported to Geneious Prime. The sequences of each mitogenome were mapped to a reference mitogenome (*Hepatozoon catesbianae* in this study) separately using the 'Map to reference' tool at its highest sensitivity. The consensus sequence generated from the mapping was extracted and annotated with the 'Annotate and predict' tool using the mitogenome of *Hepatozoon catesbianae* as the reference. CDS genes and rRNA fragments from *H. catesbianae* between 75-100% sequence similarity were used to annotate the mitogenomes of this study. The 'Find repeats' tool was used to locate repeat regions on the mitogenomes larger than 50bp. The purpose of this tool was to locate the overlap of the two ends of the linear consensus sequences to complete the circular mapping of the mitogenomes. The 'Find ORF' tool was used, set to locate ORFs (Open Reading Frames, also known as the CDS genes) on the mitogenomes larger than 500bp and similar to the genetic code of mold protozoa mitochondrial genome. The ORFs were then used to annotate the predicted start and end of the three CDS genes of each mitogenome in this study.

2.1.4. Results

2.1.4.1. Sampling

Common river frogs (*Amietia delalandii*) (n=15) were caught in the NWU Botanical Garden, of which four were infected with a *Hepatozoon* sp. Archived blood samples infected with *Hepatozoon* spp. were available in the lab for *Afrixalus fornasinii* and *Schlerophrys gutturalis*.

2.1.4.2. Mitochondrial genomes

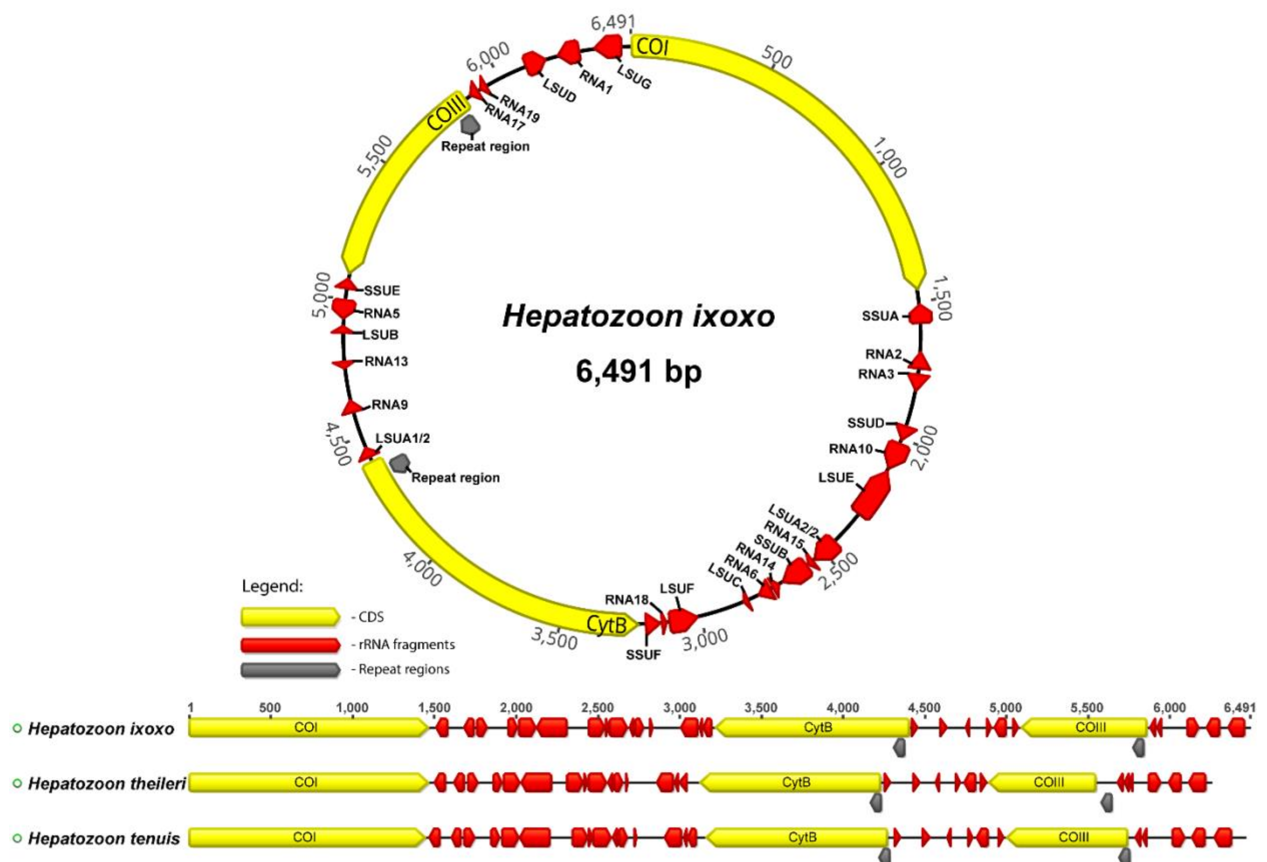


Figure 2.1.1: Complete circular mitochondrial genomes of *Hepatozoon theileri* ex *Amietia delalandii*, *Hepatozoon ixoxo* ex *Schlerophrys gutturalis* and *Hepatozoon tenuis* ex *Afrixalus fornasini*. Top: The mt-genome of H. ixoxo is illustrated in circular form. Bottom: The mt-genome of all three *Hepatozoon* spp. is illustrated in linear form. Three CDS genes (cytochrome c oxidase subunit I [COI], cytochrome c oxidase subunit III [COIII], and cytochrome B [CytB]) are encoded by the genomes separated by rRNA fragments.

The full circular mitogenomes for *H. ixoxo, H. theileri,* and *H. tenuis* were successfully amplified and sequenced (See Figure 2.1.1). The mitogenomes of *H. ixoxo* has a unit length of 6,491bp with 38.1% GC content, *H. theileri* has a unit length of 6,252bp with 37.5% GC content, and *H. tenuis* has a unit length of 6,463bp with 37.8% GC content. All three mitogenomes were aligned with the Multiple alignment tool in Geneious with a 93% cost matrix, a gap open penalty of 12,

and a gap extension penalty of 3 (Kearse *et al*., 2012). The organisation and orientation of the CDS genes and rRNA fragments relative to the COI in each mitogenome is summarized in Supp. Table 1. Between the mitogenomes of *H. ixoxo* and *H. theileri* there are 672 single nucleotide differences (SNDs), between *H. ixoxo* and *H. tenuis* there are 551 SNDs, and between *H. theileri* and *H. tenuis* there are 602 SNDs. The mitogenome of *H. theileri* is more than 200bp shorter than that of *H. ixoxo* and *H. tenuis*. The mitogenome of *H. ixoxo* has 28bp more than *H. tenuis.* Deletions from the mitogenome of *H. theileri* ranged from 1~50bp deletions dispersed throughout the genome sequence between coding genes and rRNA fragments. No clear evidence of deletions was found within protein-coding genes, however there is a 2-3bp deletion in the LSUE rRNA fragment of *H. theileri* and *H. ixoxo* compared to *H.* tenuis. No other deletions were observed in the rRNA fragments of all the mitogenomes. Fewer deletions were found throughout the mitogenome of *H. ixoxo* and *H. tenuis*, with these deletions sometimes occurring at the same sites as those deletions of *H. theileri*, and often only one of either *H. ixoxo* or *H. tenuis* having the deletion.

When compared to five other *Hepatozoon* spp. (see Table 2), the COI gene sequences (~1767bp) of all three *Hepatozoon* spp. investigated in this study shared the highest identity with *H. clamatae* (MN245142), *H. catesbianae* (NC_044466), and an undescribed *Hepatozoon* sp. (MN245143). COI gene sequences of *H. ixoxo* had a ~94% identity with these three species, while *H. theileri* had 93% and *H. tenuis* had ~94% identity. Likewise, the CytB gene sequence (~1204bp) of all three *Hepatozoon* spp. shared the highest identity with *H. clamatae* (MN245142), *H. catesbianae* (NC_044466) and the undescribed *Hepatozoon* sp. (MN245143). CytB gene sequenced of *H. ixoxo* had a ~92% identity with these three species, while *H. theileri* had ~93% and *H. tenuis* had ~93.9% identity. The identity of the COI and CytB gene sequence of *H. ixoxo, H. theileri* and *H. tenuis* fell below 70% when aligned with that of *Hepatozoon canis*. Interestingly, the COI and CytB from species of *Klossia* and *Klossiella* has a higher % identity than *Hepatozoon canis* with the investigated *Hepatozoon* spp.

Table 2: Percentage nucleotide identity between some adeleorinid COI genes (displayed in bold) and CytB gene (displayed as normal text) sequences.

	Hepatozoon ixoxo	*Hepatozoon theileri*	*Hepatozoon tenuis*	KF894962 - *Hepatozoon catesbianae*	MN245142 - *Hepatozoon clamatae*	MN245143 - *Hepatozoon* sp.	MK452388 - *Hepatozoon griseisciuri* A	MK452389 - *Hepatozoon griseisciuri* B	MT936933 - *Hepatozoon* sp.	MH615002 - *Hepatozoon canis**	MH203050 - *Klossiella equi*	NC_058857 - *Klossia helicina*	NC_058856 - *Klossia razorbacki*
Hepatzoon ixoxo		**93.775**	**92.453**	**90.890**	**91.304**	**93.404**	**86.014**	**86.685**	**84.446**	**60.635**	**69.346**	**67.456**	**69.113**
Hepatozoon theileri	93.586		**91.644**	**90.062**	**90.131**	**92.270**	**85.087**	**86.498**	**84.165**	**60.700**	**69.755**	**67.884**	**70.021**
Hepatozoon tenuis	94.250	93.801		**92.168**	**92.028**	**92.092**	**84.476**	**85.105**	**83.625**	**61.334**	**69.850**	**69.890**	**70.420**
KF894962 - *Hepatozoon catesbianae*	91.951	93.151	93.890		**99.517**	**99.645**	**85.300**	**86.611**	**86.404**	**60.360**	**70.625**	**69.386**	**70.231**
MN245142 - *Hepatozoon clamatae*	91.689	93.242	93.980	99.730		**99.291**	**85.093**	**86.680**	**86.473**	**60.560**	**70.350**	**69.117**	**69.951**
MN245143 - *Hepatozoon* sp.	91.514	93.059	93.980	99.551	99.641		**87.092**	**88.511**	**87.518**	**61.354**	**70.625**	**69.875**	**70.161**
MK452388 - *Hepatozoon griseisciuri* A	83.851	86.484	85.355	82.902	83.074	82.902		**91.958**	**89.648**	**60.893**	**70.281**	**66.601**	**69.532**
MK452389 - *Hepatozoon griseisciuri* B	84.111	87.397	86.074	86.092	86.004	86.004	88.774		**91.205**	**61.445**	**70.144**	**66.893**	**69.043**
MT936933 - *Hepatozoon* sp.	83.081	87.624	86.253	86.303	86.482	86.482	86.788	90.476		**60.626**	**69.823**	**68.598**	**70.091**
MK214283 - *Hepatozoon canis**	59.500	59.766	59.087	59.462	59.462	59.372	58.924	59.283	59.054		**60.203**	**57.839**	**60.083**
MH203050 - *Klossiella equi*	62.510	64.550	63.121	61.941	61.941	61.857	62.144	61.345	61.785	57.437		**80.829**	**81.586**
NC_058857 - *Klossia helicina*	64.212	65.437	64.539	62.737	62.651	62.995	62.947	62.137	63.008	55.402	79.839		**93.390**
NC_058856 - *Klossia razorbacki*	64.469	65.887	64.716	62.651	62.565	62.737	62.521	61.966	62.835	56.830	80.080	94.390	

Description: This table shows the %identity matrix between the COI gene sequences (presented in bold text) and CytB gene sequences (presented as normal text) currently available for *Hepatozoon* spp. separated by a diagonal line through the matrix. The table also compares the COI and CytB gene sequences of two other genera, *Klossia* and *Klossiella* against the matrix. The different host groups infected by the *Hepatozoon* spp. used in this matrix is colour coded: green represents anuran parasites, blue indicates rodent parasites, and red indicates carnivore parasites. Notice *: The two sequences used for *H. canis* has different accession numbers, due to these two sequences being separate, whereas the other sequences were obtained from the same mt-genome. Therefore, all values below the diagonal line in the MH615002 column are values for MK214283, and all values past the diagonal line in the MK214283 row is for MH615002.

The structure of the three mitogenomes is consistent with that of other *Hepatozoon* spp. mitochondrial genome sequences. These consistencies are the unit length of each mitogenome being between 6,000 – 7,000 bps, having a circular mapping, and containing 3 CDS genes (COI, COIII, CytB) separated by rRNA fragments (Hikosaka *et al.,* 2013). The organisation and orientation of the protein-coding genes and rRNA fragments more closely resemble that of other anuran *Hepatozoon* spp. (e.g. *Hepatozoon clamatae* & *Hepatozoon catesbianae*) than rodent *Hepatozoon* spp. (e.g. *Hepatozoon greisisciuri*). A comparison between the mitogenomes sequenced in this study and other Adeleorinid haemoparasites is illustrated Figure 2.1.2.

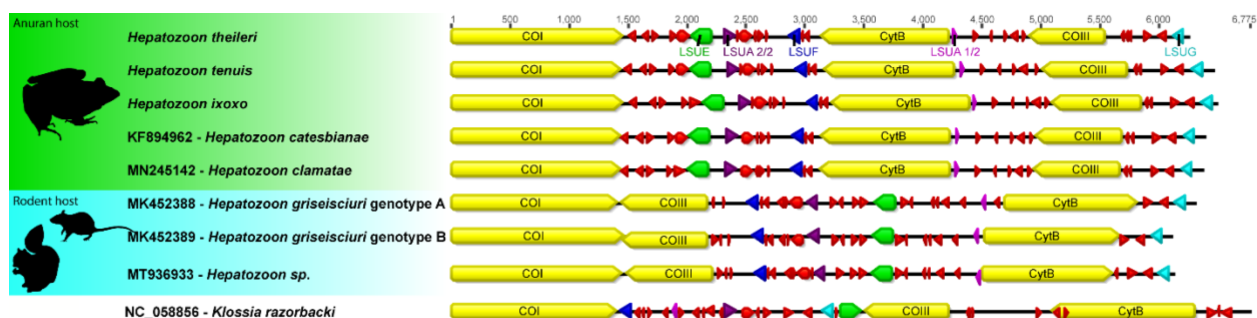


Figure 2.1.2: The mitochondrial genome structure of *H. ixoxo*, *H. theileri*, and *H. tenuis* compared against those of other *Hepatozoon* spp. and *Klossia razorbacki*. On the left, *Hepatozoon* spp. that infect anuran hosts are highlighted in green, and *Hepatozoon* spp. that infect rodent hosts are highlighted in blue. A few large subunit (LSU) rRNA fragments conserved throughout the mitochondrial genomes were chosen and colour coded to indicate, along with the protein-coding genes, the general structural changes between the mitochondrial genomes of adeleorinid coccidia.

2.1.4.3. 18S, 23S & ITS-1

The nuclear 18S rDNA, the complete ITS-1, and the 5' end of the 5.8S rDNA sequences were sequenced for all *Hepatozoon* spp. from the present study, and the pl-rDNA gene sequence was sequenced for *H. ixoxo* only. The BLAST results of the 18S rDNA sequence for Sample A (1,272bp after trimming) matched with *H. ixoxo* (MG041604) (99.76% identity and 98% query cover), Sample B (1,283bp after trimming) match with *H. theileri* (MG041605) (99.92% identity and 100% query cover), and Sample C (1,265bp after trimming) match with *H. tenuis* (MG041597) (99.92% identity and 100% query cover)

Out of the two fragments sequenced for the ITS-1 region of Sample A only one fragment (1000F primer fragment), had a positive identification as a *Hepatozoon* sp. The fragment generated from the 5.8S_R primer for Sample A (988bp untrimmed) was of very low quality (4.1% of the sequence is high quality untrimmed) and had no match to any organism on the NCBI database. The chromatogram of this fragment had mixed signalling, possibly meaning multiple untargeted organisms was amplified in the sample. It may be that the 5.8S_R primer used in this study is not specific enough to accurately amplify the ITS-1 region of apicomplexans in the

presence of other microbes. Similar observations were seen previously by our group during attempts to amplify the ITS-1 region of other haemogregarines. The fragment generated from the 1000F primer for Sample A (1,007bp) was of very high quality (66.3% of the sequence is of high quality untrimmed) and matched with a *Hepatozoon* sp. (95.64% identity and 99% query cover with *Hepatozoon clamatae*: MN244529).

Both ITS-1 fragments for Sample B identified as *Hepatozoon* sp. The fragment generated from 5.8S_R primer (995bp untrimmed) was of low quality as well (4.1% of the sequence is of high quality untrimmed) and had mixed signalling, however, the sequence was intact enough to construct contigs and align it with the 1000F primer fragment. The 1000F primer fragment for Sample B (981bp untrimmed) was of good quality (68.4% of the sequence is of high quality untrimmed) like that of Sample A and match with a *Hepatozoon* sp. The two fragments were de novo assembled, and the consensus sequence (946bp) matched with a *Hepatozoon* sp. (98.56% identity and 80% query cover: MN244530). The ITS-1 region sequence for Sample C was amplified and matched with a *Hepatozoon* sp. (95.06% identity and 92% query cover with *Hepatozoon clamantae*: MN244529).

The 23S pl-rDNA gene sequence was obtained for Sample A and -C. The 23S pl-rDNA gene sequence for Sample B was not obtained, instead, the chromosomal DNA of a bacterium was amplified and sequenced by accident. Both fragment sequences obtained for Sample A had an average length of 595bp, GC content of 34.1%, and were of high quality (77% of the sequences were of high quality). The sequences match with *Hepatozoon* catesbianae (94.88% identity and 99% query cover: MN245144) indicating that the 23S for Sample A is from a *Hepatozoon* sp., which in this case is most likely *H. ixoxo.*

2.1.5. Discussion

2.1.5.1. Mitochondrial genomes

This study presents the full mitogenomes of *Hepatozoon ixoxo, Hepatozoon theileri,* and *Hepatozoon tenuis*. The genome organisation and orientation of the protein-coding genes as well as rRNA fragments are near identical to that of other anuran *Hepatozoon* spp. (Léveillé *et al.,* 2021). However, there is a noticeable difference between the mitogenome organisation and orientation of *Hepatozoon* spp. parasitising anuran hosts when compared to those parasitising other vertebrates or those parasites from other genera. Figure 2.1.2 illustrates these similarities and dissimilarities between the mitogenomes of *Hepatozoon* spp. parasitising anurans and rodents along with that of *Klossia razorbacki* (NC_058856). It seems, with the mitogenomes of all *Hepatozoon* spp. at least, there is a conserved association between the COI CDS and a region before the start codon of the gene containing three rRNA fragments: LSUD, RNA1 and LSUG. The orientation and relative location of the RNA fragments to the COI CDS is conserved in all *Hepatozoon* mtDNA genomes available on Genbank. One other such conserved association throughout currently available *Hepatozoon* mitogenomes is that of the CytB CDS and the LSUA1/2 rRNA fragment where the rRNA fragment is in an inverse orientation relative to the CytB CDS. The significance of these conserved associations is that mitochondrial DNA of *Hepatozoon* species can be amplified from samples containing more than one genus of adeleorinid parasite. For example, if a PCR sample containing both *Hepatozoon* and *Haemogregarina* mtDNA is mixed with primers designed to aid the amplification of the protein-coding genes of adeleorinids, both parasite's mtDNA will be amplified and mixed signalling will occur. A carefully designed PCR reaction targeting these conserved associations will avoid mixed signalling of DNA from any organism that happened to be in the sample. In a sense, this information eliminates the need to design more specific primers, but rather formulate more specific PCR reactions. These types of associations can only be realized with the sequencing of the full mitogenomes of these parasites. That said, sequencing the full mitogenomes of adeleorinid haemoparasites may lead to the discovery of more of these conserved associations among groups of parasites or among different taxa. Amplifying and sequencing regions from these conserved associations may also help visualize the general organisation of the mitogenome of the *Hepatozoon* sp. in study. The ability to predict the organisation of the mitogenome of these parasites was vital when attempting to sequence the whole mitogenome instead of protein-coding genes only. Prior knowledge of the organisation of Hepatozoidae and Adeleidae mitogenomes helped to predict what PCR reactions needed to be performed to achieve the desired outcomes.

2.1.5.2. Primer walking

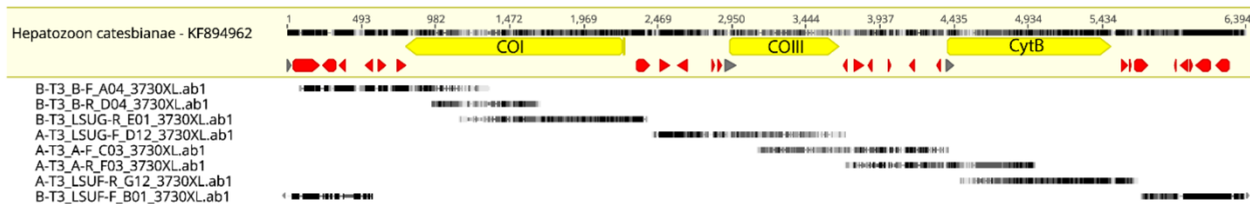


Figure 2.1.3: The primer-walking technique used to sequence the whole mt-genomes of *Hepatozoon* spp. of this study is displayed in the above diagram.

The primer walking technique is laid out in the materials and methods of this paper. With this sequencing technique the full mitogenomes of *Hepatozoon ixoxo, Hepatozoon theileri* and *Hepatozoon tenuis* was done by generating as little as 9 fragments averaging 1000bp. In some cases, gaps were formed between fragments, however these gaps were easily covered by a quick PCR covering the area. Figure 2.1.3 displays the DNA scaffolding produced by this sequencing technique. The quality of DNA sequences generated were mostly of high quality (<80%).

2.1.6. Conclusion

. In this study, it was discovered that there are conserved CDS-RNA associations within the mitochondrial genomes of *Hepatozoon*. The significance of these conserved associations is that the need for specific primers to help amplify mtDNA in a sample of mixed haemogregarine infection is not needed. Instead, a specific PCR reaction is required that target these conserved associations. The relatively small size of apicomplexan mitogenomes should serve as encouragement to follow this path. Additionally, the increased amount of data generated may aid the development of new and better primers for future studies on undescribed haemogregarines. Therefore, sequencing the full mitochondrial genomes of adeleorinid haemogregarines may be more useful than targeting CDS genes alone. In this study the primer-walking sequencing technique was used to sequence the full mitogenomes of three *Hepatozoon* spp. parasitising anurans from Southern Africa. The technique delivered good quality results in reasonable time. The fragment DNA generated by the sequencing technique constructed DNA scaffolding with good coverage across the whole mitogenomes. Each fragment generated had minimal degenerate sites within the DNA sequence, with most uncertainties located at the 5' and 3' ends of the fragments. These low-quality regions were covered by the preceding fragments that were generated, thanks to the nature of the technique used. The primers used to amplify and sequence the ITS-1 and 23S regions of the *Hepatozoon* spp. in this study produced varying results. The primers used were not able to produce fragments with reasonable quality for phylogenetic analysis. It may be that the primers are not specific enough to target the nuclear DNA of the *Hepatozoon* spp. in this study. The poor results for these fragments could also be due to inadequate sequencing conditions. Further tests should be done on these primers to conclude their usefulness.

Chapter 2.2

The complete mitochondrial genomes of two *Haemogregarina* spp. (Adeleorina: Haemogregarinidae), parasitising the serrated hinged terrapins *Pelusios sinuatis* of Southern Africa

CHAPTER 2.2

The complete mitochondrial genomes of two *Haemogregarina* spp. (Adeleorina: Haemogregarinidae), parasitising the Serrated Hinged Terrapin *Pelusios sinuatis* of Southern Africa.

Tiaan Haarhoff, Marianne Venter[2], Louis H. du Preez[1,3], and Edward C. Netherlands[1,4]*

[1] African Amphibian Conservation Research Group, Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

[2] Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom, South Africa.

[3] South African Institute for Aquatic Biodiversity, Somerset Street, Makhanda 6140, South Africa

[4] Department of Zoology and Entomology, University of the Free State, PO Box 339, Bloemfontein, 9300, South Africa.

***Author for correspondence:** EC Netherlands, E-mail: NetherlandsEC@ufs.ac.za

Word count: 3937

2.2.1. Abstract

The mitochondrial protein-coding genes in species of *Haemogregarina* (Adeleorina: Haemogregarinidae) have been examined, revealing remarkable variation even among closely related members. Recent research suggests that these protein-coding genes could be valuable for understanding the evolutionary relationships of haemogregarines. However, the unexpectedly high diversity observed among *Haemogregarina* spp. might have broader implications. In Chapter 2.2, the full mitochondrial genomes of two *Haemogregarina* spp. were sequenced with Sanger-sequencing by means of primer walking. The two mitochondrial genomes generated resemble that of other *Haemogregarina* spp. The protein-coding genes' high sequence diversity is consistent with that found in previous studies. Additionally, like other haemogregarine protein-coding genes the COIII gene has the highest sequence diversity compared to the COI and CytB genes. However, it remains unclear whether the sequence diversity of *Haemogregarina* mitochondrial protein-coding genes reduce their resolution or usefulness in phylogenetic analyses. More mtDNA data will need to be generated for the COI and CytB genes of other *Haemogregarina* spp. for comprehensive phylogenetic studies to be conducted.

Key words: *Haemogregarina,* haemogregarines, Adeleroina, mitochondrial genome, mitochondrial protein-coding genes, CDS, cytochrome *c* oxidase subunit I (COI), Sanger-sequencing

Key findings

- Two full mitochondrial genome sequences ~6200bp for two *Haemogregarina* spp. parasitising South African turtles were generated.
- The mitochondrial genome organisation is conserved among the two species. However, high sequence variability between the protein-coding genes was observed.
- 18S rDNA gene analysis reveals that the two *Haemogregarina* spp. in this study are not as closely related as thought for species parasitising a mutual host.

2.2.2. Introduction

Haemogregarina (Apicomplexa; Coccidiamorphea; Adeleorina), is one of three genera in Haemogregarinidae, which are heteroxenous, intraerythrocytic blood parasites that infect turtles and are transmitted by leeches (Telford, 2009; Rossow *et al.,* 2013; Gutierrez-Liberato *et al.,* 2021). Although often referred to as 'haemogregarines,' this is not a taxonomic classification but refers to all Adeleorinid taxa that undergo a heteroxenous life cycle (Desser, 1993). The haemogregarines include those from Haemogregarinidae, Hepatozoidae, Karyolysidae, and Dactylosomatidae.

Traditionally, the identification of haemogregarines has relied on parasite morphology of different life stages, their associations with specific hosts, and their biological traits (Barta *et al.,* 2012; Attia El Hili *et al.,* 2020; Gutierrez-Liberato *et al.,* 2021). However, this approach has proven challenging due to the similarities in morphological features among species and the limited host specificity of certain species (Leveille *et al.,* 2015). Consequently, accurately identifying haemogregarine species is difficult without a comprehensive understanding of their morphology and the significance of these morphological features (Barta *et al.,* 2012; Telford, 2009). To address the challenges in haemogregarine identification, researchers have turned to molecular markers in conjunction with morphological descriptions to resolve taxonomic and evolutionary issues within the phylum Apicomplexa (Correa *et al.,* 2022; Gutierrez-Liberato *et al.,* 2021). The conserved nuclear 18S rRNA gene is a common genetic marker for haemogregarine phylogenetics (Gutierrez-Liberato *et al.,* 2021). However, this gene may not be divergent enough to distinguish between closely related species in large groups, such as the *Hepatozoon* or *Eimeria* (Barta *et al.,* 2012; Ogedengbe *et al.,* 2011; Gutierrez-Liberato *et al.,* 2021). Therefore, the more rapidly evolving cytochrome *c* oxidase subunit I (COI) mitochondrial protein-coding gene has been suggested as an alternative marker for haemogregarines (Barta *et al.,* 2012). The COI is one of three protein-coding genes present in the mitochondrial genome of Apicomplexans, the other two being cytochrome *c* oxidase subunit III (COIII), and cytochrome b (CytB).

Sequence data for all three protein-coding genes (COI, COIII, and CytB) have been generated for various haemogregarines, including *Haemogregarina* (Gutierrez-Liberato *et al.,* 2021). In the study by Gutierrez-Liberato *et al.* (2021), these genetic markers were utilized to deduce phylogenetic relationships and evaluate evolutionary distances. However, when these markers were applied to assess the genetic diversity of *Haemogregarina* spp. in turtles from Colombia, distinct evolutionary lineages were observed within the same phenotype. This discovery poses a challenge in accurately defining species solely based on genetic markers, either individually or in combination.

The aim of this study was to obtain the complete mitochondrial genome sequences of two *Haemogregarina* spp. that infect terrapins in South Africa. Throughout this paper, the term “organisation of the mitochondrial genome” will refer to the arrangement and positioning of genes and RNA fragments in the mitochondrial genome. The objectives of this study were to 1) sequence the partial 18S rDNA sequence to determine the placement of both species within *Haemogregarina*; 2) sequence the full mitochondrial genomes of both *Haemogregarina* spp. by means of Sanger-sequencing through primer walking.

2.2.3. Materials and methods

2.2.3.1. DNA extraction

Archived whole blood samples of two serrated hinged terrapins (*Pelusios sinuatis*) was used for molecular analysis. Both blood samples were collected from northern Kwazulu-Natal. One from the Ndumo game reserve (-26.884909, 32.312398) and the other near the border of Swaziland and Mozambique (-27.014660, 32.14169). DNA extraction was performed from the whole blood samples using the NucleoSpin®Tissue Genomic DNA Tissue Kit (Macherey-Nagel, Düren, Germany). This study received the relevant ethical approval: (North- West University ethics approval no NWU-01399-23-A9).

2.2.3.2. Polymerase chain reaction

Polymerase chain reaction (PCR) was employed to amplify the 18S rRNA gene and mitochondrial genome sequences. The primers and PCR conditions are detailed in Table 3. For each PCR reaction, a final volume of 25 µl was prepared. The amount of DNA used varied between 1 µl and 5 µl, depending on the specific experiment. Two different PCR master mixes were used: DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific, Johannesburg, SA) from Thermo Fisher Scientific for target fragments below 1000bp, and Q5® High-Fidelity 2X Master Mix (New England Biolabs) from New England Biolabs for target fragments above 1000bp. PCR reactions were performed according to the manufacturer instructions, using primers at a final concentration of 0.5 mM.

2.2.3.3. Primer design

In this study, a combination of universal primers for apicomplexans and primers specifically designed for *Hepatozoon* and Adeleorinid parasites were utilised. The primers selected for aiding the amplification of the full mitochondrial genomes are listed in Table 3. During primer design, careful consideration was given to reference mitochondrial genomes such as that of *Klossiella equi, Klossia helicina, Klossia razorback, Hepatozoon catesbianae,* and *Hepatozoon griseisciuri*. These reference genomes served as a basis for primer selection. In the above-mentioned species of *Hepatozoon* mitochondrial genomes, the LSUG and LSUF rRNA regions are located on opposite poles from each-other. Initially, primers targeting these two RNA regions were used to lay a foundation for subsequent amplifications. Based on the sequences obtained from these initial primers, predictions were made regarding the organisation of the mitochondrial genomes.

Table 3: Primers used to generate the sequences of the 18S and mitochondrial genome fragments.

Barcode	Primer name	Primer sequence	Annealing temp (˚C)	Extension time	Fragment size	Reference
18S	Adel_18S_13F Adel_18S_1522R	5′-CCTGCCAGTAGTCATATGCT-3′ 5'-AYCCTATTTAGCAGGTTAAGGT-3'	56	2:00	~1300bp	Léveillé *et al.*, 2021 Léveillé *et al.*, 2020
Mitochondrial genome	Api_LSUG_UNI_F Api_LSUF_UNI_R Api_LSUG_R Api_LSUF_F Api_LSUE_UNI_F Api_LSUE_UNI_R Api_SSUE_Uni_R Api_SSUE_Uni_F Api_RNA8_F Api_RNA8_R SSUF_R SSUF_F Api_RNA10_Uni_R Api_RNA10_Uni_F Api_COI_VWXHHM_F	5′-TTGAGRCAGTTTGTTCCCTATCT-3′ 5′-ACCTGTTATCCCCGGCGWA-3′ 5′-AGATAGGGAACAAACTGYCTCAA-3′ 5′-GTWCGCCGGGGATAACAGGT-3′ 5’-GGTAAGACCCTGAGCACCT-3’ 5’-AGGTGCTCAGGGTCTTACCG-3’ 5’-CGTGACGAGCGGTGTGT-3’ 5’-ACACACCGCTCGTCACG-3’ 5’-CACYGGATTGGATACCCRG-3’ 5’CYGGGTATCCAATCCRGTG-3’ 5’-TGACRGTGAACTTGTGGCTG-3’ 5’-CAGCCACAAGTTCACYGTCA-3’ 5’-GCGGTTAAYCTTWCCTWTTCCTTAC-3’ 5’-GTAAGGAAWAGGWAAGRTTAACCGC-3’ 5’-GTNTGGGBHCAYCAYATG-3’		2:30	~1000-3000	Léveillé *et al.*, 2019c

Description: The primers used in this study to sequence the 18S markers and sequences of the mitochondrial genome of the two *Haemogregarina* spp. of this study. The target DNA, primer names, sequences, PCR conditions and references are listed.

2.2.3.4. Sequencing, genome construction and annotation

PCR samples were sent to Inqaba Biotech (a commercial sequencing company) for sequencing. Sanger-sequencing was used for all DNA samples. Sequenced data was analysed and edited using Geneious R9 software. For the alignment of 18S sequences and, in some cases, mitochondrial sequences, a DeNovo alignment approach was used. The map to reference tool was used to map mitochondrial genome sequences of *Haemogregarina* to the mitochondrial genomes of *Klossia* and *Klossiella* spp. Annotation of the complete mitochondrial genomes was performed using the annotate and predict tool within Geneious R9. During the annotation process, a sequence similarity threshold of above 60% was used, and annotations with the highest similarity were chosen. The Open Reading Frames (ORFs) tool within Geneious was used to confirm the position and size of protein-coding genes with more accuracy. The mold protozoan mitochondrial genome served as the genetic code to identify ORFs.

2.2.3.5. Phylogenetic analysis

Nuclear DNA sequences for species of *Hepatozoon* Miller, 1908, *Haemorgegarina* Danilewsky, 1885, *Hemolivia* Petit *et al.,* 1990, *Klossia* Schneider, 1875*, Klossiella* Smith and Johnson, 1902*,* and *Eimeria* Schneider, 1875, were downloaded from GenBank and aligned to the 18S rDNA sequences generated in this study. *Eimeria innocua* is a distant relative to the adeleorinids and was selected to root the phylogenetic tree. Sequences were aligned with the Muscle alignment tool implemented in Geneious R9. A model test was performed to determine the best substitution model according to the Bayesian information criterion (BIC), using jModelTest 2.1.10 (Darriba *et al.,* 2012). MrBayes v3.2.6 (Huelsenbeck & Ronquist, 2001) was used to construct the 18S phylogenetic tree using the General Time Reversal (GTR) model with gamma distribution, a portion of the sites considered invariable. The Markoc Chain Monte Carlo (MCMC) algorithm was run for 10 million generations, sampling every 100 generations. The first 25% of the trees were discarded as burn-in.

2.2.4. Results

2.2.4.1. Mitochondrial genomes

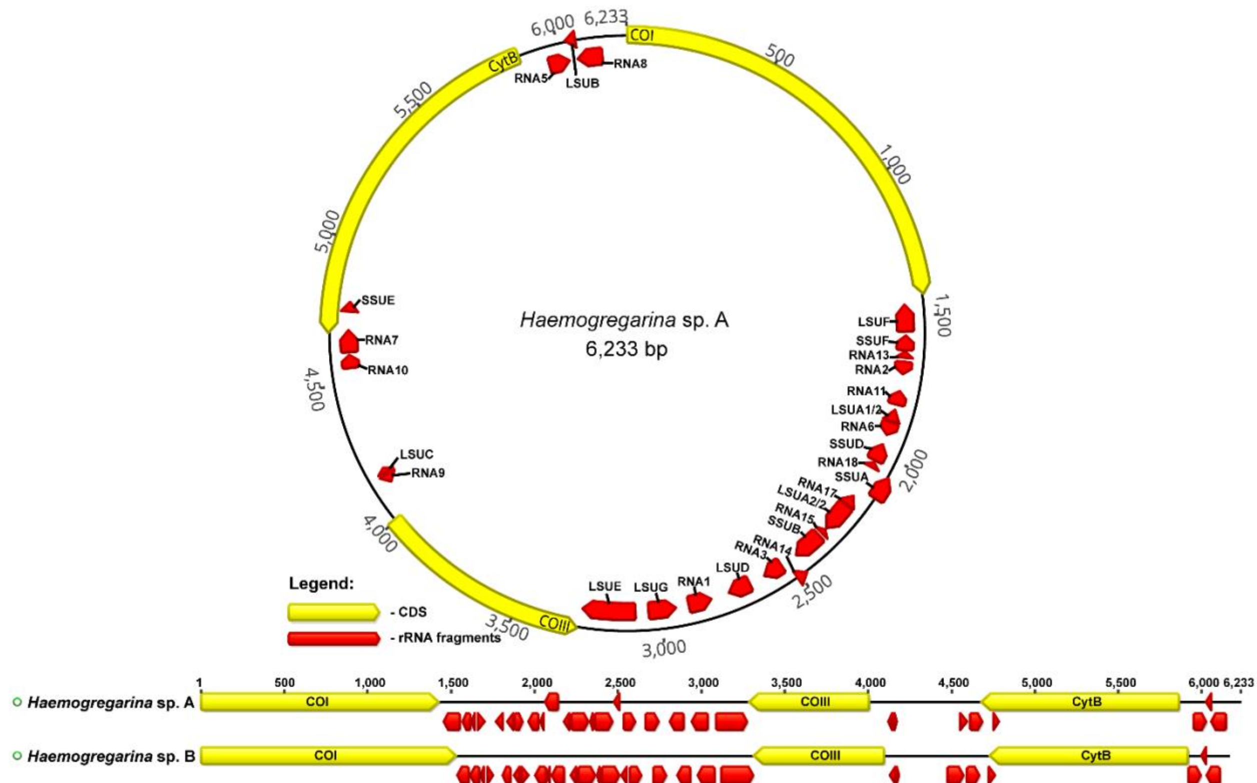


Figure 2.2.1: The mitochondrial genome of two *Haemogregarina* spp. parasitising the Serrated Hinged Terrapin *Pelusios sinuatus*. Above: The mitochondrial genome of *Haemogregarina* sp. A presented in circular form. Below: The mt-genomes of both *Haemogregarina* sp. A and B of this study is compared and presented in linear form. The genomes consist of three protein-coding genes (CDS; cytochrome c oxidase subunit I (COI), cytochrome c oxidase subunit III (COIII), cytochrome b (CytB)) and 28 ribosomal RNA fragments (rRNA). The location and orientation of CDS genes and rRNA fragments are identical for both mt-genomes although sequence similarity between the CDS genes are very low (>65%).

To construct the full mitochondrial genomes of the two *Haemogregarina* spp., approximately 33-35 overlapping sequences with an average size of 819bp were used. *Haemogregarina* sp. A from the Ndumo sample has a sequence length of 6233bp and consists of 38.4% GC base pairs, as illustrated in Figure 2.2.1. On the other hand, *Haemogregarina* sp. B has a sequence length of 6167bp and consists of 38.9% GC base pairs, also illustrated in Figure 2.2.1. The position and size of each protein-coding gene and rDNA region are summarised in Supp. Table 2. The %GC content of the protein-coding genes for *Haemogregarina* sp. A and *Haemogregarina* sp. B are 40.3% & 40.1% for COI, 35.4% & 38.3% for COIII, and 38.9% & 38.4% for CytB, respectively. There are a total of 28 rDNA fragments present, which includes 16 large subunit (LSU) rDNA, 10 small subunit (SSU) rDNA, and 2 unassigned fragments, which is similar to the findings in other adeleorinids (Léveillé *et al.,* 2019b; 2020; 2021). The mitochondrial genomes of both

Haemogregarina spp. exhibit a conserved organisation with other adeleorinids, which is also consistent with the organisation observed in *Klossia* spp. (Figure 2.2.2).

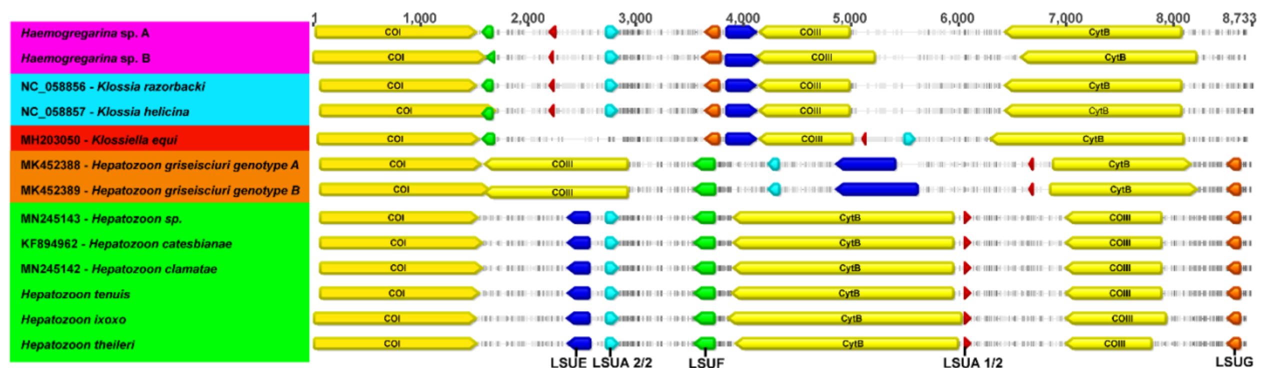


Figure 2.2.2: This illustration compares the organisation of available mitochondrial genomes of haemogregarines in adeleorina. The first nucleotide of every mitochondrial genome was set at the start of the COI gene and aligned. All three protein-coding genes are indicated and coloured along with some of the RNA fragments (LSUF, LSUG, LSUA 1 & 2, and LSUE). These orientations and positions of the protein-coding genes and RNA fragments represent the similarities between closely related species and dissimilarities between genera. Species within the green box represent species of *Hepatozoon* parasitising anuans hosts and orange those parasitising rodent hosts. The two *Haemogregarina* spp. of this study is highlighted by the pink box. The blue and red box indicate species of *Klossia* and *Klossiella* respectively.

The protein-coding genes of both isolated genomes were aligned and compared as shown in Figure 2.2.3. The geneious alignment tool of Geneious R9 was used to align the COI, COIII and CytB mitochondrial protein-coding gene sequences of both *Haemogregarina* spp. with a 93% similarity (5.0/-9.026168) from the present study. The sequence similarity is visualized by base pair colouration, where identical base pairs are coloured black, while dissimilar base pairs are coloured white, resulting in a barcode-like appearance.

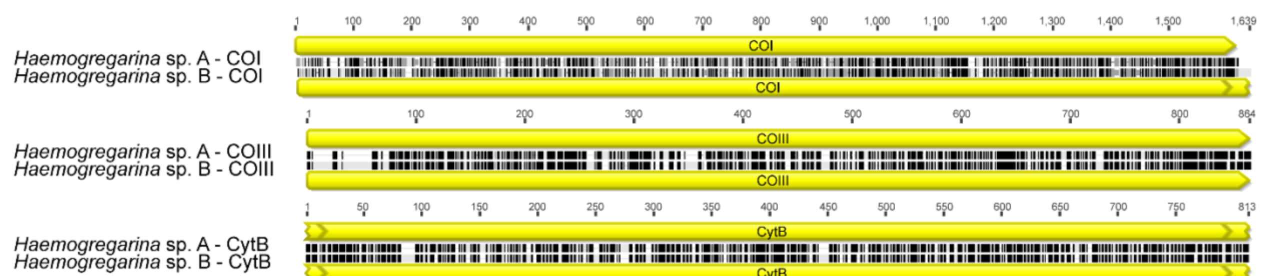


Figure 2.2.3: The CDS genes of the two mt-genomes generated in this study is aligned individually with 93% similarity cost matrix. The sequences are visualized and stylised to indicate similarity between each CDS pair. Black lines indicate similar base pairs, while white lines indicate dissimilar base pairs. Top: The COI CDS of both *Haemogregarina* spp. have 62.2% sequence similarity. Middle: The COIII CDS of both *Haemogregarina* spp. have 58.9% sequence similarity. Bottom: The CytB CDS of both *Haemogregarina* spp. have a 62.5% sequence similarity.

The pairwise identity between the two samples was calculated for each gene. The COI gene has a pairwise identity of 62.2%, 58.9% for the COIII gene, 62.5%. for the CytB gene. It was difficult

to obtain high quality sequence data for the COI gene regions. Sequenced data from either the start or end of the gene region is of high quality for the first 800 basepairs, afterwhich sequence quality drops.

2.2.4.2. 18S rDNA phylogenetic analysis:

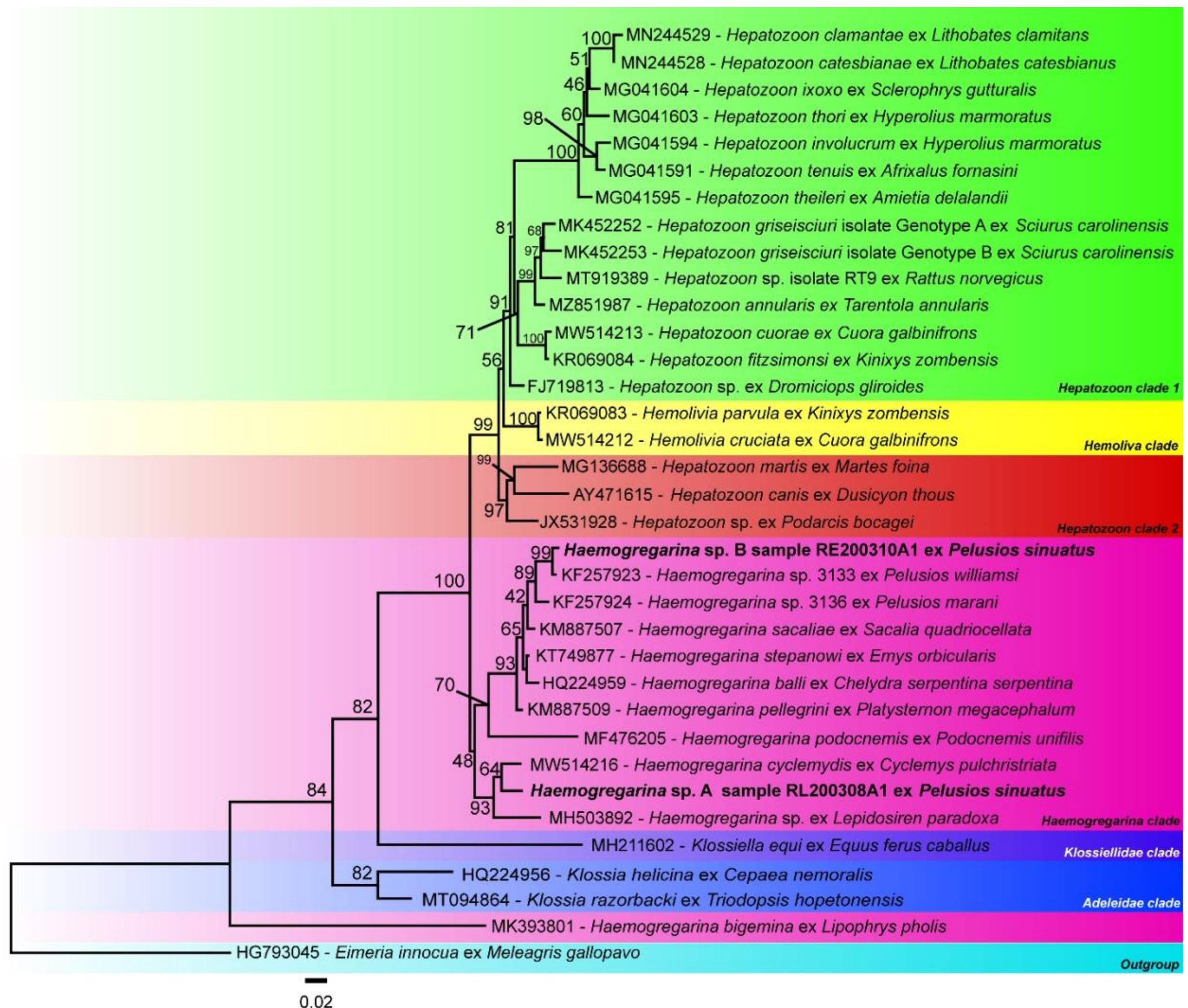


Figure 2.2.4: The 18S rDNA sequence of multiple adeleorinid haemoparasites was used to construct the phylogenetic tree above. An *Eimeria innocua* (HG793045) 18S rDNA sequence was used to root the tree. The two *Haemogregarina* spp. of this study are highlighted in bold. The 18S rDNA sequences for *Haemogregarina* used in the tree formed a well-supported clade, except for *Haemogregarina bigemina* (MK393801). The *Haemogregarina* clade is split into two poorly supported subclades with *Haemogregrina* sp. A grouped with species parasitising a turtle and a lungfish, and *Haemogregarina* sp. B grouped with species parasitising turtles, especially other *Pelusios* spp.

The nuclear 18S rDNA region for both *Haemogregarina* spp. A and B was successfully isolated and sequenced. The partial 18S rDNA sequence obtained for species A (1249bp) showed a similarity of 98.24% with *Haemogregarina cylemydis* (MW514216) isolated from an Eastern

Black-bridged-Leaf Turtle (*Cyclemys pulchristriata*) in Vietnam, with complete query cover. Similarly, the partial 18S rDNA sequence of species B (1253bp) has a 99.66% identity with an unclassified *Haemogregarina* sp. (KF257923) isolated from a Williams' mud turtle (*Pelusios williamsi*) in Kenya. The 18S rDNA sequences were used to construct a phylogenetic tree that included sequences from various apicomplexan parasites such as *Hepatozoon, Hemolivia, Haemogregarina, Klossia, Klossiella,* and *Eimeria*, obtained from the NCBI database (Figure 2.2.4). Despite both *Haemogregarina* spp. from this study parasitising the same host, they are placed within different subclades within the larger *Haemogregarina* clade. *Haemogregrina* sp. A is placed with two *Haemogregarina* spp., one parasitising a lungfish and the other a turtle. *Haemogregarina* sp. B is placed with *Haemogregarina* spp. parasitizing *Pelusios* hosts, unlike *Haemogregarina* sp. A. The analysis also showed that species of *Haemogregarina,* with the exception of *Haemogregarina bigemina,* formed a well-supported monophyletic clade.

2.2.5. Discussion

Several limitations have been faced by molecular phylogenetic studies using the 18S rDNA sequence and with the use of mitochondrial genetic markers for apicomplexan haemoparasites, primarily due to a lack of available data. Certain parasites within this group, like *Haemogregarina* spp., have not received as much attention as medically significant apicomplexans such as *Plasmodium* spp., contributing to the lack of comprehensive information in the field. In the present study, we sought to bridge the knowledge gap by conducting full mitochondrial genome sequencing of unclassified *Haemogregarina* spp., thereby contributing valuable insights to the field. The mitochondrial genome organisation of both *Haemogregarina* spp. from the present study closely resembles that of other *Haemogregarina* spp. (Léveillé*,* 2019a). The genomes have similar sizes ~6500bp and possess a similar %GC content. The circular mapping of the mitochondrial genomes contains the three protein-coding genes (COI, COIII, and CytB), and 28 rDNA fragments common in apicomplexan haemogregarines (Leivelle *et al.,* 2015).

The COI protein-coding gene for *Haemogregarina* sp. A was difficult to sequence. Despite all attempts to gather high quality sequences for the middle region of the gene, some nucleotides within the sequence remain unresolved. After extensive comparisons with the complete COI gene of *Haemogregarina* spp. B and the DNA fragments generated for the region, no real sense could be made of the region to perhaps predict the nucleotides. This could have been due to a few possabilities such as the interference of mixed infections, degraded DNA, or PCR inhibitors. The blood samples from which DNA was extracted were not fresh, which could mean that the DNA was degraded. It is difficult to pin point the cause of the problem without further analyses to explore different hypotheses. It is unlikely to have been mixed infection since the same phenomenon would have been observed in other regions such as the COIII and CytB, but no interference were observed. Furthermore, the signals generated at the poor quality regions are scattered and speratic, rather than coherent and forming double peaks. Because of these reasons it is likely that degraded DNA or PCR inhibitors are to blame. At the start of this study the mitogenome of *Haemogregarina* sp. A was generated through next generation sequencing (NGS) however there were some implications that made the sequence unusable. Nonetheless, the protein-coding genes were sequenced and were useful to compare with the mitogenome of sp. A generated in this study. Surprisingly, the NGS sequence of the genome also had degenerate base calls in the same region of the COI gene. The fragment DNA sequenced for the COI gene of sp. A all displayed the same characteristic. For each fragment the first 500bp of the 1000bp fragment was very high quality followed by an immediate drop in quality for the remaining 500bp. This drop in sequence quality overlapped the middle region of the COI gene for sp. A in all cases. Why this is the case can be due to several reasons but is probably best explained by mixed signalling or degraded DNA.

The protein-coding genes of the two mitochondrial genomes were compared (Figure 2.3.3). The pairwise identity showed high variability in gene sequences between these two species, with COI exhibiting 62.2% pairwise identity, COIII showing 58.9% pairwise identity, and CytB displaying 62.5% pairwise identity. The respective alignment of the three protein-coding genes within each mitochondrial genome in this study showed more variability than expected from closely related species, with approximately 37.6% to 39% pairwise variability across the genes. To demonstrate the pairwise sequence similarity between closely related species, two *Hepatozoon* spp. that infect anuran hosts from Canada was used, specifically *Hepatozoon catesbianae* and *Hepatozoon clamantae* (Figure 2.2.5).

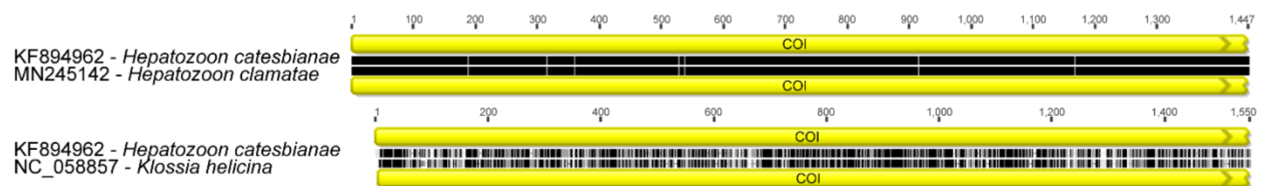


Figure 2.2.5: Top: The COI protein-coding genes of *Hepatozoon clamantae* and *Hepatozoon catesbianae* is compared. The two sequences have a 99.5% pairwise indicated by the similar black regions with dissimilar white regions in between. **Bottom:** The COI protein-coding genes of *Klossia helicina* and *Hepatozoon catesbianae* is compared to indicate pairwise identity between the two genes of two different genera. These genes are 71.7% identical, represented by the scattered white and black lines, black indicating similarity and white indicating dissimilarity.

In this example, the pairwise similarity of the COI protein-coding gene between the two closely related *Hepatozoon* spp. is higher (99.5% pairwise identity) than what was observed in the *Haemogregarina* spp. of this study. Furthermore, while not presented in this context, similar pairwise similarity patterns were observed for the other two protein-coding genes of *Hepatozoon catesbianae* and *Hepatozoon clamantae*. Only when comparing the protein-coding genes of distantly related species, such as the COI of *Klossia helicina* with *Hepatozoon catesbianae* (see Figure 2.2.5), is pairwise identities below 70% seen. The pairwise identities between the CDS genes of the *Haemogregarina* spp. of this study with *Klossia helicina* is similar to those between the CDS genes of *Klossia helicina* and *Hepatozoon catesbianae*. In theory, this could mean that the two *Haemogregarina* spp. in this study are not of the same genus, however the organisation of the mitogenome contradicts this.

The mitochondrial genome organisation along with the phylogenetic analysis of the 18S rDNA suggest the two *Haemogregarina* spp. in this study are of the same genus. In Figure 2.2.2, multiple adeleorinid mitochondrial genomes are compared to illustrate conserved genome organisation among closely related species within the same genus. The mitochondrial genome organisation of *Hepatozoon* spp. that infect anurans is conserved, as is the genome organisation of *Hepatozoon* spp. that infect rodents. Similarly, the mitochondrial genome organisation of

adeleorinids in the genus *Klossia* is conserved. However, when compared, the mitogenome organisations across different genera have notable differences. This may suggest that the mitogenome organisation is conserved among species from the same genus. This statement does contradict the data displayed since *Hepatozoon* spp. parasitising anurans and *Hepatozoon* spp. rodents are still of the same genus, but the genome organisation differs between the two. In short, the genus *Hepatozoon* is a large haemogregarine group currently under revision to resolve its monophyly. It is likely that the group will be split into multiple new genera, unfortunately due to a lack of compelling evidence the issue has not yet been resolved. The differences observed between the mitogenome organisations of different genera and species may help point taxonomists in the right direction. It is also worth noting that the mitogenome organisation of the *Haemogregarina* spp. in this study is similar to that of the unpublished mitogenomes sequenced by Léveillé *et al.* (2019a). Thus, based on the mitogenome organisation the two haemogregarines of this study are from the genus *Haemogregarina*. Interestingly, the mitogenome organisation of *Klossia* spp. is similar to that of *Haemogregarina* spp. and, to a lesser extent, *Klossiella* spp. mitochondrial genomes. It could possibly be that mitochondrial genome organisation and classification are linked, which could mean the two *Haemogregarina* spp. in this study have shared evolutionary history with the *Klossia* spp.

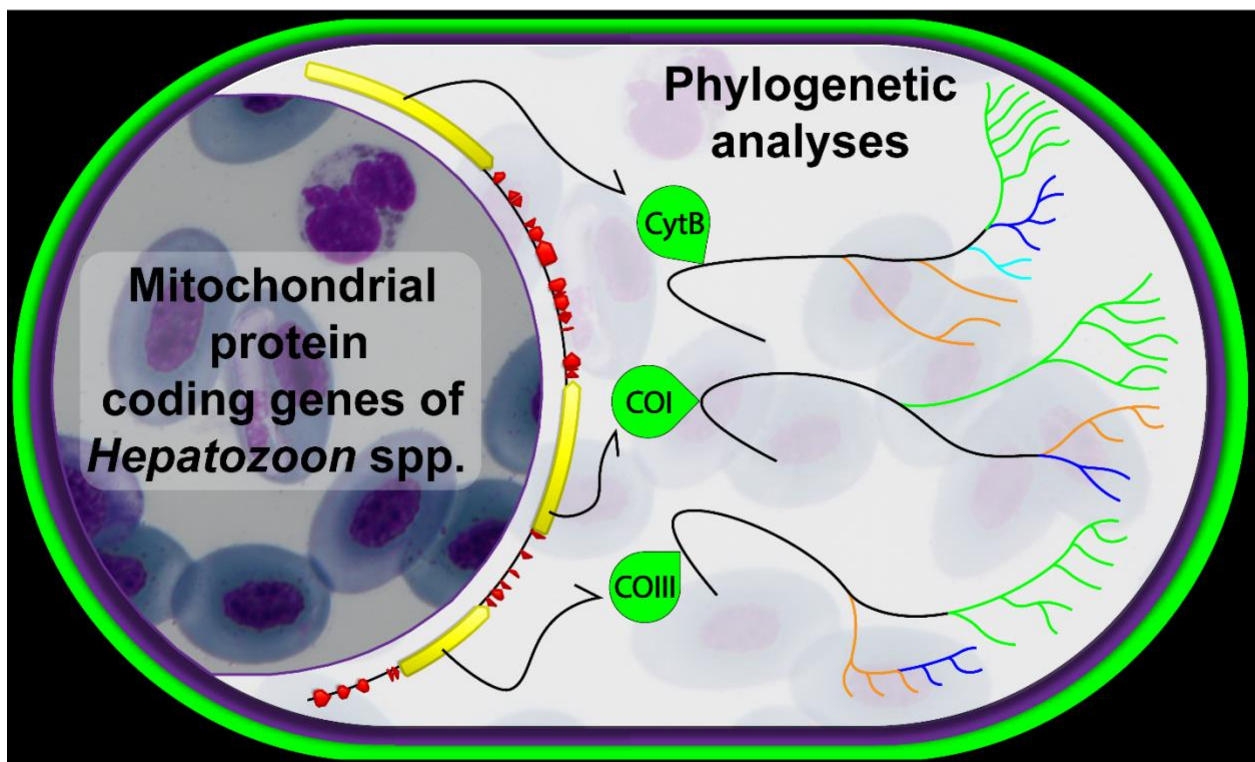
The 18S rDNA gene is useful for inferring phylogenetic relationships among genera and species. In the present study, the 18S rDNA sequences of both *Haemogregarina* spp. were used to compare their relatedness to other adeleorinids. The phylogenetic analysis revealed that both sequences clustered within the *Haemogregarina* clade (see Figure 2.2.4). Furthermore, the *Haemogregarina* clade itself is divided into two poorly supported subclades, with *Haemogregarina* sp. A and B falling into separate subclades. This division could potentially explain the observed diversity in the protein-coding genes between the two species. The sequence diversity in the protein-coding genes may indicate unique evolutionary patterns within the *Haemogregarina* genus, underscoring the importance of incorporating additional molecular markers and comprehensive analyses to further elucidate the evolutionary relationships and taxonomic classification within this group of haemoparasites. Overall, the 18S rDNA gene remains a valuable marker for studying the phylogenetic relatedness of *Haemogregarina* spp., while the mitochondrial protein-coding genes provide additional insights that should be considered alongside other genetic and morphological data to fully understand the evolutionary relationships among *Haemogregarina* spp. and other haemogregarines.

2.2.6. Conclusion

The analysis of mitochondrial gene sequences has revealed high diversity among *Haemogregarina* spp. highlighting potential challenges using these genes for species identification or phylogenetic analysis. The addition of more mitochondrial DNA from these and other haemogregarines may aid in the analysis of these genes as either additional or supplementary barcodes for phylogenetic analyses. The high sequence diversity between the CDS genes of *Haemogregarina* spp. raises some concerns. The main concern would be if despite the high sequence diversity, the CDS genes of *Haemogregarina* spp. still poses phylogenetic or evolutionary information. It may be that they do, but how this will affect phylogenies remains to be fully explored. The 18S rDNA phylogenetic analysis of the two *Haemogregarina* spp. in this study suggests that these two haemogregarines are two different species of the same genus parasitising the same host. It is worth investigating the host specificity of *Haemogregarina* and how this relates to other haemogregarines. The aim of this investigation was to expand the dataset of haemogregarine mitochondrial genetics by providing full genomes of two additional *Haemogregarina* spp. This data will contribute to more robust assessments of the mitochondrial gene diversity and its implications for species identification and phylogenetic analysis within the Adeleorina group. Such efforts will enhance our knowledge of haemogregarine evolution and aid in resolving the complex relationships among different species and genera.

CHAPTER 2.3

A comparative phylogenetic analysis of mitochondrial protein-coding genes from *Hepatozoon* spp. (Adeleorina: Hepatozoidae), parasitising South African vertebrates



CHAPTER 2.3

A comparative phylogenetic analysis of mitochondrial protein-coding genes from *Hepatozoon* spp. (Adeleorina: Hepatozoidae), parasitising South African vertebrates.

Tiaan Haarhoff, Marianne Venter[2], Louis H. du Preez[1,3], and Edward C. Netherlands[1,4]*

[1] African Amphibian Conservation Research Group, Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa

[2] Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom, South Africa.

[3] South African Institute for Aquatic Biodiversity, Somerset Street, Makhanda 6140, South Africa

[4] Department of Zoology and Entomology, University of the Free State, PO Box 339, Bloemfontein, 9300, South Africa.

***Author for correspondence:** EC Netherlands, E-mail: NetherlandsEC@ufs.ac.za

Word count: 9876

CHAPTER 2.3

2.3.1. Abstract

Of the Apicomplexans, the genus *Hepatozoon* (Adeleorina: Haemogregarinidae) has mostly been placed on the parasite back burner in favour of other more medically important parasites. However, *Hepatozoon* have been recorded in all vertebrate classes and merit more attention. Nevertheless, *Hepatozoon* spp. are major herpetological parasites, accounting for the majority of reported infections in Southern Africa. Currently, the genus comprises of more than 300 species worldwide of which the phylogenetic placement of certain species have been somewhat of a debate. Although it is evident that the genus needs to be subdivided into multiple genera, the methodology for precisely delineating these divisions remains uncertain. The cytochrome *c* oxidase subunit I (COI) mitochondrial protein-coding gene sequence has been proposed as an alternative genetic marker to the traditional nuclear 18S rRNA gene sequence. The gene has proven useful in delineating recent speciation events in other apicomplexan parasites. In this study the mitochondrial protein-coding genes COI, cytochrome *c* oxidase subunit III (COIII), and cytochrome B (CytB) were amplified and sequenced to analyse their ability to resolve recent speciation events within *Hepatozoon*. A total of 34 protein-coding gene sequences were generated and analysed. The investigation revealed that combining the COI and COIII gene sequences proves most effective in resolving the deeper nodes of the phylogenetic tree. Combining these two gene sequences with the 18S and CytB sequences further aid in resolving these deeper nodes. More data will be needed to determine whether the same conclusions is reached with a larger pool of samples. Considering the challenges posed by the high sequence variability of the COIII gene, it may prove beneficial to undertake whole mitochondrial genome sequencing of *Hepatozoon* spp., particularly given the difficulties encountered in designing primers for the COIII region.

Key words: *Hepatozoon*, mitochondrial protein-coding genes, cytochrome *c* oxidase subunit, COI, COIII, CytB, phylogenetic analysis

2.3.2. Introduction

The genus *Hepatozoon* Miller, 1908 (Apicomplexa: Adeleorina) is a group of intraerythrocytic or intraleukocytic apicomplexans parasitising all vertebrate classes, including mammals, fish, birds, reptiles, and amphibians. Species of *Hepatozoon* are generally referred to as haemogregarines since they follow a heteroxenous life cycle, parasitising a vertebrate intermediate host and a definite hematophagous invertebrate host or vector (Desser, 1993; Netherlands *et al.,* 2018).

Hepatozoon are frequently described from amphibians and reptiles (Cook *et al.,* 2018), and are the most reported haemogregarines from anurans (Netherlands *et al.,* 2018). Although there are more than 300 species of *Hepatozoon* (Smith, 1996) only a few have been described from South African reptiles and amphibians. Those of reptiles include *H. varani* Laveran, 1905, *H. refringens* Sambon and Seligmann, 1907, *H. bitis* Fantham, 1925, *H. fitzsimonsi* Dias, 1953, *H. langii* Van As *et al.,* 2013, *H. vacuolatus* Van As *et al.,* 2013, *H. affluomaloti* Van As *et al.,* 2015, *H. angeladaviesae* Cook, Netherlands *et al.,* 2018, and *H. cecilhoarei* Cook, Netherlands *et al.,* 2018. Those of amphibians include *H. theileri* Laveran, 1905, *H. ixoxo* Netherlands *et al.,* 2014a, *H. involucrum* Netherlands *et al.,* 2017, *H. tenuis* Netherlands *et al.,* 2017, *H. thori* Netherlands *et al.,* 2017.

Historically, the identification and description of haemogregarines was based on the morphological characterisation of the different stages of life within infected host cells (Barta *et al.,* 2012; Smith & Desser, 1997). However, according to Barta *et al.* (1997), the morphology of haemogregarines is unreliable in accurately classifying these parasites. In response, molecular phylogenetic analysis was suggested to supplement classification with the added benefit that knowledge with respect to complex life cycles is not needed (Barta *et al.,* 1997). The 18S rDNA sequence has been used in various studies and has proven to be useful for phylogenetic classification and analysis of Apicomplexans (Criado-Fornelio *et al.,* 2006; Kopečná *et al.,* 2006; Rubini *et al.,* 2006; Ogedengbe *et al.,* 2011; Barta *et al.,* 2012; Maia *et al.,* 2016a; O'Donoghue, 2017; Ogedengbe *et al.,* 2018). Although useful, the 18S rDNA gene sequence is relatively conservative and has been shown to be unable to resolve closely related species or clades within a large genus, such as *Hepatozoon,* into monophyletic clades (Ogedengbe *et al.,* 2011; Barta *et al.,* 2012).

The mitochondrial cytochrome *c* oxidase subunit I (COI) protein-coding gene is a faster evolving gene compared to the 18S rDNA gene and has been shown to be capable of delineating closely related species (Lane, 2009; Ogedengbe *et al.,* 2011). Due to some disputes regarding the monophyly of *Hepatozoon*, more focus has been given to COI as a molecular marker to perhaps delineate closely related species within the group. Several complete mitochondrial genomes have been sequenced for *Hepatozoon* spp. that infect anurans and rodents in Canada and the Czech Republic, and partial mitochondrial genome fragments have been sequenced for

Hepatozoon canis (Léveillé *et al.,* 2015; Léveillé *et al.,* 2020; Léveillé *et al.,* 2021; Hrazdilová *et al.,* 2021; Léveillé *et al.,* 2019b). Thus, there are an estimated 12 COI sequences available for phylogenetic analysis of *Hepatozoon* spp. Three full mitochondrial genomes are available for anurans: *H. catesbianae* (NC_044466), *H. clamatae* (MN245142), and an unclassified *Hepatozoon* sp. (MN245143) parasitising *Lithobates pipiens*. Seven full mitochondrial genomes are available for rodents: *H. griseisciuri* genotypes A and B (MK452388 & MK452389), and five unidentified *Hepatozoon* spp. (MT936929 - MT936933) parasitising *Rattus* spp. of which four are near identical. Two sequences of COI for *H. canis* are available (MH557087 & MH615002). For most, if not all of these species, the COIII and CytB protein-coding genes from the mitochondrial genome are also available.

There is a clear lack of mitochondrial genome data for *Hepatozoon* and considering the polyphyly of the genus and its impact on the conclusions of the biology and evolution of the species within the genus. The aim of this study was to generate additional mitochondrial DNA for *Hepatozoon* spp. to be used for phylogenetic analyses. The objectives of this study were to 1) design primers that would aid in amplification of the three mitochondrial protein-coding genes (COI, COIII, and CytB); 2) amplify and sequence the protein-coding genes of eight *Hepatozoon* spp.; 3) analyse the phylogenetic properties and usefulness of each protein coding gene.

2.3.3. Materials and Methods

2.3.3.1. DNA extraction

DNA was extracted from whole blood samples from (n = 4) snakes, (n = 2) turtles, and (n = 2) frogs. Some archived samples were made available as extracted DNA thanks to ECN. When required, DNA was extracted with a NucleoSpin®Tissue Genomic DNA Tissue Kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions, and using 25–50 µl whole blood samples stored in 100% alcohol. This study received the relevant ethical approval: (North-West University ethics approval no NWU-01399-23-A9).

The NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) kit was used as instructed by the manufacturer to isolate PCR products from otherwise 'dirty' PCR results that had undesired amplified DNA. Approximately 6 µl of PCR product was loaded onto a 1% agarose gel stained with a standard gel stain and run for 35 min at 50V and 400A. A 1kb ladder was loaded in a separate well as a reference. After electrophoresis, gels were viewed under a UV light source with a protective UV filter cover. The desired PCR bands were carefully cut from the gel and then processed with the Clean-up kit. From these extracted DNA, standard PCR procedures were performed to amplify the desired DNA.

2.3.3.2. PCR procedures

Two working volumes were used throughout the duration of this study. For troubleshooting and testing purposes, a mixture volume of 12.5 µl was used and 25 µl mixture volumes for sequencing. In some cases, the PCR tubes used were faulty, which caused the mixtures to leak during temperature cycles. The faults were easily observed with 12.5 µl mixtures but hardly noticeable with 25 µl mixtures. Reaction mixtures of 12.5 µl consisted of: 6.25 µl master mix, 0.625 µl of each forward and reverse primer, DNA template (usually 1-2 µl), and nuclease free water (usually 3-4 µl) at a volume that brings the final reaction volume to 12.5 µl. Reaction mixtures of 25 µl uses twice the volume of reagents used in 12.5 µl reaction volumes, however, the volume of DNA template was kept between 1-3 µl and nuclease-free water was added to obtain the final reaction volume. Two different master mixes were used throughout, DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific, Johannesburg, SA) and Q5® High-Fidelity 2X Master Mix (New England Biolabs). DreamTaq was used to amplify DNA fragments shorter than 1500 bp, and High-Fidelity was used to amplify DNA fragments longer than 1500 bp. All primer concentrations were 10 µM. The PCR product was sent to Inqaba Biotechnical Industries (Pretoria, South Africa) for sequencing. The chosen sequencing method was Sanger-sequencing due to its accuracy and reliability as well as ease of use.

2.3.3.3. *Primer design and CDS gene sequencing*

The amplification of 18S rRNA gene sequences was performed with the primers Adel_18S_13F 5'-CCTGCCAGTAGTCATATGCT-3′ and Adel_18S_1522_R 5'-AYCCTATTTAGCAGGTTAAGGT-3'. The amplification of the CytB gene was done with Hep_CytB_70F 5'-AATTTYATGTGGAATTTYGGGTTT-3' or Hep_CytB_391F 5'-CCWTGGGGTCAGATGAGTT-3' as the forward primer and Hep_CytB_1025R 5'-GTACACTGAGGTAATTGRCT-3' as the reverse primer. Primers used for the COI region are Hep_COI_329_R 5'-GCWGAGCCTAAGGAACAT-3' and Hep_COI_1733_F 5'-TCCATTTAAMTTCCACTTCTC-3', however Api_LSUG_UNI_R 5'-AGATAGGGAACAAACTGYCTCAA-3' was also used as an alternative to Hep_COI_1733_F. Amplification of the COIII region was done with multiple primers. Samples 07H2 and 07H6 were amplified with primers LSUD_R 5'-CCGGCARATGMCA-3' and LSUB_F 5'-TGTGCYTGGDAGCTGTAATC-3'. Sample 27A1 was amplified with primers Hep_COIII_273_F 5'-GTCTAGCTATTAATCARGTYAAAGCTCATTTACAA-3' and Hep_COIII_971_R 5'-GCCAMACTAATTCWACYAAATGCCAGTA-3'. Sample 16B1 was amplified with primers Hep_COIII_273_F 5'-GTCTAGCTATTAATCARGTYAAAGCTCATTTACAA-3' and Hep_COIII_351_R 5'-AAYTCWACRAAATGCCAGTA-3'. Various primers were designed to supplement the primers available, along with the PCR conditions used. The primers designed in this study are listed in Table 7. Primer design was done using Geneious R9 9.1.8 software (https://www.geneious.com), which utilises Primer3 (Untergasser *et al.,* 2012; Koressaar & Remm, 2007). Some primers were designed manually. Primers were designed according to the process of primer design discussed by Bustin and Huggett (2017). Conserved regions among multiple adeleorinid mitochondrial genomes were chosen to design primers from. Primers were designed that targeted conserved regions among multiple adeleorinid mitochondrial genomes within the COI, COIII, both parts of the LSUA, LSUB, LSUD, and SSUD. The rDNA fragments LSUA, LSUB, LSUD, and SSUD are well conserved among adeleorinids, have low variability, low AT content, and are large enough to design primers of adequate length. These conserved regions were extracted from whole mitochondrial genome sequences and aligned to design primers with the best desired properties. These were that primers should: 1) all have a Tm (annealing temperature) between 54 °C and 59 °C, 2) not be greater than 22 bp, 3) have a GC content above 30%, 4) have a hairpin Tm more than 5 °C lower than the annealing Tm, 5) have a degeneracy lower than 8. The primers were designed close to or within these parameters.

Sequenced DNA chromatograms were visualised, edited, and annotated with Geneious R9. Various alignment tools were used to determine the position of sequenced DNA fragments in mitochondrial genomes of adeleorinid. Each PCR product was sequenced as two overlapping

fragments. A sequence was generated for each forward and reverse primer in a PCR product and aligned with the DeNovo assemble tool. In some cases, mixed signalling causes degenerate base calls in both fragments, resulting in failed assemblies. For these fragments, the Geneious alignment tool was used with a 65% cost matrix, 12 gap penalty, and 3 gap extension penalty. Some fragments were larger than 1600bp, resulting in no overlapping of the two associated sequences. In a case like these, the Map to reference tool was used to map the two sequences to an existing mitochondrial genome from a *Hepatozoon* sp. Each DNA sequence generated in this study was matched with the NCBI database to determine its origin and identity.

2.3.3.4. Phylogenetic analyses

Nuclear and mitochondrial DNA sequences for species of *Hepatozoon* were downloaded from GenBank and aligned to the sequences generated in this study. *Adelina dimidiata* Schneider, 1875, *Haemogregarina sacaliae* Dvořáková, Kvičerová, Hostovský, and Široký, 2015, *Dactylosoma ranarum* Lankester, 1882, were selected as the outgroup for the 18S phylogenetic tree. *Klossia helicina* Schneider, 1875, *Klossia razorback* Zeldenrust and Barta, 2021, and *Klossiella equi* Baumann, 1946 were used as the outgroup for phylogenetic trees constructed with mitochondrial CDS genes. Genetic sequences were aligned separately with ClustalW in Geneious R9 with a 35 gap open cost and a 0.73 gap extend cost. The aligned sequences were then concatenated. Molecular markers were concatenated in various configurations as follows: 18S-COI, 18S-COIII, 18S-CytB, COI-COIII, COI-CytB, COIII-CytB, 18S-COI-COIII-CytB, and COI-COIII-CytB. A model test was performed on each alignment and concatenated alignment to determine the best substitution model according to the Bayesian information criterion (BIC), using jModelTest 2.1.10 (Darriba *et al.,* 2012). Phylogenetic trees were constructed with single genetic markers and concatenated alignments using the best model parameters determined by jModelTest. MrBayes v3.2.6 (Huelsenbeck & Ronquist, 2001) was used to construct an 18S phylogenetic tree with the General Time Reversal (GTR) model with gamma distribution, a portion of the sites considered invariable. The analysis was run with the Markov Chain Monte Carlo (MCMC) algorithm for 10 million generations, sampling every 100 generations. The first 25% of the trees were discarded as 'burn-in'. All protein-coding gene phylogenetic trees and concatenated phylogenetic trees were constructed through a two-step process. First a maximum parsimony tree was constructed of each alignment with a heuristic search strategy through PAUP* 4.0a169 (implemented from within Geneious R9) (Swofford and Sullivan, 2009). From the resulting maximum parsimony trees, Bayesian inferred (BI) analyses was run with MrBayes (implemented from within Geneious R9). BI analyses used a GTR substitution model, gamma distributed variation, 4 discrete gamma categories, a portion of sites considered invariable, 50 thousand generations were sampled every 100 generations, and 25% of trees discarded as 'burn-

in'. For a Bayesian inferred analysis it is mostly accepted that the placement of a taxon or clade is well-supported if the posterior probability is >90% (Netherlands *et al.,* 2020). Barta *et al.* (2012) considers a posterior probability of >70% to be well supported. In this study a posterior probability of >80% will be considered as well-supported.

Partition homogeneity (ILD) tests were performed on each concatenated alignment to determine the degree of congruency among genetic markers. PAUP* 4.0a169 was used for the analysis. The analysis criterion within PAUP* was set to 'Parsimony' and an alignment nexus file was imported. The gene partitions were set according to the lengths of each genetic marker in the alignment. Uninformative sites were excluded from the analyses. The ILD test was run for 5000 replicates with a heuristic starting seed of 365, and a random addition sequence at 10 replicates with a 3658 seed. One tree was held at each step. The process was repeated for each concatenated alignment file.

The patristic distance matrix for each phylogenetic tree was exported from Geneious R9 to be used in Excel. Within Excel the top-right half of each distance matrix was discarded since it's only a duplication of the bottom-left half. The data of each tree were then transposed into separate columns and sorted from the smallest to the largest branch lengths. A line graph was drawn from the data to visualise the resolution of each phylogenetic tree.

All images were edited and constructed in Adobe Illustrator CC v23.0.1 (Adobe Inc., 2019)

2.3.4. Results

2.3.4.1. Molecular data

Mitochondrial protein-coding genes (COI, COIII, and CytB) of (n=8) *Hepatozoon* spp. were sequenced and are listed in Table 4. Of the protein-coding genes sequenced, there are four partial COI sequences, five partial COIII sequences, and eight partial CytB sequences. Unit lengths for the COI sequences range between 1266 bp & 1388 bp with 36.4–39.0% GC content. COI and COIII data were obtained for isolate 16B1, 27A1, 07H2, and 07H6. Two partial COIII sequences were obtained for isolate 27A1. Partial CytB sequences were obtained for all *Hepatozoon* spp. analysed in this study. The unit lengths of the COIII sequences ranged from 533–1278 bp with an average GC content of 36.2%. The unit lengths of CytB are between 606 bp and 966 bp with an a 35.2–38.5% GC content. The BLAST results of all mitochondrial DNA (mtDNA) sequences obtained matched *Hepatozoon* spp. with 70.5–93.4% sequence similarity. 18S rDNA sequence fragments between 1200 and 1700 bp were sequenced from *Hepatozoon* spp. in isolates 07A1, 27A1, 16B1, 19F1, 18B1, and 20A1. The BLAST results of the 18S rDNA sequences are as follows: 07A1 shares 95.4% sequence similarity with *Hepatozoon cecilhoarei* (MG519504), 27A1 shares 99.0% sequence similarity with *Hepatozoon ophisauri* (MN723845), 16B1 shares 98.5% sequence similarity with *Hepatozoon ayorgbor* (EF157822), 19F1 shares 98.1% sequence similarity with *Hepatozoon cuorae* (MW514214), 18B1 share 99.2% sequence similarity with *Hepatozoon cuorae* (MW514213), and 20A1 shares 99.5% sequence similarity with *Hepatozoon cuorae* (MW514213). All 18S rDNA sequence BLAST results had a query cover between 98 and 100%. Isolates 07H2 and 07H6 have 18S data from a previous study available on GenBank under accession numbers MG041594 (*Hepatozoon involucrum*) and MG041603 (*Hepatozoon thori*) respectively.

Table 4: 34 Mitochondrial protein-coding gene sequences were generated from 8 *Hepatozoon* spp. and are listed. Isolate 27A1 has two COIII fragments generated and are listed within one cell, one atop the other.

Isolate	Parasite	Host	Fragment length (bp)			GC content %			BLAST results						
										% identity			% cover		
			CytB	COI	COIII	CytB	COI	COIII	Match	CytB	COI	COIII	CytB	COI	COIII
16B1	*Hepatozoon* sp.	*Dipsadoboa aulica*	892	1303	696	38.5	39	37.8	*H. griseisciuri* - MK452388	91.46	89.46	89.44	99	83	100
20A1	*Hepatozoon* sp.	*Kinixys zombensis*	611			35.7			*Hepatozoon* sp - MT936932	78.64			79		
27A1	*Hepatozoon* sp.	*Naja annulifera*	609	1271	533 540	38.3	38.1	35.3 36.7	*H. griseisciuri* - MK452388	93.43	92.03	89.33 90.49	100	99	94 99
18B1	*Hepatozoon* sp.	*Stigmochelys pardalis*	620			36.3			*Hepatozoon* sp - MT936932	78.50			79		
19F1	*Hepatozoon* sp.	*Naja mossambica*	606			38.4			*Hepatozoon* sp - MT936931	78.95			100		
07A1	*Hepatozoon* sp.	*Philothamnus semivariegatus*	611			35.5			*Hepatozoon* sp - MT936932	84.89			99		
07H2	*H. involucrum*	*Hyperolius marmoratus*	966	1266	1256	36.1	36.8	35.4	*H.* clamantae - MN245142	92.14	93.21	89.44	98	100	100
07H6	*H. thori*	*Hyperolius marmoratus*	967	1388	1278	35.2	36.4	35.6	*H.* clamantae - MN245142	93.08	93.44	88.96	98	100	95

Description: The table above lists the 34 mitochondrial protein-coding gene sequences generated from 8 different *Hepatozoon* spp. parasitising vertebrates from South Africa. The three protein-coding genes (COI, COIII and CytB) were only sequenced for 4 of the 8 *Hepatozoon* spp. where the remaining 4 *Hepatozoon* spp. only CytB sequences were obtained. The table includes the host from which the *Hepatozoon* spp. were isolated, length of each sequence fragment, and %GC content of each sequence obtained. The table also includes the BLAST results of each sequence fragment, which includes the species match, %identity, and %query cover.

2.3.4.2. Phylogenetic analyses

Using MrBayes set to 10 million generations took approximately 14-17 hours to produce a consensus tree, whereas constructing a maximum parsimony tree followed by a BI analysis with 50 000 generations produced a consensus tree within 10 minutes. This method produced phylogenetic trees equal to or better than the trees generated with 10million generations.

2.3.4.2.1. Genetic marker characteristics, incongruency test and patristic distances

The characteristics of each genetic marker are listed in Table 5. Of the four markers, COIII is the most informative, most variable, and least conserved, whereas the 18S rDNA sequences are the least informative, the least variable, and the most conserved.

Table 5: Phylogenetic analysis of genetic markers 18S, COI, COIII, and CytB.

Marker	Pi	V	C	Size (bp)
18S	14%	24%	73%	894
COI	39%	56%	43%	1212
CytB	47%	64%	35%	669
COIII	52%	75%	23%	740

Description: The table above provides additional information of the genetic markers analysed in this study. The portion of informative sites (Pi), portion of variable sites (V) and the portion of conserved sites (C) indicated as a percentage. The total unit length of each marker is indicated.

The incongruency of the genetic markers was tested with the partition homogeneity (ILD: Incongruence Length Distance) test in PAUP*4. The null hypothesis of the test is congruency among compared genetic markers, and the alternative hypothesis is incongruency among compared genetic markers (Farris *et al.,* 1994). P-values were calculated for each combination of genetic markers and are listed in Table 6. Two different conditions were measured. Condition A excluded *Hepatozoon canis* and *Klossiella equi* sequences from the analyses and subsequent trees tested. Condition B included *Hepatozoon canis* and *Klossiella equi* sequences in the analyses and subsequent trees. A P-value below 0.05 rejects the null hypothesis. All calculated P values are above 0.05 indicating congruency among all markers compared. Since P-values vary depending on the markers compared, a value closer to 1 indicates a higher degree of congruency. The 18S-COIII concatenation shows the highest degree of congruency, followed by the 18S-COI concatenation. The 18S-CytB concatenation and the COI-COIII concatenation indicate the lowest degree of congruency. The removal of the *Hepatozoon canis* and *Klossiella equi* sequences from the data matrices reduced the degree of congruency for most markers by ~0.04 up to ~0.76. The COI-COIII concatenation had minimal change between conditions A and B.

Table 6: Results of the Partition homogeneity (ILD) test for both conditions A and B of each genetic markers compared.

	A	B
Markers tested	P value	P value
18S vs COIII	0.74	0.95
18S vs COI	0.44	0.91
COI vs CytB	0.38	0.9
18S vs COI vs COIII vs CytB	<0.1	0.86
COI vs COIII vs CytB	<0.1	0.76
COIII vs CytB	0.24	0.75
COI vs COIII	0.42	0.46
18S vs CytB	0.42	0.33

Description: The P-values for the incongruency test of genes is displayed above. Column A includes the P-values of trees excluding *Klossiella* and *Hepatozoon canis* taxa. Column B includes the P-values of trees including *Klossiella* and *Hepatozoon canis* taxa. A P-value closer to 1 indicates a higher degree of congruency whereas a P-value closer to 0 indicates the opposite.

To visualise the resolution of the trees constructed with the genetic markers, a patristic distance graph of each tree was drawn. The different graphs are illustrated in Figure 2.3.1. The patristic distance is the sum of the branch lengths that connect two nodes on a phylogenetic tree. Only the patristic distances of the phylogenetic trees constructed under conditions B are illustrated. Four graphs are displayed, labelled A, B, C, and D. Graph A is a combination of all the data sets analysed, B is the grouping of the 18S-CDS trees, C is the grouping of the CDS trees along with the 18S-COI-COIII-CytB tree, and D is the grouping of the concatenated CDS gene trees. The concatenated CDS genes show the highest patristic distances among the data sets. Trees constructed with a single CDS gene are second to the concatenated CDS trees. Concatenated 18S-CDS trees have the lowest patristic distances. Within graph B, 18S-COIII and 18S-COI have the highest patristic distances compared to the 18S-CytB and 18S trees. Of the four genetic markers, COIII produces trees with the highest patristic distances, followed by the COI, then the CytB and lastly 18S. The patristic distances of the COI and COIII trees are near similar in most cases. Accordingly, the data indicate that the COI and COIII genes have the highest resolutions of the four markers. The resolution in this case is the patristic distance. The concatenated trees resemble the resolution of the combined genetic markers used. Each tree graph has regions where the data 'jumps'. These jumps are a sudden increase of the y-axis value at certain points along the x-axis. These jumps occur at roughly the same position on the x-axis of each tree graph. The magnitude of the jump depends on the genetic markers. Trees with CDS genes display the largest jumps, and trees containing the COI and COIII genes have the largest jumps among these. Almost no jumps are observed in the 18S rRNA gene tree except for a very small jump (compared to the jumps of the CDS trees) of the y-axis near the end of the x-axis.

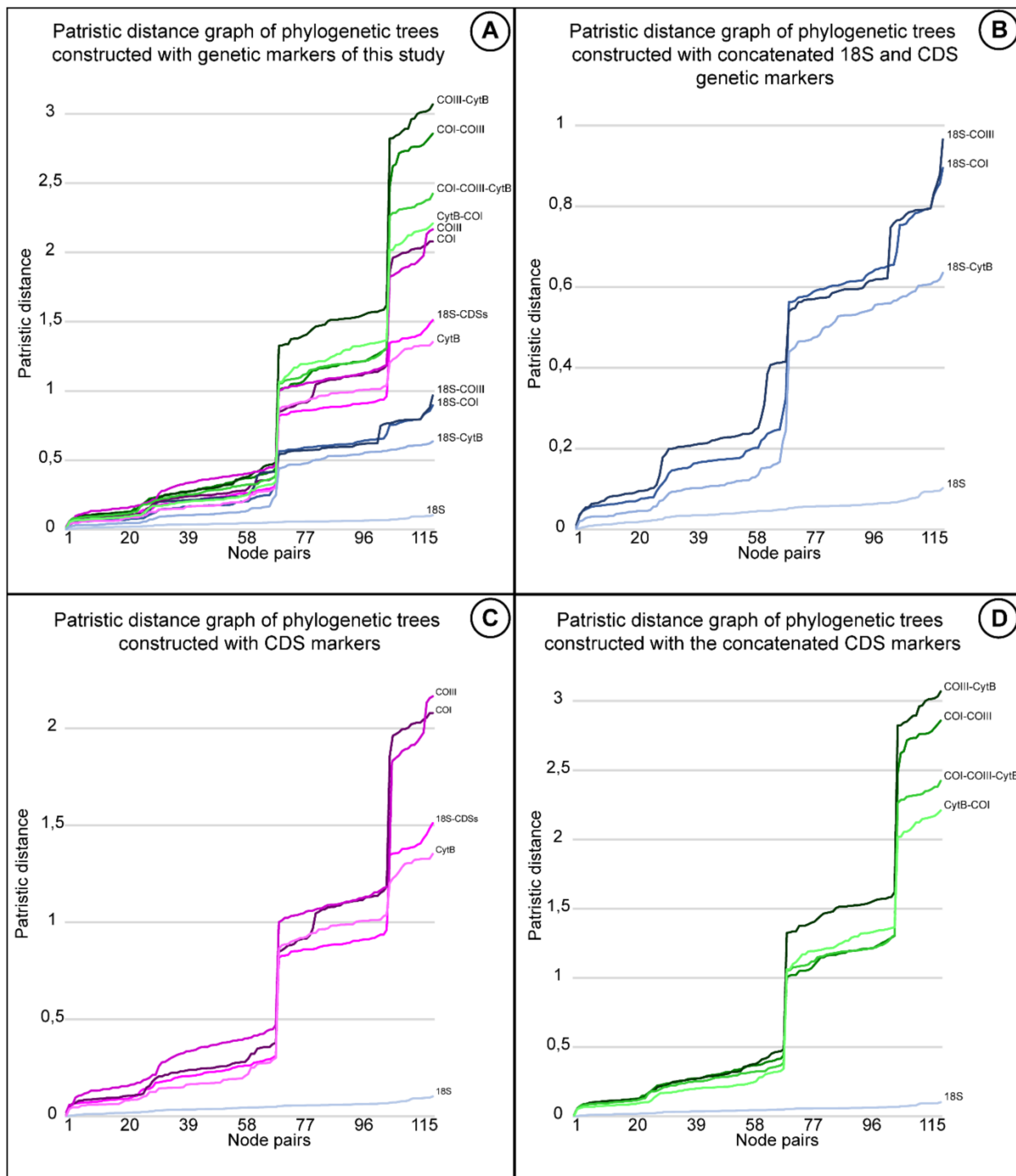


Figure 2.3.1: Patristic distance graphs. The patristic distance matrix of each phylogenetic tree was extracted and plotted on an xy-axis. The y-axis is the patristic distance between two nodes of a phylogenetic tree and the x-axis are the node pairs compared placed from smallest to largest patristic distances. A: The patristic distances of all 12 phylogenetic trees are plotted. B: The patristic distances of 18S-COI, 18S-COIII, and 18S-CytB trees are plotted with the 18S tree. C: The patristic distances of the COI, COIII, and CytB trees are plotted with the 18S-COI-COIII-CytB and 18S trees. D: The patristic distances of each concatenated CDS tree is plotted with the 18S tree. Among all the trees, the COI and COIII sequences produce trees with the highest patristic distances compared to the CytB.

2.3.4.2.2. 18S rDNA tree

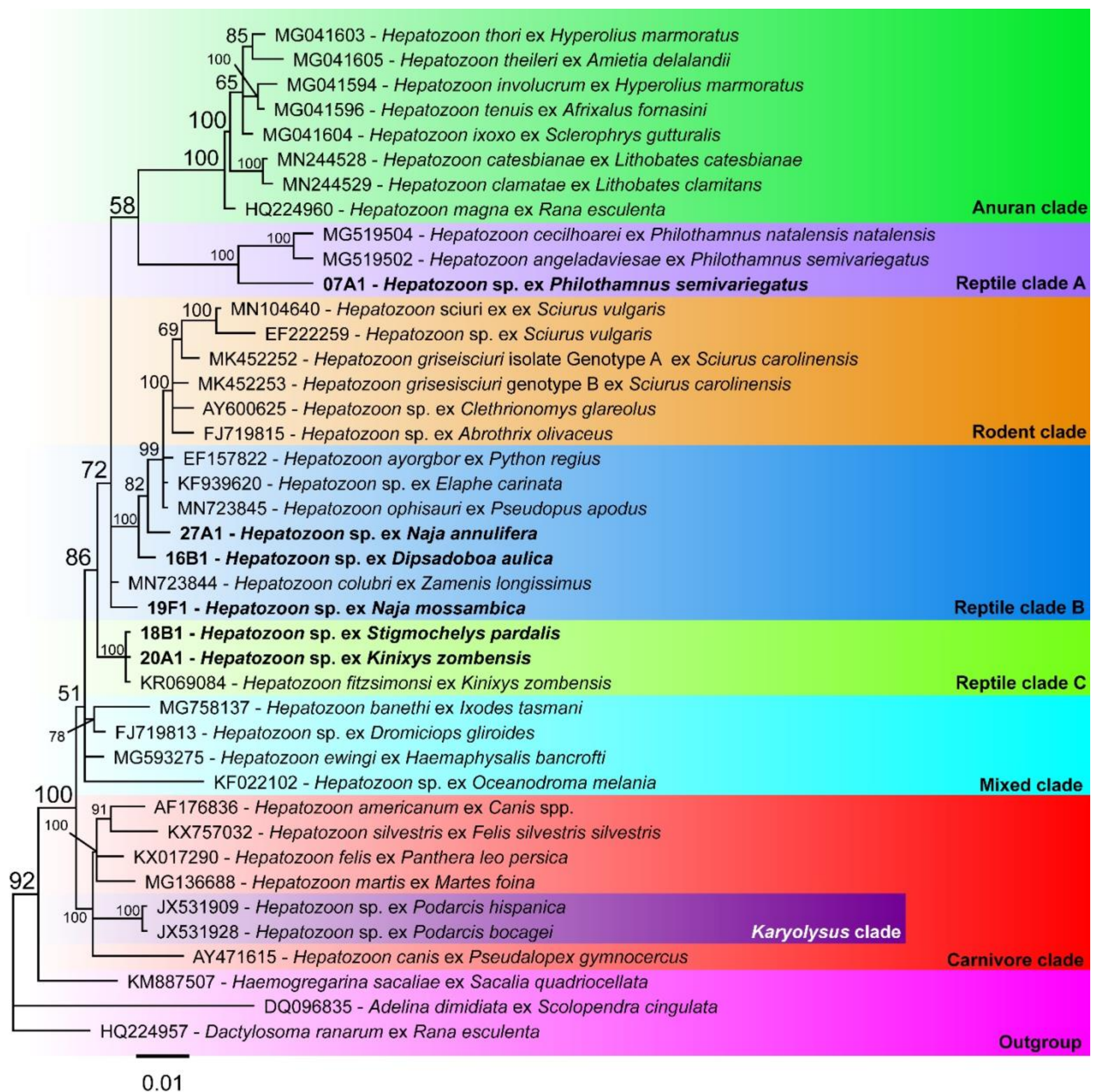


Figure 2.3.2: The 18S rDNA phylogenetic tree includes species of *Hepatozoon* available on GenBank. Most taxa used to construct the tree are the same as those used by Netherlands *et al.* (2020). The placement of taxa and posterior probabilities are similar to that of Netherlands *et al* (2020), however with some differences. The sequences generated in this study are added to the phylogenetic analyses and are highlighted as BOLD text. Isolates 07A1, 27A1, 16B1, 19F1, 18B1, and 20A1 correctly cluster with other *Hepatozoon* spp. parasitizing reptile hosts.

The sequences used to construct the phylogenetic trees are listed in Supp. Table 3. For the 18S rDNA tree, 6/41 candidates were sequenced in this study and the rest were downloaded from GenBank. From the 18S rDNA phylogenetic analysis, *Hepatozoon* species diverge into seven clades (Figure 2.3.2). The *Hepatozoon* clades are as follows: the anuran clade representing *Hepatozoon* spp. parasitising anuran hosts; reptile clades A, B, and C represent *Hepatozoon* spp.

parasitising snakes, turtles, and lizards; the rodent clade represents *Hepatozoon* spp. parasitising rodent hosts; the mixed clade represents *Hepatozoon* spp. isolated from ticks and a bird; and the carnivore clade represents *Hepatozoon* spp. parasitising mammalian carnivore hosts. According to Netherlands *et al*. (2020), the two *Hepatozoon* spp. parasitising *Podarcis* spp. are placed with *Karyolysus* spp. in a separate clade from carnivore *Hepatozoon* spp. The positions of these two *Hepatozoon* spp. are indicated as the *Karyolysus* clade within the carnivore clade. Reptile clade A is sister to the anuran clade, although not well-supported, and separates the anuran and rodent clade. The rodent clade shares a common ancestor with reptile clade B, although the monophyly of the rodent clade is well-supported. Reptile clade C forms a well-supported monophyletic sister clade to reptile clade B. The placement of mixed clades between the carnivore clade and the reptile clade C is not well-supported. The monophyly of the carnivore clade is well-supported and is a sister to the outgroup and the mixed clade. The 18S rDNA sequences generated in this study form well-supported associations with other *Hepatozoon* spp. parasitising reptile hosts. Isolate 07A1 parasitising *Philothamnus semivariegatus,* clade with other *Hepatozoon* spp. parasitising *Philothamnus* hosts. Isolates 27A1, 16B1, and 19F1 fall within reptile clade B closely related to other *Hepatozoon* spp. parasitising snake hosts along with a *Hepatozoon* spp. parasitising a legless European lizard. Isolates 18B1 and 20A1 parasitizing turtle hosts form a well-supported monophyletic clade with *Hepatozoon fitzsimonsi* as reptile clade C.

2.3.4.2.3. Concatenated 18S-CDS trees

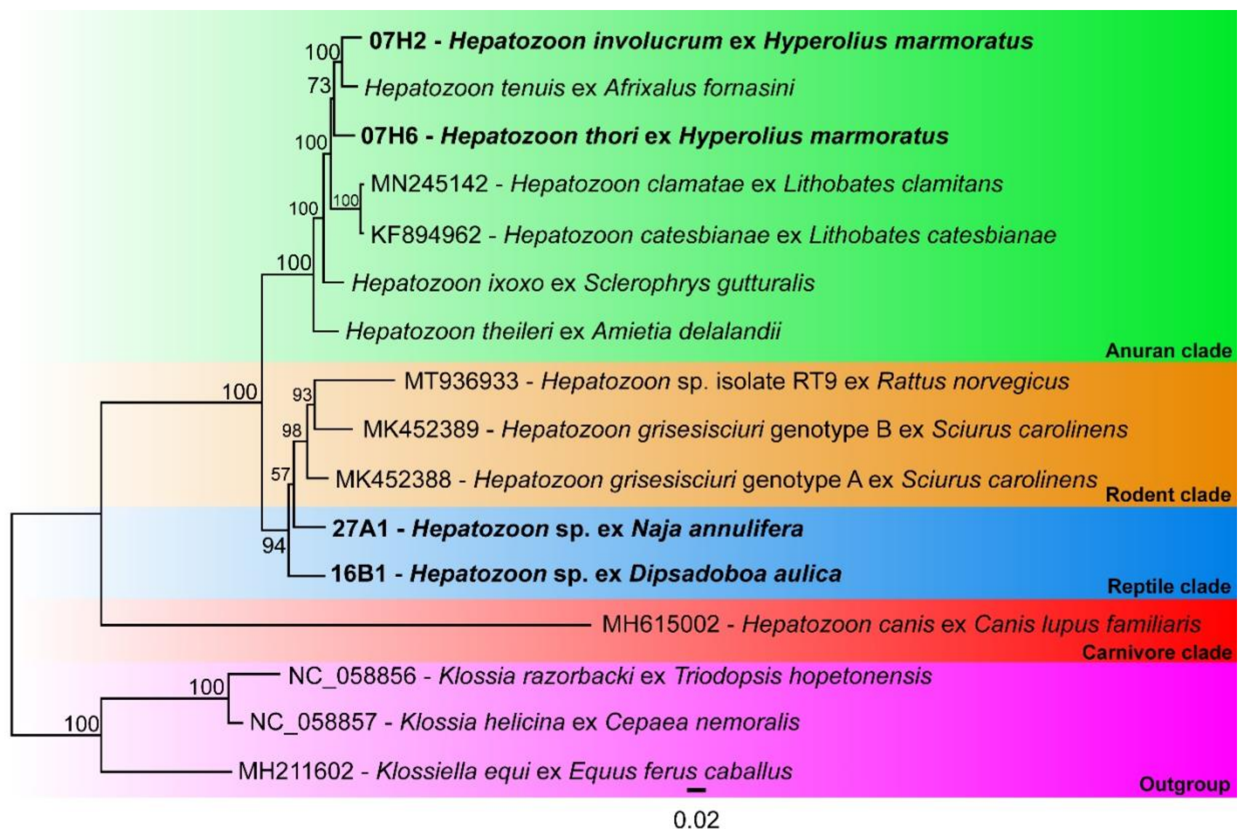


Figure 2.3.3: The 18S COI phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together alongside *Hepatozoon* spp. parasitising rodent hosts. The marker combination places taxa in the correct order according to the 18S tree.

The 18S rDNA sequence was concatenated with each of the CDS genes separately. 18S-COI is illustrated in Figure 2.3.3, 18S-COIII in Figure 2.3.4, and 18S-CytB in Figre 2.3.5. The 18S-COI and 18S-COIII trees split *Hepatozoon* spp. into four clades. The four clades formed in both trees are the anuran clade, rodent clade, reptile clade, and carnivore clade. The 18S-CytB has more taxa represented in the tree and split the *Hepatozoon* spp. into seven clades. The seven clades consist of an anuran clade, three reptile clades (A, B, C, and D), a rodent clade, and a carnivore clade. Reptile clade B of the 18S-CytB tree is a synonym of the reptile clade in the 18S-COI and 18S-COIII trees. The monophyly of the anuran clade is well-supported in all three trees and contains isolates 07H2 and 07H6. The trees suggest that *H. involucrum* is closely related to *H.* tenuis. Only the 18S-CytB tree suggests that *H. thori* and *H. theileri* share a common ancestor. All three trees agree that *H. catesbianae* and *H. clamatae* share a common ancestor. In the 18S-CytB tree, reptile clade A is well-supported containing one taxon, isolate 07A1 parasitising *Philothamnus semivariegatus*, and is sister to the anuran clade, separating the rodent clade from the anuran clade. The 18S-COI and 18S-CytB trees both indicate that the anuran clade is

evolutionarily closer to the rodent clade than to the reptile clade (Reptile clade B in the 18S-CytB tree). The 18S-COIII tree contradicts this by placing the reptile clade sister to the anuran clade. In all three trees, the rodent and reptile clades are subclades of a larger monophyletic clade and share a common ancestor. The reptile clade in all three trees contain isolates 27A1 and 16B1, however, the monophyly is not resolved. In the 18S-CytB tree, the position of the rodent-reptile clade B group is well-supported between reptile clade A and reptile clade C. The reptile clade C contains isolate 19F1 and its position is well-supported. The reptile clade D is sister to the reptile clade C and contains isolates 18B1 and 20A1. The carnivore clade is wedged between the reptile clade D and the outgroup.

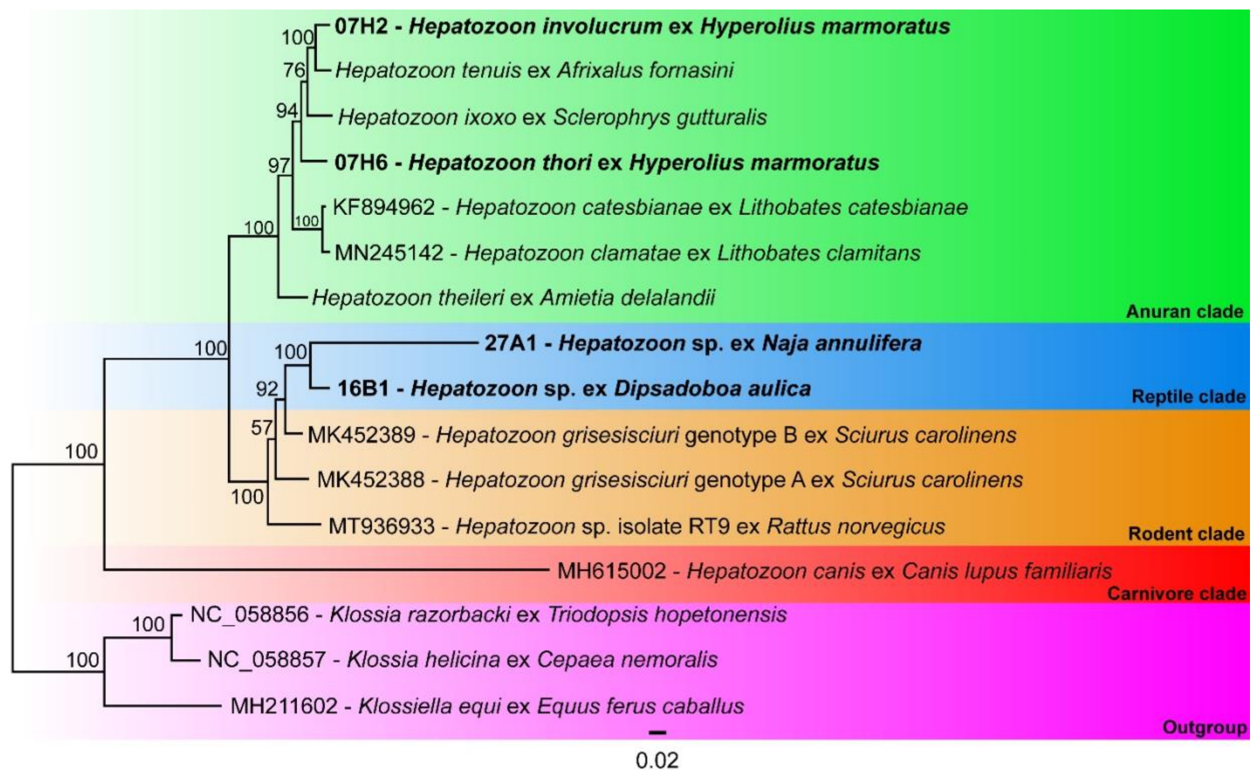


Figure 2.3.4: The 18S COIII phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together alongside *Hepatozoon* spp. parasitising rodent hosts. The marker combination does not place taxa in the correct order according to the 18S tree. Isolates 27A1 and 16B1 are placed close to the anuran clade instead of close to the carnivore clade.

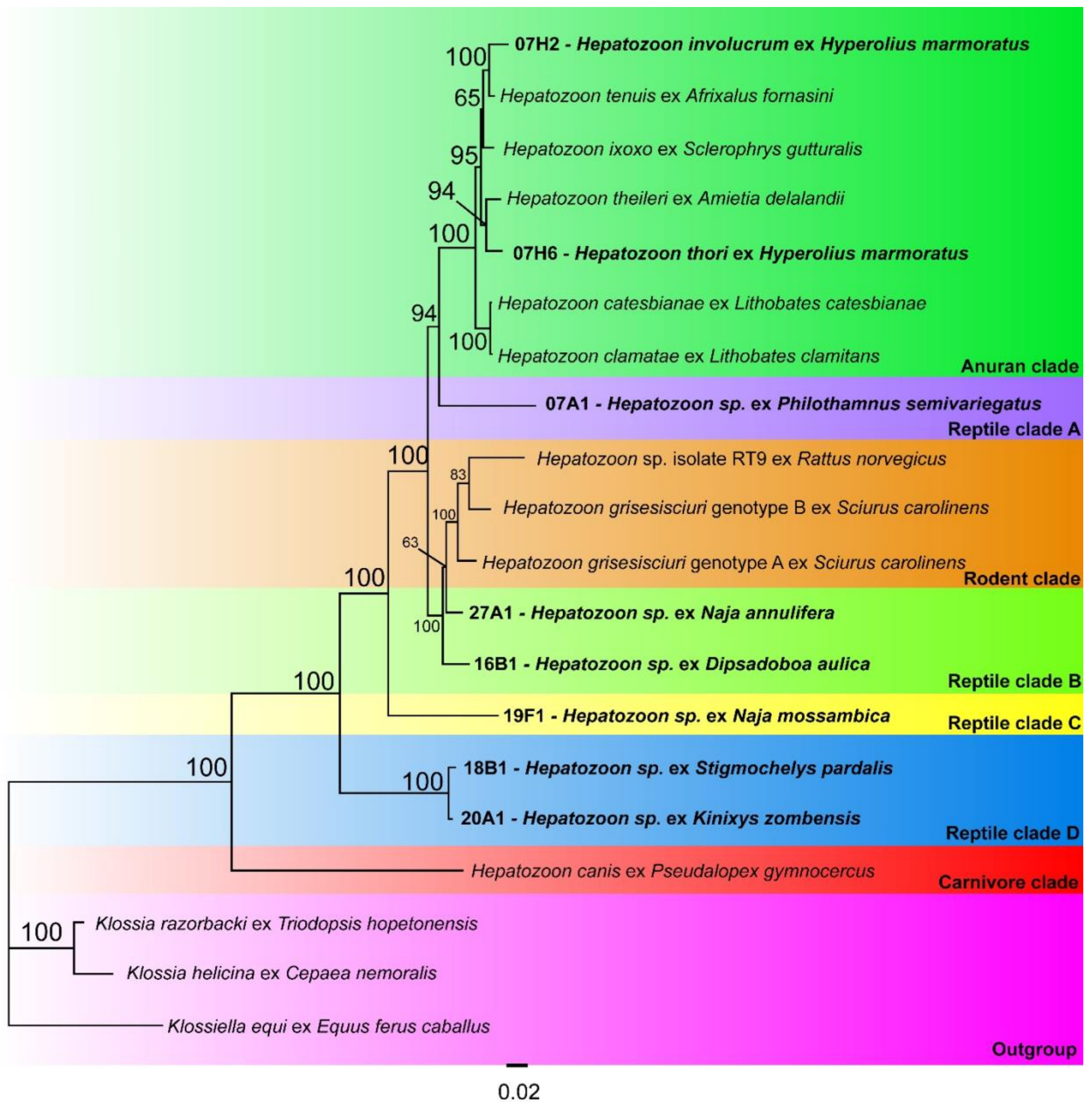


Figure 2.3.5: The 18S CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolate 07A1 is placed alone alongside the anuran clade. Isolates 27A1 and 16B1 clade together alongside *Hepatozoon* spp. parasitising rodent hosts. Isolate 19F1 is placed on its own close to the rodent-reptile clade. Isolates 18B1 and 20A1 are placed within their own clade next to the carnivore clade. The marker combination places taxa in the correct order according to the 18S tree.

2.3.4.2.4. Concatenated 18S-COI-COIII-CytB tree:

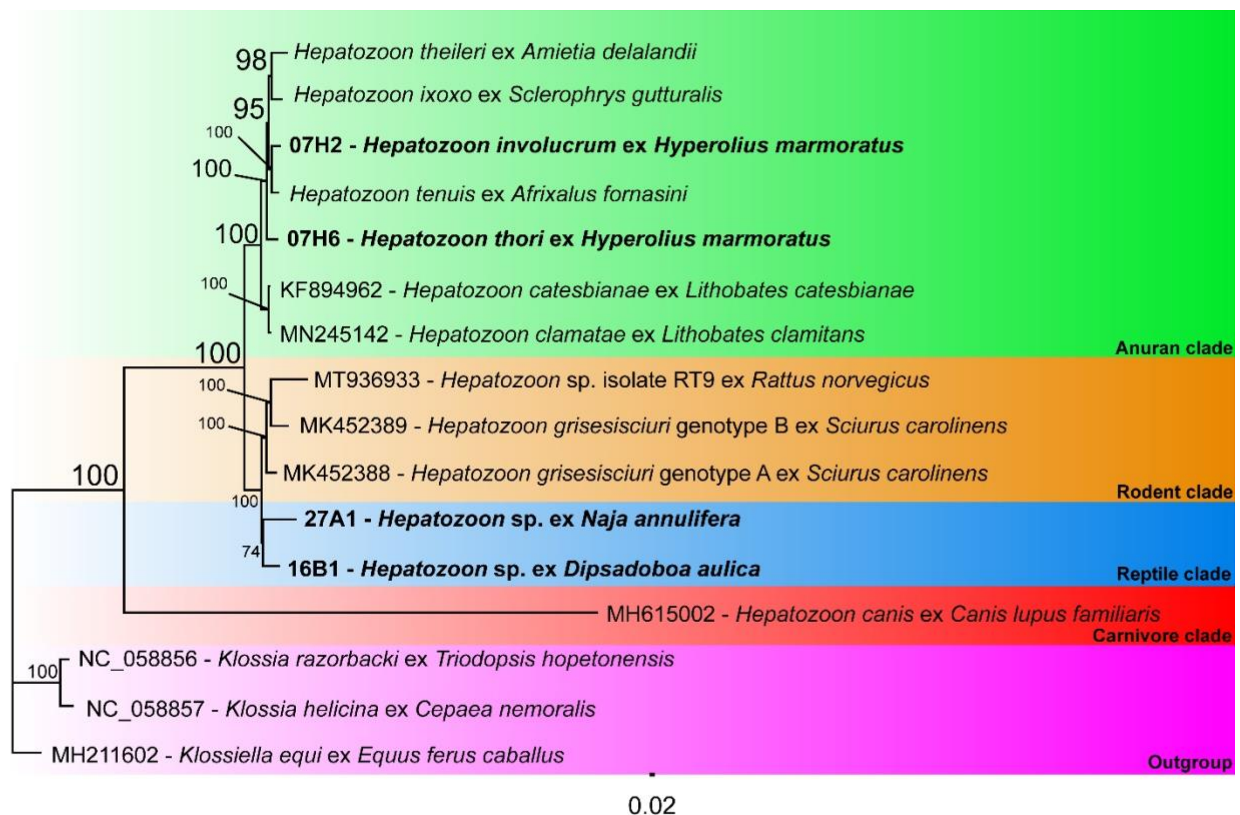


Figure 2.3.6: The 18S-COI-COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade alongside *Hepatozoon* spp. parasitising rodent hosts. The marker combination places taxa in the correct order according to the 18S tree. The marker combination is able to resolve the monophyly of nodes deep within a phylogenetic tree of *Hepatozoon* spp.

The 18S rDNA sequence was concatenated with all protein-coding genes from *Hepatozoon* spp. in this study (Figure 2.3.6). In this tree *Hepatozoon* spp. diverge into four clades. The four clades formed are the anuran, rodent, reptile, and carnivore clade. The anuran clade is well-supported, containing isolates 07H6 and 07H2. The reptile and rodent clade are well-supported subclades of a larger well-supported clade. Unlike the 18S-COI, 18S-COIII, and 18S-CytB trees, the monophyly of the reptile clade is resolved. The reptile clade shares a common ancestor with the rodent clade. The rodent clade is more closely related to the anuran clade, and the reptile clade is more closely related to the carnivore clade. The monophyly of the carnivore clade as sister to the outgroup is well-supported. The subsequent addition of all CDS genes to the 18S rDNA sequence compared to the addition of a single CDS gene produced a tree with higher resolution and increased posterior probabilities. The patristic distance graphs corroborate this observation.

2.3.4.2.5. CDS trees

The CDS sequences of *Hepatozoon* spp. were used for phylogenetic analysis and diverge into four clades. The four clades are the anuran clade, rodent clade, reptile clade, and carnivore clade. The topology of the COI (Supp. Figure 1), COIII (Supp. Figure 2), and CytB (Supp. Figure 3) trees differ, as well as their ability to delineate taxa. Of the three genes, the COI shows the most resemblance to the 18S tree. All three genes place isolates 07H2 and 07H6 within the anuran clade with 07H2 alongside *H. tenuis*. Isolates 16B1 and 27A1 are placed alongside the rodent clade in all three trees, and alongside other reptile taxa in the CytB tree. Isolates 18B1, 19F1, and 16B1 are also placed within the reptile clade with the CytB gene.

With the COI gene, the anuran clade is monophyletic and well-supported. The rodent and reptile clades are subclades of a larger clade and share a common ancestor. The posterior probability of the subclades is not well-supported. The division of the rodent-reptile clade from the anuran clade is well-supported. The monophyly of the carnivore clade is well-supported and genetically the most distant taxa in the tree. The anuran clade forms the youngest branches of the tree. The rodent clade is more closely related to the anuran clade, while the reptile clade is more closely related to the carnivore clade.

With the COIII, the rodent and reptile taxa form a well-supported clade. The monophyly of the anuran clade is well-supported and is the youngest of the clades. The reptile clade is a well-supported subclade of the larger rodent-reptile clade. The carnivore clade is sister to the outgroup. Unlike the COI, the reptile clade is sister to the anuran clade. Also, *H. catesbianae* and *H. clamatae* are the youngest of the anuran *Hepatozoon* spp. in this tree compared to the COI and CytB trees where they are some of the oldest.

With the CytB gene the reptile clade is divided into reptile clades A and B. Reptile clade A is a well-supported monophyletic clade that contains isolates 18B1, 20A1, 19F1 and 07A1. Isolates 16B1 and 27A1 form a paraphyletic clade as reptile clade B. The monophyly of the rodent clade is not resolved and is paraphyletic. The anuran clade is separated from the rodent clade by the reptile clade. The anuran clade is also monophyletic and the youngest in the tree. The carnivore clade is also monophyletic and sister to the outgroup.

2.3.4.2.6. Concatenated CDS trees

The three protein-coding genes were concatenated into four possible concatenation combinations and phylogenetic trees were constructed of each: COI-COIII (Supp. Figure 4), COIIII-CytB (Supp. Figure 5), COI-CytB (Supp. Figure 6), and COI-COIII-CytB (Supp. Figure 7).

Concatenated trees COI-COIII, COIII-CytB, and COI-CytB split the *Hepatozoon* spp. into four clades, the anuran clade, reptile clade, rodent clade, and carnivore clade. In all three trees the reptile clade contains isolates 27A1 and 16B1 and is monophyletic but is only well supported

in the COI-COIII tree. The anuran clade is monophyletic and well-supported in all three trees except in the COI-CytB tree. In all three trees isolates 07H6 and 07H2 clade with other anuran *Hepatozoon* with 07H2 and *H. tenuis* placed together. The COI-COIII and COI-CytB tree place *H. theileri* and *H. ixoxo* together, however the support for this placement is not strong in the COI-COIII tree. The rodent clade is monophyletic and well supported in the COI-COIII tree but is paraphyletic in the other two trees.

COI, COIII, and CytB sequences were aligned and concatenated to construct a phylogenetic tree (Supp. Figure 7). *Hepatozoon* spp. diverge into four clades, the anuran, rodent, reptile and carnivore clades. The topology and posterior probabilities of the tree is identical to that of the 18S-COI-COIII-CytB tree.

2.3.4.3. Primers designed

A total of 22 primers were designed in this study, of which 8 were not fully tested. These primers are listed in Table 7 along with additional information such as primer sequence, quality of sequences produced, GC content, degeneracy, size, annealing temperature, and hairpin loop forming temperature.

Table 7: Primers designed.

Quality	Primer name	Primer sequence (5' → 3')	GC%	De	Size (bp)	Annealing temp	Hairpin Tm
Good	Api_COIII_351_R	AAYTCWACRAAATGCCAGTA	33.3	8	20	48.9 – 53.9	None
	SSUF_Haem_F	TGACRGTGAACTTGTGGCTG	52.6	2	20	57 – 59.4	None
	Hep_COI_329_F	ATGTTCCTTAGGCTCWGC	50	2	18	52.5	None
	Hep_COI_329_R	GCWGAGCCTAAGGAACAT					
	Hep_COI_1733 F	TCCATTTAAMTTCCACTTCTC	35	2	21	50-51.8	None
	Hep_COI_1733 R	GAGAAGTGGAAMTTAAATGGA					
Moderate	Hep_COIII_273_F	GTCTAGCTATTAATCARGTYAAAGCTCATTTACAA	30.3	4	35	61.1 – 63.6	41
	Hep_COIII_273_R	TTGTAAATGAGCTTTRACYTGATTAATAGCTAGAC					
	LSUB_F	TGTGCYTGGDAGCTGTAATC	50	6	20	54.1 – 58.3	50.7 – 53.7
	LSUB_R	GATTACAGCTHCCARGCACA					
Bad	LSUD_F	TGKCATYTGCCCGG	57	4	14	47.8 – 53.5	59.3
	LSUD_R	CCGGCARATGMCA					
	Hep_COIII_971_F	TACTGGCATTTRGTWGAATTAGTKTGGC	38.5	8	28	60 – 64.3	None
	Hep_COIII_971_R	GCCAMACTAATTCWACYAAATGCCAGTA					
Not tested	SSUD_F	TGGAAGTAGAAATCGTTASTAA	31.8	2	22	51	None
	SSUD_R	TTASTAACGATTTCTACTTCCA					
	LSUA2_F	GACCGDACCTTGGHAT	57.1	9	16	47.2 – 54.4	39.6
	LSUA2_R	ATDCCAAGGTHCGGTC					
	Api_LSUD_F	CAACACAAGACWCTGATAACTTTGA	36	2	25	56.1 – 56.6	None
	Api_LSUD_R	TCAAAGTTATCAGWGTCTTGTGTTG					None
	LSUA1_F	CYAKAAAGCATGTGATCTAATT	30	4	22	48.2 – 53.1	31.3
	LSUA1_R	AATTAGATCACATGCTTTMTRG					

Description: The table lists the primers designed in this study to amplify and sequence the protein-coding genes of *Hepatozoon* spp. Some primers were tested, and some were not. Those primers that were tested yielded either good, moderate, or bad quality sequence fragments. The name of each primer is listed along with its DNA sequence written from 5' to 3'. The optimal PCR conditions of each primer pair is listed along with additional information regarding hairpin Tm. De = The degeneracy value of each primer.

2.3.5. Discussion

2.3.5.1. A peculiar COIII

The COIII of isolate 27A1 was sequenced, however, one large and one small fragment were obtained instead of one complete fragment. The PCR product was obtained through a nested PCR approach. Each fragment had two smaller fragments, which were sequenced DNA from the primers. Interestingly, the smaller fragment (~211bp) made a scaffolding nearly identical to that of the larger fragment (~950bp). When the PCR product was run through gel electrophoresis, two bands were observed, one of about 200 bp and the other of about 1000 bp in size, which corresponds with the sequenced data. The two large fragments are expected to align with minimal differences. A de novo assembly reveals that both fragments have a 124 bp region at the 5' end with 99.2% sequence similarity. A pairwise alignment with a cost matrix of 93% similarity reveals that the fragments have 76.7% sequence similarity across 489 bps of both fragments. The two smaller fragments were DeNovo assembled and showed 95.1% sequence similarity across 150bps of both fragments. When the large and small fragments were DeNovo assembled, the smaller fragments aligned with 97.9% sequence similarity to the larger fragments near the middle of the assemblage. The DeNovo assemblage of both fragments did not make much sense since four repeat regions formed in the consensus of the alignment (~950bp). Among the mitochondrial genomes of adeleorinids, it is uncommon for the COIII gene to be larger than 800bp, with most being around 720bp. In this case, it is unlikely that the DeNovo assembled sequence is a true reflection of the COIII gene in isolate 27A1. Instead, it seems that both large fragments need to be treated separately and considered to have originated from two different organisms.

2.3.5.2. Primers

The primers used to amplify and sequence the COIII fragments from isolate 27A1 were designed in this study. Additional primers were also used and designed for other CDS genes in the mitogenome of *Hepatozoon* spp. Various primers are available for use in apicomplexan mitochondrial genomes (Leivelle *et al.,* 2019c). However, when sequencing the mitochondrial genomes of apicomplexans or their protein-encoding genes more is always better. Not every apicomplexan mitochondrial genome is the same in sequence or organisation and depends highly on the phylogeny of that parasite. For example, when the mitochondrial genomes of two *Haemogregarina* spp. where sequenced the limited primers significantly delayed the PCR and sequencing process (see Chapter 2.2). This was due to the variability of the mitochondrial genome organisation between genera forcing the use of primer pairs that produce large DNA fragments instead of pairs that would produce smaller, more specific fragments. In most cases, primers for smaller fragments were not available or limited to the CytB gene and a small region of the COI gene. No universal primers for apicomplexan COIII were available. We therefore

designed primers to target both ends of the COI gene and the COIII gene of *Hepatozoon* spp. Furthermore, we designed additional primers to target small and large subunit RNA fragments that are conserved among apicomplexans.

2.3.5.3. *Phylogenetic analysis:*

Several tests were performed to determine the usefulness of each CDS gene compared to the 18S rRNA gene in phylogenetic analyses, as well as how best to use each gene. The first test analysed how informative, variable, and conserved each gene is compared to the 18S rDNA sequence. The 18S rRNA gene has been found to be too conserved and slowly evolving, resulting in low-resolution phylogenetic trees (Ogedengbe *et al.,* 2011; Barta *et al.,* 2012). The phylogenetic analyses reported here support this assertion. The 18S rRNA gene carries 25-38% less informative sites than the CDS genes in this study. Also, the 18S is 30-50% more conserved than the CDS genes and 11-51% less variable. These findings are consistent with literature (Rubinoff and Holland 2005). Of the three CDS genes, the COIII is the most informative, most variable, and least conserved. This came at a surprise, since it was thought that the high variability would interfere with the phylogenetic weight of the gene. Consequently, it was hypothesised that the high degree of informative sites could be a coincidence and merely a 'false positive' due to the high variability of the gene. The Bayesian inferred phylogenetic trees of the COIII gene further fuelled these suspicions, since the topology of the tree was scrambled and did not resemble any of the other trees constructed at the time. A partition homogeneity test or an incongruency length distance test was performed on each of the genetic markers to test congruency. The idea was that the ILD test would reveal if the informative sites of the CDS genes correspond to the informative sites of the 18S rRNA gene. Through this a numerical value could be applied to gene pairs which would lay a foundation for further observations. The p-values revealed that the COIII is the most congruent gene to the 18S rDNA with a p-value of 0.98 (take note of the null hypothesis of the test in the results section, p-values are inversed). The ILD test revealed that all genetic markers are congruent with each other to varying degrees. Interestingly, the CytB gene has the least congruency with the 18S rRNA gene despite having the second highest portion of informative sites. It could be that the informative sites of the CytB is a false representation of its actual usefulness, however, this is only an assumption.

The genetic markers of this study were subject to two different conditions during the ILD tests. In one condition the DNA sequences for *H. canis* and *Klossiella equi* was removed from the analysis. The purpose was to evaluate the effect less genera represented in the test would have on the results. The idea comes from the mtDNA being useful for phylogenetic studies within a genus, but not across multiple genera (Caterino *et al.,* 2001; Holland *et al.,* 2004; Reed and Sperling, 1999; Rubinoff and Sperling, 2002). Removing *H. canis* and *Klossiella equi* from the

analyses reduced the p-values significantly for some trees or had no change in others. The resulting trees constructed under this condition had no remarkable differences when compared to the trees under condition B. These observations may indictate that species of *Hepatozoon, Klossia,* and *Klossiella* are related enough that there won't be much of a difference if they are included into a phylogenetic analysis of *Hepatozoon* with mtDNA.

Phylogenetic trees with genetic markers were constructed in various concatenated configurations. The focus of this study is the combined use of CDS genes with the 18S rRNA gene. However, other configurations were also tested, including separate trees drawn with single CDS genes and concatenated CDS genes. The 18S-CDS trees drawn told similar stories: 1) the anuran clade is well-supported and monophyletic; 2) *Hepatozoon* spp. from reptiles and rodents form a well-supported monophyletic clade; 3) the monophyly of the subclades in the reptile-rodent clade is unresolved for either the rodent clade or reptile clade depending on the CDS gene used. The only notable difference among the trees is the 18S-COIII tree placing the reptile clade as a sister to the anuran clade. As mentioned before, the 18S-COIII tree has the highest congruence, which suggests that its topology is the most accurate representation of the actual topology of the clades. It would seem, at least in this case, that it is not desirable to have a genetic marker that tells the same story as the 18S rRNA gene. Despite the additional information given by the CDS genes, the monophyly of the reptile clade is unresolved within the larger rodent-reptile clade. The most desirable outcome would be to have genetic markers that resolve a clade as deep into a tree as in this case. Surprisingly, out of the three CDS genes the COI on its own is the only sequence able to resolve the monophyly of the reptile clade; however, the support is low. Why this is not observed in the 18S-COI tree could be that the information from the 18S rDNA overwhelms that from the COI. The posterior probability at the node which splits the rodent clade from isolate 27A1 is low, which may be evidence for this assumption. The 18S-COI-COIII-CytB tree resolves the monophyly of the reptile clade and is well-supported. The COI-CytB does not resolve the monophyly of the reptile clade, which means that CytB does not contribute to the data. The COI-COIII does resolve the monophyly of the reptile clade and increases the posterior probability of the branching. This means that the COI splits the reptile-rodent clade into monophyletic subclades, and the COIII strengthens the probability of the split. The addition of the CytB gene or the 18S rDNA gene to COI and COIII reduces the posterior probability of some nodes. Using all four genes together in a tree increases the posterior probabilities. All of this supports the literature that suggests that each mitochondrial gene plays a different role in the story being told (Kowalczyk *et al.,* 2021). The 18S rRNA gene is already known to be good at sorting larger groups or clades of multiple genera. It could then be that COI filters closely related taxa into their appropriate clades, while the COIII and CytB genes cement the placements of taxa within these clades. Additional evidence of these roles can be seen with the anuran clade. The combination of genetic markers can distinguish between anuran *Hepatozoon* from South Africa

and those from Canada, with higher posterior probabilities than the 18S rRNA gene tree. In the case of this study, the use of the 18S rRNA gene is irrelevant since COI-COIII-CytB generates the same tree. The use of 18S rDNA would become increasingly more important as the different genera are added to the analysis.

2.3.6. Conclusion

This study provides new data that will supplement the research conducted around the phylogeny of haemogregarines, but more specifically *Hepatozoon.* Most articles surrounding the phylogeny of *Hepatozoon,* and the different solutions proposed to resolve it have come to the same conclusion: more data is needed. Mitochondrial protein-coding genes, especially COI, have been proposed as alternative or supplementary genetic markers that would perhaps resolve nodes within *Hepatozoon* that the 18S rRNA gene cannot. In the current study, it was discovered that all three protein-coding genes are sufficient to be used as supplementary genetic material with the traditional 18S rRNA genes for phylogenetic analyses. It was discovered that each gene carries different information on the placement of taxa in a phylogenetic tree. Although all three genes alone with the 18S can resolve the monophyly of larger groups within *Hepatozoon,* it is only when the three genes are used together with the 18S that deep phylogeny is resolved. In other words, no one gene tells the full story of the *Hepatozoon*, but it seems that together they create a more complete picture. This conclusion is supported by the opinions of Rubinoff and Holland (2005). This is that the use of mtDNA as lone standing barcodes are not desired, but that the combined use of these genetic markers with other genetic markers such as those from the nuclear genome are argued for. Due to this, it would be advantageous to sequence all three protein-coding genes of *Hepatozoon* spp. For these types of analyses. Primers that target the CytB CDS work well and produce high-quality fragments. Of the three CDSs, CytB was the easiest to amplify and sequence. Primers designed for the COI CDS were effective in aiding amplification, but to varying degrees of success. The primers only bound in *Hepatozoon* closely related to *Hepatozoon* from rodent hosts and for species of *Hepatozoon* from anuran hosts. The COI genes of *Hepatozoon* from turtles and two snakes were not amplified or sequenced. The reason for this could be that the primers are not degenerate enough to account for the change between groups of *Hepatozoon*. It was difficult to design primers for the COIII gene due to its high variability and low GC content. One primer designed for this region worked well, unfortunately it's the only one that could be designed. The design of primers is time consuming and complex, especially when not much is known about the region for which primers are being designed. The data gathered from this study will hopefully aid future studies design more robust primers for these regions. Although more data is needed to fully realise the divergence of taxa within *Hepatozoon*, the data gathered from this study along with those already available have painted a clearer picture. The mitochondrial protein-coding genes of *Hepatozoon* spp. used together with the common 18S rRNA gene are the key to determining the genetic divergence of *Hepatozoon*. That said, it seems that species of *Hepatozoon* are more likely to diverge into genera for every major vertebrate class parasitised by these parasites.

2.3.6.1. Challenges/Limitations

Here follows comments and observations made during the design process of primer pairs:

2.3.6.1.1. Primers COIII_351_R, COIII_273_F & _R, COIII_971_F & _R

Of the three mitochondrial protein-coding genes, designing primers for the COIII gene proved to be most challenging. The gene presents high variability among *Hepatozoon* spp. and has a high percentage of AT content. These two factors led to various primer design problems, with primers being too long (>20bp), too low in GC content (~<40%), highly degenerate, or having high hairpin Tm's (Bustin and Huggett, 2017). Of the five primers designed for the COIII protein-coding gene, COIII_351_R was most effective. The primer is located 351 bps from the assumed start of the gene. Although the primer only has a 33.3% GC content and a degeneracy of 8, the primer is 20bp long and has a comfortable annealing temperature of 48.9 - 53.9℃ with no hairpin Tm. These factors might have aided in the success of this primer above the other four, not only with its efficiency but also with the quality of sequenced fragments. It is clear why the COIII_971 primers performed worse than the COIII_273 primers, since in most cases the primer does not align with COIII gene sequences of available adeleorinid mitochondrial genomes unless a high mismatch parameter is set. This can only mean that the primer is poorly designed to target a region that is not sufficiently conserved among adeleorinids. However, the primer COIII_351_R is the optimisation of COIII_971 and targets a region that is well conserved among available adeleorinid mitochondrial genomes. These primers were able to aid amplification of the COIII protein-coding genes of two anuran *Hepatozoon* spp. and two snake *Hepatozoon* spp., although they were unable to aid amplification of the COIII gene of two other snake *Hepatozoon* and two terrapin *Hepatozoon.* The inconsistency of the primers could be due to a flaw only later realised with primers COIII_273, explanation follows. For primer design, most of the COIII gene sequences used were from anuran *Hepatozoon* spp. (n=6) with some rodent *Hepatozoon* spp. (n=3) and a *Haemogregarina* spp. The primer is designed to target a region 10 bps from the assumed start of the gene. This was problematic since this region falls within a repeat region found in the CytB gene and the COIII gene of *Hepatozoon* spp. that infect anurans. This means that the primer would target both COIII and CytB genes of any *Hepatozoon* spp. that would happen to share this repeat region trait in these two genes. Additionally, this also means that the primers will not bind to DNA in any *Hepatozoon* spp. or adeleorinid that do not have these repeat regions. Therefore, it is likely that *Hepatozoon* spp. in isolates 20A1, 18B1, 19F1, and 07A1 do not retain repeat regions in the COIII and CytB genes as in anuran *Hepatozoon* spp., but this requires further investigation. However, some questions remain unanswered, such as Why did the primers bind in two snake *Hepatozoon* spp. which clade with rodents but not on any of the other snake *Hepatozoon*? Would the primers bind when using different annealing temperatures than those

used to amplify the COIII gene? When these primers were used during amplification of genes in isolates 20A1, 07A1, 19F1, and 18B1 no bands or smears were observed on the gels, which ruled out nonspecific binding and amplification. The COIII fragments obtained for isolate 27A1 was done through a nested PCR starting with primers RNA15_R & COI_VWXHHM_F followed by primers COIII_273_F & COIII_971_R. The same methodology was used for all other snake/reptile *Hepatozoon* isolates but with no positive results.

2.3.6.1.2. Primers COI_329 and COI_1733

Two primers that were available at the start of this study are BarHep_COI_273R, which targets a conserved region at the assumed start of the COI gene, and Api_COI_VWXHHM_F, which targets a conserved region near the middle of the COI gene. The problem with these two primers is that their amplification directions are facing away from each other. Alternative primers were sought, however, in all attempts to amplify the COI gene of *Hepatozoon* spp. none worked. The COI_1733 and COI_329 primers were then designed as a result. The COI_1733 primer targets a conserved region near the assumed start among anuran and rodent *Hepatozoon* spp. COI genes, while the COI_329 primer targets a conserved region near the assumed end of the gene. Unlike the COIII primers previously discussed, the COI primers were designed to only target *Hepatozoon* mitochondrial COI genes. Consequently, it is unlikely that these primers will target any genus outside of *Hepatozoon*. These primers were theoretically tested with bioinformatics software on *Hepatozoon canis* (MH615002), *Klossiella equi* (MH203050), and a *Haemogregarina* mitochondrial COI gene; however, no primer would bind with acceptable mismatch tolerance. These primers could be viable if they were redesigned with some degeneracy to include other genera. Interestingly, these primers were unable to amplify the COI gene of 20A1, 18B1, 19F1, and 07A1, the same isolates for which the COIII gene sequnce was not obtained.

2.3.6.1.3. Large subunit (LSU) primers LSUA2, LSUA1, Api_LSUD, LSUD, and LSUB:

As mentioned previously, alternative primers were used to amplify desired regions within the mitochondrial genomes of *Hepatozoon* spp. This was important since the organisation of mitochondrial genomes among *Hepatozoon* spp. varies depending on the host they infect. For this reason, one primer combination may work for one *Hepatozoon* sp. but not for another. Additional LSU primers were designed as a result, since these RNA fragments are conserved among apicomplexans, and in this case among *Hepatozoon* spp. Unlike the COI primers, the LSU primers were designed with other genera in adeleorina along with *Hepatozoon* in mind. The organisms used to design the LSU primers are two *Haemogregarina* spp., two anuran *Hepatozoon* spp. (*H. ixoxo* and *H. clamatae* (MN245142)), two rodent *Hepatozoon* spp. (*H. griseisciuri* genotype A (MK452388) and *H.* sp (MT936933)), two *Klossia* spp. (*K. helicina*

(NC_058857) and *K. razorback* (NC_058856)), and *Klossiella equi* (MH203050). Out of these primers only the primers of Api_LSUD and LSUB gave acceptable results. Unfortunately, the usefulness of the primers LSUA2 and LSUA1 was not adequately tested. Both primers were used in seven PCR experiments, but no results were obtained. In most PCR runs, the result was multiple bands and, in some cases, no bands. It was concluded that with some of the runs that produced no visible results, the PCR tubes were faulty and caused the PCR mixture to leak. Those PCR runs with multiple bands would probably have benefited from optimised conditions; however, time did not allow for these experiments to be conducted. The LSUD primer was poorly designed due to a lack of knowledge surrounding the primer design process at the time. The LSUD primer has a hairpin Tm of (59.3˚C) higher than the aneling temperature of the primer (47.8 – 53.5˚C). In other words, the primer will form a hairpin loop that cannot bind to the DNA template before it is at its binding (annealing) temperature. Surprisingly, the COIII sequence of 07H2 was generated with the LSUD primer with remarkable quality. The annealing temperature used for the PCR run was 55˚C, well below the hairpin Tm. The primer was redesigned in hopes of optimising it, resulting in the Api_LSUD primer. Api_LSUD was used in one isolate (07H6), which resulted in a prominent single ± 1200bp fragment in the gel electrophoresis visualisation. The fragment size was consistent with the expected size of the fragment; however, upon sequencing the Api_LSUD fragment failed to initiate sequencing. The reason why the primer failed to aid sequencing of the target is unclear and could be due to several reasons, but more experiments will be needed to test its usefulness. The primer was also used during amplification of mitochondrial DNA of isolates 18B1 and 20A1, but to no avail. It could be that the mitochondrial genome of 18B1 and 20A1 is larger than 7000 bps, which would be too large a fragment to amplify under the PCR conditions used. At best the conditions 18B1 and 20A1 were run with Api_LSUD as one of the primers the largest fragment length that would have been produced would have been 5000 bps, yet no bands were visible on the gel.

2.3.6.1.4. Small subunit (SSU) primer SSUD:

Like the LSUs found in the mitochondrial genomes of adeleorinids, SSUs are also present. The SSU SSUD was chosen as amplification target, for which primers were designed. Unfortunately, the primer was not tested because of a lack of sufficient time. The primers are likely to perform well in DNA amplification of the adeleorinid mitochondrial genome, but this has yet to be fully tested.

Chapter 3

General discussion

CHAPTER 3

3.1. Discussion

Studies that deal with the phylogeny of organism groups represented by numerous taxa often struggle to resolve nodes deep into the phylogenetic tree. This is especially true with *Eimeria* and some haemogregarines such as *Hepatozoon*. The 18S has done wonders for molecular biology and phylogenetics in general, however as was seen with the process of resolving the monophyly of the reptile-rodent group within *Hepatozoon* its usefulness is limited. In this study, it was again shown that the 18S rRNA gene carries only a small portion of informative sites. While this works well with sorting larger classifications, the region loses resolution when moving closer to the species level. This in part is why *Haemogregarina* and *Hepatozoon* have such a long history of reclassification and revision. The COI protein-encoding gene of the mitochondrial genome was suggested as a barcode that would be useful for the identification and delimitation of species of apicomplexans (Ogedengbe *et al.,* 2011). According to Ferri *et al.* (2009), barcodes need to be variable between species, short, and contain adequate sites for universal primer design. In the present study, Chapter 2.3 delved into the analysis of mitochondrial protein-coding genes from *Hepatozoon* spp. revealing that the COI gene indeed proves informative for species-level phylogenetic analysis, in line with prior suggestions (Ogedengbe *et al.,* 2011). Additionally, the COIII and CytB genes emerged as more informative than the COI gene, with the COIII gene standing out as the most informative. While the COI gene is considered the best for phylogenetic comparison due to its ability to resolve deep nodes, the drawback lies in the low posterior probabilities for trees constructed with this gene. The CytB gene has been used in phylogenetic analyses of haemosporidians and has been one of the main barcodes used for taxonomy in these parasites (Maia *et al.,* 2016b; Perkins & Schall, 2002). However, it seems that for the *Hepatozoon,* this gene may not serve well as a barcode. While the gene is short enough and has sufficient conserved regions for universal primer design, it seems to lack the phylogenetic signal needed to accurately delimitate nodes. Chapter 2.3 describes how the trees generated by this gene fall short of those drawn with the COI or COIII genes. Although phylogenetic analyses of the CytB gene suggest that the gene is better and more informative than the COI gene, the evidence is sadly not reflected when put to the test. In one occasion the CytB genes of the *Haemogregrina* spp. of Chapter 2.2 was analysed with the CytB gene sequences of those taxa in Supp. Figure 3 and it was observed that the topology of the resulting phylogenetic tree is more resolved. However, the resulting tree was not accurate according to the 18S tree, nonetheless an interesting observation. It could mean that the CytB gene sequence is best used in large data sets, but this remains to be tested once more data is avalible. Even though there were twice as many CytB gene sequences used in the analyses of this study, the marker was still not able to delineate the represented taxa coherently. Furthermore, the gene seems to decrease the nodal support of trees when concatenated with other genetic markers. The COIII gene was also analysed and was found to

be the best of the three mitochondrial genes. Unfortunately, the gene will not be suitable as a barcode for these organisms since it is too variable for primer design. The COI gene does not present these issues and may well be the best of the three coding genes for phylogenetic use because of its ability to resolve deep nodes. However, the posterior probabilities for trees constructed with this gene are low. It was instead discovered that the best way to use the mitochondrial protein-coding genes of *Hepatozoon* is to use them in conjunction. Only when the genes were concatenated was resolution increased. The mitochondrial genomes of haemogregarines, approximately 6500bp in size, are among the smallest in eukaryotes. The small size of these genomes is convenient for generating data for haemogregarine mitochondrial DNA. Chapter 2.1 and 2.2 focused on the amplification and whole genome sequencing of the mitochondrial genomes of two haemogregarine genera, *Hepatozoon* and *Haemogregarina.* Whole genome sequencing of the mitochondrial genome for haemogregarines is beneficial. Considering the results in Chapter 2.3, this method produces all three protein-coding genes without the need for primers specific to the gene regions. Furthermore, sequencing the complete mitochondrial genome present an accurate approach when encountering mixed infections. The genomes also retain a conserved organisation among closely related species, which could aid in species and generic classification or delineation. It can be a laborious process, sequencing the complete mitochondrial genome, but the ~6kb mitogenomes of apicomplexans are small enough to make the task viable. This study is the first attempt to use mitochondrial DNA in conjunction with nuclear DNA for phylogenetic analyses of *Hepatozoon.* The data from this study build on the move towards larger mitochondrial DNA datasets to aid in providing better resolution and more robust phylogenetic analyses of the haemogregarines. There is still much that we do not understand about how the different mitochondrial protein-coding genes describe the evolution of haemogregarines. Currently, it is unclear what the effect of additional sequences would have on the analyses. In some cases, it was discovered that the removal or addition of DNA sequences to the data sets changed the resulting tree topologies. Testing the datasets under the two different conditions revealed that for the COIII tree and the COI-COIII-CytB tree the accuracy of tree topology is dependent on the presence of *Hepatozoon canis* sequences in the analyses. The limited data of this study hinted at the possible roles each protein-coding gene plays in sorting the topologies of phylogenetic trees of *Hepatozoon*. It would be interesting to see how the different roles of each gene are revealed as data sets increase, especially for other genera. Of the haemogregarines species of *Hepatozoon* are the only to have had their mitochondrial genomes sequence. This study is among the first to sequence the mitochondrial genomes of more than one species of *Haemogregarina*. Unfortunately, this also means the lack of mtDNA data available for this genus limits the phylogenetic analyses that can be conducted. This is unfortunate since the mitochondrial genome of *Haemogregarina* has intriguing sequence diversity and its organisation is similar to those of *Klossia* species. Phylogenetic analysis with the protein-coding genes of these

adeleorinids would shed light on the phylogenetic usefulness of the protein-coding genes, how *Haemogregarina* and *Klossia* are related, and possibly why the mitochondrial genomes of adeleorinids have so many diverse properties. It may also be that the way the *Haemogregarina* protein-coding genes are used for phylogeny differs significantly from how those from *Hepatozoon* are used. This may also be true for every other adeleorinid genera. Our understanding of adeleorinid mitochondrial DNA is limited to four genera of 15 that comprise the adeleorinids. Within the haemogregarines mitochondrial DNA is lacking for *Hemolivia, Karyolysus, Babesiosoma, Dactylosoma, Cyliria* and *Desseria* (see Figure 3.1).

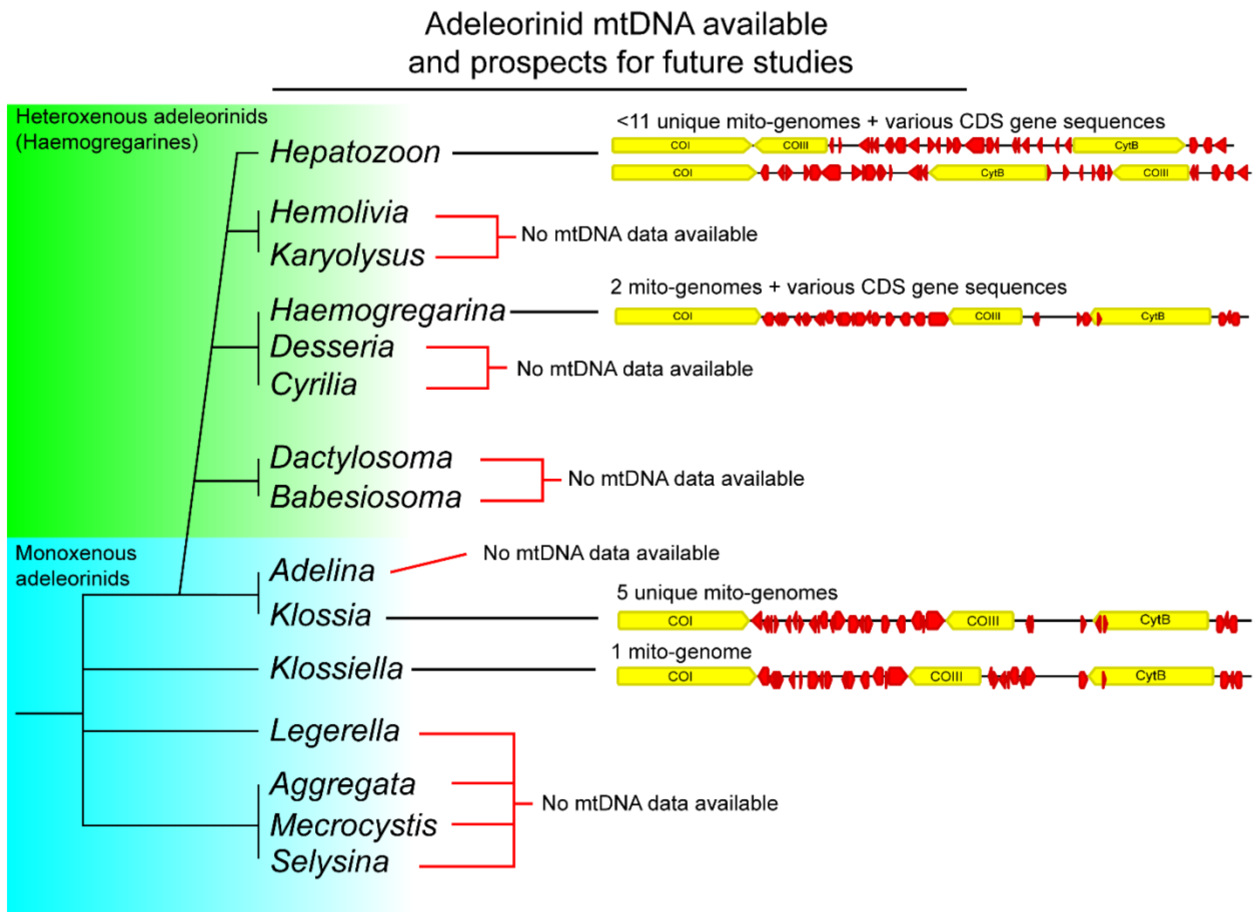


Figure 3.1: An illustration of taxa within adeleorina, the mitochondrila DNA that has been gathered and those that are not yet available. All-in-all there is approximately nineteen unique mitochondrial genomes available for four adeleorinid coccidia out of fifteen, and various protein-coding gene sequences. The figure does not include repeated genomes, or genomes from the same organism.

The lack of mtDNA data is a testament to the infancy of studies surrounding the mitochondrial DNA of adeleorinids. Despite this great lack of data, some interesting observations have been made of the years and in this study. However some questions remain to be fully explained. With the introduction of mtDNA for these genera to those of *Hepatozoon* and *Haemogregarina* some of these quastions could be answered. These include the organisation, size and topology of mt-

genomes, if genome organisation differs between genera, if they are much bigger than the ±6500 bp average for adeleorinids and if there are any topology differences between genera. It may be found that a difference in mt-genome organisation is significant for genus identification, which would ultimately mean that *Hepatozoon* spp. from rodents are of a different genus than anuran *Hepatozoon*. The insights gained through the study of adeleorinid mitochondrial genomes may broaden our understanding of the evolution of mitochondrial genomes within a group. By studying and revealing the evolutionary mechanisms of mitochondrial genomes across taxa in adeleorina, our understanding of how mitochondrial genomes evolve in other apicomplexans may broaden. This may especially be valuable for the drug development for virulent protozoa, such as *Plasmodium*. The studies done here raised more exciting questions than those they attempted to answer. In a way, these questions are an answer in themselves. The use of mitochondrial protein-coding genes as genetic markers within adeleorina may not be as straightforward and may just be a delicate dance between different pockets of evidence, expertly stitched together to decipher the rich history of these parasites.

REFERENCE LIST

Adobe Inc. 2019. *Adobe Illustrator*. https://adobe.com/products/illustrator Date of access: 15 Nov. 2023.

Attia El Hili, R., Achouri, M.S. and Verneau, O. 2020. The genetic diversity of blood parasites within the freshwater turtles *Mauremys leprosa* and *Emys orbicularis* in Tunisia reveals coinfection with *Haemogregarina* spp. *Parasitology research*, **119**:3315-3326

Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic acids research*, **27**(8):1767-3326. Doi: 10.1007/s00436-020-06821-1

Barta, J.R. 2000. Suborder Adeleorina Leger, 1911. *Illustrated guide to the protozoa,* **1**:308-318.

Barta, J.R., Martin, D.S., Liberator, P.A., Dashkevicz, M., Anderson, J.W., Feighner, S.D., Elbrecht, A., Perkins-Barrow, A., Jenkins, M.C., Danforth, H.D. and Ruff, M.D. 1997. Phylogenetic relationships among eight *Eimeria* species infecting domestic fowl inferred using complete small subunit ribosomal DNA sequences. *The Journal of parasitology*, **83**(2):262. Doi: 10.2307/3284453

Barta, J.R., Ogedengbe, J.D., Martin, D.S. and Smith, T.G. 2012. Phylogenetic position of the Adeleorinid coccidia (Myzozoa, Apicomplexa, Coccidia, eucoccidiorida, Adeleorina) inferred using 18S rDNA sequences. *Journal of eukaryotic microbiology*, **59**(2):171-180. Doi:10.1111/j.1550-7408.2011.00607.x

Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic acids research*, **27**(8):1767-1780. Doi:10.1093/nar/27.8.1767

Brands, S.J. 1989-present. The Taxonomicon. Universal Taxonomic Services, Zwaag, The Netherlands. http://taxonomicon.taxonomy.nl/ Date of access: 15 Nov. 2023

Brown T.A. 2002. Chapter 16: Molecular Phylogenetics. (*In* Brown, TA, *ed*. Genomes. 2nd edition. Oxford: Wiley-Liss. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21122/ Date of access: 15 Nov. 2023.

Bustin, S. and Huggett, J. 2017. QPCR Primer Design revisited. *Biomolecular detection and quantification*, **14**:19-28. Doi:10.1016/j.bdq.2017.11.001

REFERENCE LIST

Caterino, M.S., Reed, R.D., Kuo, M.M. and Sperling, F.A. 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Systematic biology*, **50**(1):106-127. Doi: 10.1080/10635150119988

Carreno, R.A. and Barta, J.R. 1999. An eimeriid origin of isosporoid coccidia with Stieda bodies as shown by phylogenetic analysis of small subunit ribosomal RNA gene sequences. *The Journal of parasitology*:77-83.

Cardoso, W.A., Perles, L., Picelli, A.M., Correa, J.K., André, M.R., and Viana, L.A. 2022. *Hepatozoon* parasites (Apicomplexa: Hepatozoidae) in fish *Hoplias aimara* (Characiformes, Erythrinidae) from the Eastern Amazon, Brazil. *Parasitology research*, **121**(3):1041-1046. Doi:10.1007/s00436-022-07462-2

Cavalier-Smith, T. 2014. Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa. *European journal of protistology*, **50**(5):472-495. Doi: 10.1016/j.ejop.2014.07.002

Cook, C.A., Netherlands, E.C. and Smit, N.J. 2016. Redescription, molecular characterisation and taxonomic re-evaluation of a unique African Monitor Lizard Haemogregarine *Karyolysus paradoxa* (Dias, 1954) n. comb. (Karyolysidae). *Parasites and vectors*, **9**(1):1-13. Doi: 10.1186/s13071-016-1600-8

Cook, C.A., Netherlands, E.C., Smit, N.J. and Van As, J. 2018. Two new species of *Hepatozoon* (Apicomplexa: Hepatozoidae) parasitising species of *Philothamnus* (Ophidia: Colubridae) from South Africa. *Folia parasitologica*, **65**. Doi:10.14411/fp.2018.004

Correa, J.K.C., Picelli, A.M., da Silva, M.R.L., Valadão, R.M., Hernández-Ruz, E.J. and Viana, L.A. 2022. Phylogenetic analysis of chelonian hemogregarines reveals shared species among the Amazonian freshwater turtle *Podocnemis* spp. and provides a description of two new species of *Haemogregarina*. *Parasitology research*, **121**(2):691-701. Doi: 10.1007/s00436-021-07405-3

Criado-Fornelio, A., Martinez-Marcos, A., Buling-Saraña, A., and Barba-Carretero, J.C. 2003a. Molecular studies on *Babesia*, *Theileria* and *Hepatozoon* in southern Europe: part I. Epizootiological aspects. *Veterinary parasitology*, **113**(3-4):189-201. Doi:10.1645/ge-464r.1

Criado-Fornelio, A., Martinez-Marcos, A., Buling-Sarana, A., and Barba-Carretero, J.C. 2003b. Molecular studies on *Babesia*, *Theileria* and *Hepatozoon* in southern Europe: Part

II. Phylogenetic analysis and evolutionary history. *Veterinary parasitology*, **114**(3):173-194. Doi: 10.1016/s0304-4017(03)00141-9

Criado-Fornelio, A., Ruas, J.L., Casado, N., Farias, N.A.R., Soares, M.P., Müller, G., Brum, J.G.W., Berne, M.E.A., Buling-Saraña, A. and Barba-Carretero, J.C. 2006. New molecular data on mammalian *Hepatozoon* species (Apicomplexa: Adeleorina) from Brazil and Spain. *Journal of parasitology*, **92**(1):93-99. Doi:10.1645/ge-464r.1

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, **9**(8):772. Doi: 10.1038/nmeth.2109

Davies, A.J. and Johnston, M.R.L. 2000. The biology of some intraerythrocytic parasites of fishes, amphibia and reptiles. *Advances in parasitology*, **45**:1-107. Doi: 10.1016/s0065-308x(00)45003-7

Davies, A.J., Smit, N.J., Hayes, P.M., Seddon, A.M., and Wertheim, D. 2004. *Haemogregarina* bigemina (protozoa: Apicomplexa: Adeleorina) - past, present and future. *Folia parasitologica*, **51**(2-3):99-108. Doi: 10.14411/fp.2004.015

Desser, S.S. 1993. The Haemogregarinidae and Lankesterellidae. *Parasitic protozoa*:247-272. Doi:10.1016/b978-0-12-426014-6.50009-0

Desser, S.S., Hong, H., and Martin, D.S. 1995. The life history, ultrastructure, and experimental transmission of *Hepatozoon catesbianae* n. comb., an apicomplexan parasite of the Bullfrog, *Rana catesbeiana* and the mosquito, *Culex territans* in Algonquin Park, Ontario. *The Journal of parasitology*, **81**(2):212. Doi: 10.2307/3283922

Esteves-Silva, P.H., da Silva, M.R., O'Dwyer, L.H., Tavares-Dias, M., and Viana, L.A. 2019. *Haemogregarina daviesensis* sp. nov. (Apicomplexa: Haemogregarinidae) from South American Lungfish *Lepidosiren paradoxa* (Sarcopterygii: Lepidosirenidae) in the eastern Amazon Region. *Parasitology research*, **118**(10):2773-2779. Doi: 10.1007/s00436-019-06430-7

Farris, J.S., Källersjö, M., Kluge, A.G., and Bult, C. 1994. Testing significance of incongruence. *Cladistics*, **10**(3):315-319. Doi:10.1111/j.1096-0031.1994.tb00181.x

Feagin, J.E. 1994. The extrachromosomal DNAs of apicomplexan parasites. *Annual review of microbiology*, **48**(1):81-104. Doi: 10.1146/annurev.mi.48.100194.000501

REFERENCE LIST

Ferri, G., Alu, M., Corradini, B., Licata, M., and Beduschi, G. 2009. Species identification through DNA ‘‘barcodes’’. *Genetic testing and molecular biomarkers*, **13**:421-426. Doi: 10.1089=gtmb.2008.0144

Gutierrez-Liberato, G.A., Lotta-Arévalo, I.A., González, L.P., Vargas-Ramírez, M., Rodríguez-Fandiño, O., Cepeda, A.S., Ortiz-Moreno, M.L. and Matta, N.E. 2021. The genetic and morphological diversity of *Haemogregarina* infecting turtles in Colombia: Are mitochondrial markers useful as barcodes for these parasites? *Infection, genetics and evolution*, **95.** Doi: 10.1016/j.meegid.2021.105040

Haklová-Kočíková, B., Hižňanová, A., Majláth, I., Račka, K., Harris, D.J., Földvári, G., Tryjanowski, P., Kokošová, N., Malčeková, B. and Majláthová, V. 2014. Morphological and molecular characterization of *Karyolysus* – a neglected but common parasite infecting some European lizards. *Parasites and vectors*, **7**(1). Doi: 10.1186/s13071-014-0555-x

Hikosaka, K., Kita, K. and Tanabe, K. 2013. Diversity of mitochondrial genome structure in the phylum Apicomplexa. *Molecular and biochemical parasitology*, **188**(1):26-33. Doi: 10.1016/j.molbiopara.2013.02.006

Holland, B.S., Dawson, M.N., Crow, G.L. and Hofmann, D.K. 2004. Global phylogeography of *Cassiopea* (Scyphozoa: Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. *Marine biology*, **145**:1119-1128. Doi: 10.1007/s00227-004-1409-4

Hrazdilová, K., Červená, B., Blanvillain, C., Foronda, P. and Modrý, D. 2021. Quest for the type species of the genus *Hepatozoon* – phylogenetic position of hemogregarines of rats and consequences for taxonomy. *Systematics and biodiversity*, **19**(6):622-631. Doi: 10.1080/14772000.2021.1903616

Huelsenbeck, J.P., and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics,* **17**:754-755. Doi: 10.1093/bioinformatics/17.8.754

Karadjian, G., Chavatte, J.M. and Landau, I. 2015. Systematic revision of the adeleid haemogregarines, with creation of *Bartazoon* ng, reassignment of *Hepatozoon argantis* Garnham, 1954 to *Hemolivia*, and molecular data on *Hemolivia stellata*. *Parasite*, **22.** Doi: 10.1051/parasite/2015031

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. and Thierer, T. 2012. Geneious Basic: an

integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**(12):1647-1649. Doi: 10.1093/bioinformatics/bts199

Kopečná, J., Jirků, M., Oborník, M., Tokarev, Y.S., Lukeš, J. and Modrý, D. 2006. Phylogenetic analysis of coccidian parasites from invertebrates: search for missing links. *Protist*, **157**(2):173-183. Doi: 10.1016/j.protis.2006.02.005

Koressaar, T., and Remm, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**(10):1289-1291.

Kowalczyk, M., Staniszewski, A., Kamiñska, K., Domaradzki, P., and Horecka, B. 2021. Advantages, possibilities, and limitations of mitochondrial DNA analysis in molecular identification. *Folia biologica (kraków)*, **69**(3):101-111. Doi: 10.3409/fb_69-3.12

Lane, N. 2009. Biodiversity: On the origin of Barcodes. *Nature*, **462**(7271):272-274. Doi: 10.1038/462272a

Le, T.H., Blair, D., Agatsuma, T., Humair, P.F., Campbell, N.J., Iwagami, M., Littlewood, D.T., Peacock, B., Johnston, D.A., Bartley, J., Rollinson, D., Herniou, E.A., Zarlenga, D.S., and McManus, D.P. 2000. Phylogenies inferred from mitochondrial gene orders — a cautionary tale from the parasitic flatworms. *Molecular biology and evolution*, **17**:1123-1125. Doi: 10.1093/oxfordjournals.molbev.a026393

Léveillé, A.N. 2019a. Scratching the surface: Diversity among the first sequenced extrachromosomal genomes of parasites in the suborder Adeleorina (Apicomplexa) with a focus on *Hepatozoon* species. Ontario: University of Guelph (Thesis - PhD).

Léveillé, A.N., Baneth, G. and Barta, J.R. 2019b. Next generation sequencing from *Hepatozoon canis* (Apicomplexa: Coccidia: Adeleorina): Complete apicoplast genome and multiple mitochondrion-associated sequences. *International journal for parasitology*, **49**(5):375-387. Doi: 10.1016/j.ijpara.2018.12.001

Léveillé, A.N., Bland, S.K., Carlton, K., Larouche, C.B., Kenney, D.G., Brouwer, E.R., Lillie, B.N., and Barta, J.R. 2019c. *Klossiella equi* infecting kidneys of ontario horses: Life cycle features and multilocus sequence-based genotyping confirm the genus *Klossiella* belongs in the Adeleorina (apicomplexa: Coccidia). *Journal of parasitology*, **105**(1):29-40. Doi: 10.1645/18-80

Léveillé, A.N., El Skhawy, N. and Barta, J.R. 2020. Multilocus sequencing of *Hepatozoon* cf. *griseisciuri* infections in Ontario eastern gray squirrels (*Sciurus carolinensis*) uncovers two

genotypically distinct sympatric parasite species. *Parasitology research*, **119**(2):713-724. Doi: 10.1007/s00436-019-06583-5

Léveillé, A.N., Ogedengbe, M.E., Hafeez, M.A., Tu, H-H. (A.) and Barta, J.R. 2014. The complete mitochondrial genome sequence of *Hepatozoon catesbianae* (Apicomplexa: Coccidia: Adeleorina), a blood parasite of the green frog, *Lithobates* (formerly *Rana*) *clamitans*. *Journal of parasitology*, **100**(5):651-656. Doi: 10.1645/13-449.1

Léveillé, A.N., Zeldenrust, E.G. and Barta, J.R. 2021. Multilocus genotyping of sympatric *Hepatozoon* species infecting the blood of Ontario ranid frogs reinforces species differentiation and identifies an unnamed *Hepatozoon* species. *The Journal of parasitology*, **107**(2):246-261. Doi: 10.1645/20-18

Levine, N.D. 1988. The Protozoan Phylum Apicomplexa (CRC, Boca Raton, FL).

Maia, J.P., Carranza, S. and Harris, D.J. 2016a. Comments on the systematic revision of adeleid haemogregarines: are more data needed? *Journal of Parasitology*, **102**(5):549-552. Doi: doi.org/10.1645/15-930

Maia, J.P., Harris, D.J. and Carranza, S. 2016b. Reconstruction of the evolutionary history of Haemosporida (Apicomplexa) based on the cyt b gene with characterization of *Haemocystidium* in geckos (Squamata: Gekkota) from Oman. *Parasitology international*, **65**(1):5-11. Doi: 10.1016/j.parint.2015.09.003

Maia, J.P., Perera, A. and Harris, D.J. 2012. Molecular survey and microscopic examination of *Hepatozoon* Miller, 1908 (Apicomplexa: Adeleorina) in lacertid lizards from the western Mediterranean. *Folia parasitologica*, **59**(4):241-248. Doi: 10.14411/fp.2012.033

Martynova-Vankley, A., Syvyk, A., Teplova, I., Hume, M., and Nalian, A. 2008. Rapid detection of avian *Eimeria* species using denaturing gradient gel electrophoresis. *Poultry science,* **87**(9):1707-1713

Mathur, V., Kwong, W.K., Husnik, F., Irwin, N.A., Kristmundsson, Á., Gestal, C., Freeman M. and Keeling, P.J. 2020. Phylogenomics identifies a new major subgroup of Apicomplexans, Marosporida *class nov.*, with extreme apicoplast genome reduction. *Genome biology and evolution*, **13**(2). Doi:10.1093/gbe/evaa244

Morrison, D.A., Bornstein, S., Thebo, P., Wernery, U., Kinne, J. and Mattsson, J.G. 2004. The current status of the small subunit rRNA phylogeny of the Coccidia (Sporozoa). *International journal for parasitology*, **34**(4):501-514. Doi: 10.1016/j.ijpara.2003.11.006

Netherlands, E.C., Cook, C.A. and Smit, N.J. 2014a. *Hepatozoon* species (Adeleorina: Hepatozoidae) of African bufonids, with morphological description and molecular diagnosis of *Hepatozoon ixoxo* sp. nov. parasitising three *Amietophrynus* species (Anura: Bufonidae). *Parasites and vectors*, **7**(1):1-12. Doi: 10.1186/s13071-014-0552-0

Netherlands, E.C., Cook, C.A., Du Preez, L.H., Vanhove, M.P., Brendonck, L. and Smit, N.J. 2018. Monophyly of the species of *Hepatozoon* (Adeleorina: Hepatozoidae) parasitizing (African) anurans, with the description of three new species from hyperoliid frogs in South Africa. *Parasitology*, **145**(8):1039-1050.

Netherlands, E.C., Cook, C.A., du Preez, L.H., Vanhove, M.P.M., Brendonck, L. and Smit, N.J. 2020. An overview of the Dactylosomatidae (Apicomplexa: Adeleorina: Dactylosomatidae), with the description of *Dactylosoma kermiti* n. sp. parasitising *Ptychadena anchietae* and *Sclerophrys gutturalis* from South Africa. *International journal for parasitology: parasites and wildlife*, **11**:246-260. Doi: 10.1016/j.ijppaw.2019.12.006

Netherlands, E.C., Cook, C.A., du Preez, L.H., Vanhove, M.P.M., Brendonck, L. and Smit, N.J. 2017. Monophyly of the species of *Hepatozoon* (Adeleorina: Hepatozoidae) parasitizing (African) anurans, with the description of three new species from hyperoliid frogs in South Africa. *Parasitology*, **145**(8):1039-1050. Doi: 10.1017/s003118201700213x

Netherlands, E.C., Cook, C.A., Kruger, D.J.D., du Preez, L.H. and Smit, N.J. 2015. Biodiversity of frog haemoparasites from sub-tropical northern KwaZulu-Natal, South Africa. *International journal for parasitology: parasites and wildlife*, **4**(1):135-141. Doi: 10.1016/j.ijppaw.2015.01.003

Netherlands, E.C., Cook, C.A., Smit, N.J. and du Preez, L.H. 2014b. Redescription and molecular diagnosis of *Hepatozoon theileri* (Laveran, 1905) (Apicomplexa: Adeleorina: Hepatozoidae), infecting *Amietia quecketti* (Anura: Pyxicephalidae). *Folia parasitologica*, **61**(4):293-300. Doi: 0.14411/fp.2014.046

O'Donoghue, P. 2017. Haemoprotozoa: making biological sense of molecular phylogenies. *International journal for parasitology: parasites and wildlife*, **6**(3):241-256. Doi: 10.1016/j.ijppaw.2017.08.007

Ogedengbe, J.D., Hanner, R.H. and Barta, J.R. 2011. DNA barcoding identifies *Eimeria* species and contributes to the phylogenetics of coccidian parasites (Eimeriorina, Apicomplexa, Alveolata). *International journal for parasitology*, **41**(8):843-850. Doi: 10.1016/j.ijpara.2011.03.007

Ogedengbe, M.E., El-Sherry, S., Ogedengbe, J.D., Chapman, H.D. and Barta, J.R. 2018. Phylogenies based on combined mitochondrial and nuclear sequences conflict with morphologically defined genera in the eimeriid coccidia (Apicomplexa). *International journal for parasitology*, **48**(1):59-69. Doi:10.1016/j.ijpara.2017.07.008

Okamoto, N. and Keeling, P.J. 2014. The 3D structure of the apical complex and association with the flagellar apparatus revealed by serial TEM tomography in *Psammosa pacifica*, a distant relative of the Apicomplexa. *Plos one*, **9**(1):e84653. Doi: 10.1371/journal.pone.0084653

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukeš, J., Mann, D.G., Mitchell, E.A.D., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G.B., Smirnov, A.V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindel, D. and de Vargas, C. 2012. CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLos biology*, **10**(11):e1001419. Doi: 10.1371/journal.pbio.1001419

Perkins, S.L. and Schall, J. 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *Journal of parasitology*, **88**(5):972-978. Doi: 10.2307/3285540

Reed, R.D. and Sperling, F. 1999. Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus *Papilio*. *Molecular biology and evolution*, **16**(2):286-297. Doi: 10.1093/oxfordjournals.molbev.a026110

Rossow, J.A., Hernandez, S.M., Sumner, S.M., Altman, B.R., Crider, C.G., Gammage, M.B., Segal, K.M. and Yabsley, M.J. 2013. Haemogregarine infections of three species of aquatic freshwater turtles from two sites in Costa Rica. *International journal for parasitology: parasites and wildlife*, **2**:131-135. Doi: 10.1016/j.ijppaw.2013.02.003

Rubinoff, D. and Holland, B.S. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic biology*, **54**(6):952-961. Doi: 10.1080/10635150500234674

Rubinoff, D. and Sperling, F.A. 2002. Evolution of ecological traits and wing morphology in *Hemileuca* (Saturniidae) based on a two-gene phylogeny. *Molecular phylogenetics and evolution*, **25**(1):70-86. Doi: 10.1016/S1055-7903(02)00213-0

Rubini, A.S., dos Santos Paduan, K., Perez, R.R., Ribolla, P.E.M. and O'Dwyer, L.H. 2006. Molecular characterization of feline *Hepatozoon* species from Brazil. *Veterinary Parasitology*, **137**(1-2):68-171. Doi:10.1016/j.vetpar.2005.12.008

Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., Cavalier-Smith, T., Guiry, M.D. and Kirk, P.M. 2015. A higher level classification of all living organisms. *Plos one*, **10**(4):e0119248. Doi: 10.1371/journal.pone.0130114

Siddall, M.E. 1995. Phylogeny of adeleid blood parasites with a partial systematic revision of the haemogregarine Complex. *The Journal of eukaryotic microbiology*, **42**(2):116-125. Doi: 10.1111/j.1550-7408.1995.tb01551.x

Sloboda, M., Kamler, M., Bulantová, J., Votýpka, J., and Modrý, D. 2007. A new species of *Hepatozoon* (Apicomplexa: Adeleorina) from *Python regius* (Serpentes: Pythonidae) and its experimental transmission by a Mosquito vector. *Journal of parasitology*, **93**(5):1189-1198. Doi:10.1645/ge-1200r.1

Smith, T.G. 1996. The genus *Hepatozoon* (Apicomplexa: Adeleina). *The Journal of parasitology,* **82**(4):565-585. Doi:10.2307/3283781

Smith, T.G. and Desser, S.S. 1997. Phylogenetic analysis of the genus *Hepatozoon* Miller, 1908 (Apicomplexa: Adeleorina). *Systematic parasitology*, **36**(3):213-221. Doi:10.1023/a:1005721501485.

Smith, T.G., Kim, B., Hong, H. and Desser, S.S. 2000. Intraerythrocytic development of species of *Hepatozoon* infecting ranid frogs: Evidence for convergence of life cycle characteristics among apicomplexans. *The Journal of parasitology*, **86**(3):451-458. Doi:10.2307/3284856.

Swofford, D.L. and Sullivan, J. 2009. Phylogeny inference based on parsimony and other methods using PAUP. *The Phylogenetic handbook*:267-312. Doi:10.1017/cbo9780511819049.010.

Teletchea, F. 2010. After 7 years and 1000 citations: Comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*, **21**(6):206-226. Doi: 10.3109/19401736.2010.532212

Telford Jr, S.R., Norton, T.M., Moler, P.E. and Jensen, J.B. 2009. A new *Haemogregarina* species of the alligator snapping turtle, *Macrochelys temminckii* (Testudines:

Chelydridae), in Georgia and Florida that produces macromeronts in circulating erythrocytes. *Journal of parasitology*, **95**(1):208-214.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. 2012. Primer3—new capabilities and interfaces. *Nucleic acids research*, **40**(15):e115. Doi:10.1093/nar/gks596.

Wang, X-C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J-H., Cai, D. and Li, J-Q. 2014. ITS1: A DNA barcode better than ITS2 in eukaryotes? *Molecular ecology resources*, **15**(3):573-586. Doi: 10.1111/1755-0998.12325

Zhao, X., Duszynski, D.W. and Loker, E.S. 2001. Phylogenetic position of *Eimeria antrozoi*, a bat coccidium (Apicomplexa: Eimeriidae) and its relationship to morphologically similar *Eimeria* spp. from bats and rodents based on nuclear 18S and Plastid 23S rDNA sequences. *The Journal of parasitology*, **87**(5):1120-1123. Doi: 10.2307/3285244

APPENDIX A: SUPPLEMENTARY DATA

Chapter 2.1

Supp. Table 1: COI and RNA fragments of the mitochondrial genomes generated for *Hepatozoon ixoxo, H. theileri* and *H. tenuis.*

		Hepatozoon ixoxo			*Hepatozoon theileri*			*Hepatozoon tenuis*			
Content	Type	Min	Max	Length (bp)	Min	Max	Length (bp)	Min	Max	Length (bp)	Direction
COI	CDS	1	1467	1467	1	1467	1467	1	1476	1476	Forward
SSUA	rRNA	1514	1585	72	1503	1574	72	1523	1594	72	reverse
RNA2	rRNA	1683	1748	66	1622	1687	66	1660	1721	62	reverse
RNA3	rRNA	1765	1831	67	1705	1771	67	1739	1805	67	forward
SSUD	rRNA	1955	2013	59	1861	1919	59	1903	1961	59	forward
RNA10	rRNA	2020	2125	106	1926	2030	105	1968	2075	108	forward
LSUE	rRNA	2129	2318	190	2034	2223	190	2079	2268	190	reverse
LSUA 2	rRNA	2444	2547	104	2315	2418	104	2398	2501	104	forward
RNA15	rRNA	2548	2575	28	2419	2446	28	2502	2529	28	forward
SSUB	rRNA	2576	2688	113	2447	2559	113	2530	2642	113	forward
RNA14	rRNA	2695	2726	32	2566	2597	32	2649	2680	32	reverse
RNA6	rRNA	2730	2785	56	2601	2656	56	2684	2739	56	forward
LSUC	rRNA	2822	2843	22	2672	2693	22	2777	2798	22	forward
LSUF	rRNA	3011	3118	108	2861	2968	108	2966	3073	108	reverse
RNA18	rRNA	3123	3147	25	2973	2997	25	3082	3106	25	reverse
SSUF	rRNA	3148	3202	55	2998	3052	55	3107	3161	55	reverse
CytB	CDS	3224	4411	1188	3130	4236	1107	3219	4331	1113	reverse
LSUA 1	rRNA	4423	4464	42	4257	4298	42	4371	4412	42	forward
RNA9	rRNA	4595	4644	50	4429	4478	50	4543	4592	50	forward
RNA13	rRNA	4753	4782	30	4569	4598	30	4694	4723	30	reverse
LSUB	rRNA	4879	4912	34	4691	4724	34	4820	4853	34	forward
RNA5	rRNA	4929	5005	77	4744	4820	77	4874	4950	77	reverse
SSUE	rRNA	5041	5081	41	4845	4885	41	5005	5045	41	forward
COIII	CDS	5092	5865	774	4896	5555	660	5059	5799	741	reverse
RNA17	rRNA	5880	5917	38	5680	5717	38	5844	5881	38	reverse
RNA19	rRNA	5923	5955	33	5723	5755	33	5886	5918	33	reverse
LSUD	rRNA	6106	6185	80	5871	5950	80	6071	6150	80	forward
RNA1	rRNA	6228	6311	84	5993	6076	84	6193	6276	84	reverse
LSUG	rRNA	6358	6463	106	6119	6224	106	6330	6435	106	reverse

Description: This table describes the structure of the mitochondrial genome of the three *Hepatozoon* spp. investigated in this study. The first column describes the content of the mitochondrial genomes followed by the type of content being described in the second column. The content of the mitochondrial genomes is mainly CoDing sequences (CDS) and rRNA fragments. In each mitochondrial genome there are three CDSs and multiple large subunit (LSU) and small subunit (SSU) rRNA fragments. The positions of these CDSs and rRNA fragments are indicated by an min (minimum) and max (maximum) position relative to the first nucleotide of the genome sequence. In the case of this study the assumed start codon of the COI gene sequence was chosen as the start of each genome. The length of each CDS and rRNA is indicated by a number representing the unit length in base pairs. The last column of the table describes the orientation of a CDS or rRNA fragment relative to the direction of the COI gene sequence of each genome.

APPENDIX

Chapter 2.2

Supp. Table 2: Annotation of fragmented ribosomal RNA (rRNA) genes encoded in the mitochondrial genomes of *Haemogregarina* sp. A and B based on pairwise comparisons with conserved annotated genes from *Klossiella equi* (MH203050), and *Hepatozoon catesbianae* (KF894962).

		A			B		
Fragmented rDNA[a]	Type rRNA	Beginning (bp)	End (bp)	Length (bp)	Beginning (bp)	End (bp)	Length (bp)
SSUA	SSU	3,96	4,045	86	3,931	4,015	85
SSUB	SSU	4,261	4,375	115	4,232	4,344	113
SSUD	SSU	3,864	3,931	68	3,833	3,904	72
SSUE	SSU	408	448	41	391	437	47
SSUF	SSU	3,466	3,521	56	3,445	3,503	59
RNA1	LSU	4,711	4,798	88	4,685	4,77	86
RNA2	LSU	3,56	3,608	49	3,545	3,589	45
RNA3	LSU	4,434	4,508	75	4,399	4,476	78
RNA5	SSU	1,61	1,688	79	1,589	1,667	79
RNA6	LSU	3,781	3,832	52	3,751	3,802	52
RNA7	-	265	347	83	258	340	83
RNA8	SSU	1,713	1,805	93	1,696	1,782	87
RNA9	SSU	6,019	6,052	34	5,96	5,998	39
RNA10	LSU	208	257	50	144	246	103
RNA11	LSU	3,666	3,717	52	3,638	3,692	55
RNA13	LSU	3,522	3,549	28	3,504	3,533	30
RNA14	SSU	4,375	4,415	41	4,348	4,388	41
RNA15	-	4,233	4,253	21	4,205	4,225	21
RNA17	SSU	4,076	4,115	40	4,046	4,085	40
RNA18	LSU	3,936	3,96	25	3,908	3,927	20
LUSA 1/2	LSU	3,735	3,779	45	3,706	3,746	41
LSUA 2/2	LSU	4,118	4,228	111	4,089	4,2	112
LSUB	LSU	1,682	1,719	38	1,664	1,702	39
LSUC	LSU	6,058	6,079	22	6,001	6,022	22
LSUD	LSU	4,565	4,648	84	4,542	4,624	83
LSUE	LSU	4,987	5,18	194	4,949	5,147	199
LSUF	LSU	3,355	3,456	102	3,334	3,436	103
LSUG	LSU	4,841	4,945	105	4,807	4,917	111

[a]Nomenclature follows that of Léveillé *et al.* (2020).

APPENDIX

CHAPTER 2.3

Supp. Table 3: Some of the sequence data used to construct the phylogenetic trees of this study was extracted from Genbank and are listed.

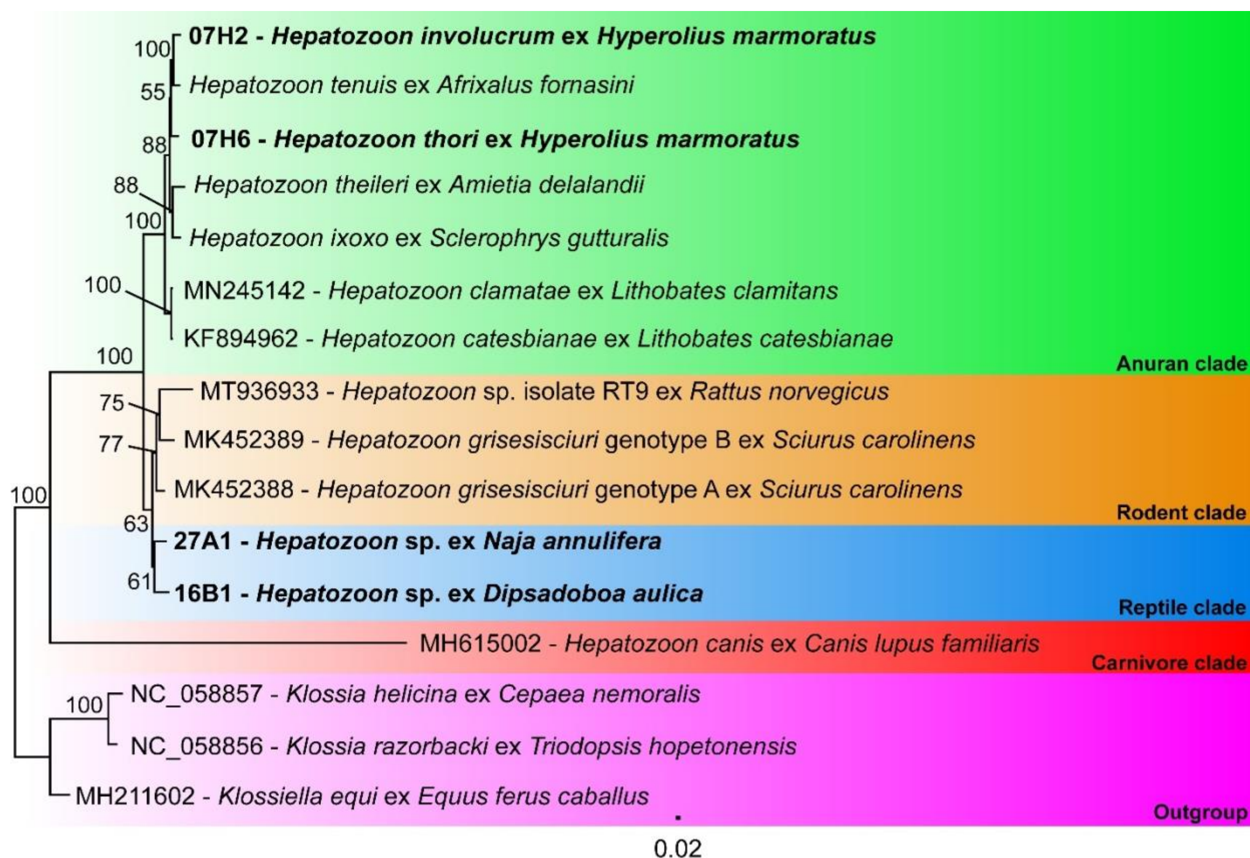
Taxon		Accession #			
Organism	Host	18S	COI	COIII	CytB
Hepatozoon thori	*Hyperolius marmoratus*	MG041603	SAMPLE	SAMPLE	SAMPLE
Hepatozoon involucrum	*Hyperolius marmoratus*	MG041594	SAMPLE	SAMPLE	SAMPLE
Hepatozoon tenuis	*Afrixalus fornasini*	MG041596	SAMPLE	SAMPLE	SAMPLE
Hepatozoon ixoxo	*Sclerophrys gutturalis*	MG041604	SAMPLE	SAMPLE	SAMPLE
Hepatozoon catesbianae	*Aquarana catesbeiana*	MN244528	KF894962	KF894962	KF894962
Hepatozoon clamatae	*Aquarana clamitans*	MN244529	MN245142	MN245142	MN245142
Hepatozoon theileri	*Amietia delalandii*	MG041605	SAMPLE	SAMPLE	SAMPLE
Hepatozoon magna	*Pelophylax lessonae*	HQ224960	N/D	N/D	N/D
Hepatozoon cecilhoarei	*Philothamnus natalensis natalensis*	MG519504	N/D	N/D	N/D
Hepatozoon angeladaviesae	*Philothamnus semivariegatus*	MG519502	N/D	N/D	N/D
Hepatozoon sp. - 07A1	*Philothamnus semivariegatus*	SAMPLE	N/D	N/D	SAMPLE
Hepatozoon sciuri	*Sciurus vulgaris*	MN104640	N/D	N/D	N/D
Hepatozoon sp.	*Sciurus vulgaris*	EF222259	N/D	N/D	N/D
Hepatozoon grisesisiciuri A	*Sciurus carolinensis*	MK452252	MK452388	MK452388	MK452388
Hepatozoon grisesisiciuri B	*Sciurus carolinensis*	MK452253	MK452389	MK452389	MK452389
Hepatozoon sp.	*Clethrionomys glareolus*	AY600625	N/D	N/D	N/D
Hepatozoon sp.	*Abrothrix olivaceus*	FJ719815	N/D	N/D	N/D
Hepatozoon ayorgbor	*Python reguis*	EF157822	N/D	N/D	N/D
Hepatozoon sp.	*Elaphe carinata*	KF939620	N/D	N/D	N/D
Hepatozoon ophisauri	*Pseudopus apodus*	MN723845	N/D	N/D	N/D
Hepatozoon sp. - 27A1	*Naja annulifera*	SAMPLE	SAMPLE	SAMPLE	SAMPLE
Hepatozoon sp. - 16B1	*Dipsadoboa aulica*	SAMPLE	SAMPLE	SAMPLE	SAMPLE
Hepatozoon colubri	*Zamenis longissimus*	MN723844	N/D	N/D	N/D
Hepatozoon sp. - 19F1	*Naja mossambica*	SAMPLE	N/D	N/D	SAMPLE

APPENDIX

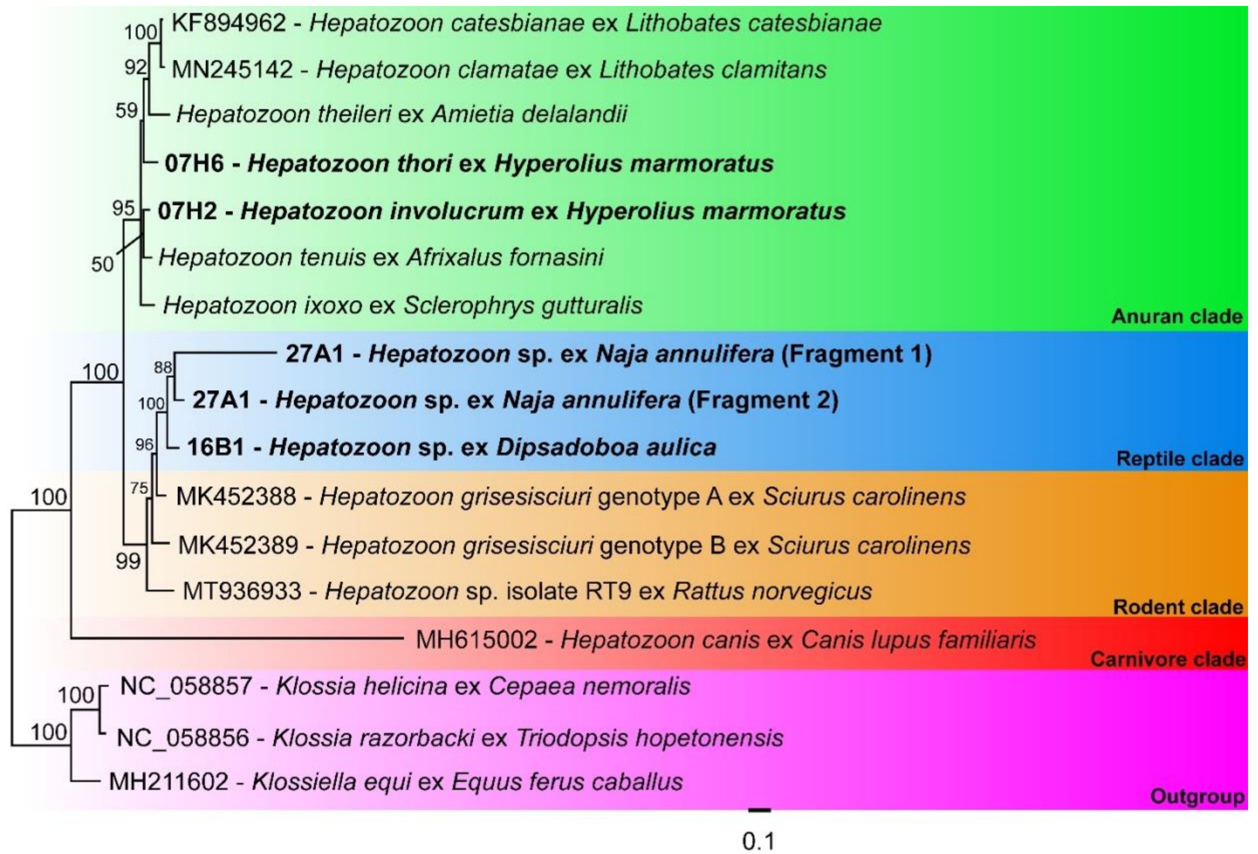
Supp. Table 3: Continued.

Taxon		Accession #			
Hepatozoon sp. - 18B1	Stigmochelys paradlis	SAMPLE	N/D	N/D	SAMPLE
Hepatozoon sp. - 20A1	Kinixys zombensis	SAMPLE	N/D	N/D	SAMPLE
Hepatozoon fitzsimonsi	*Kinixys zombensis*	KR069084	N/D	N/D	N/D
Hepatozoon banethi	*Ixodes tasmani*	MG758137	N/D	N/D	N/D
Hepatozoon sp.	*Dromiciops gliroides*	FJ719813	N/D	N/D	N/D
Hepatozoon ewingi	*Haemaphysalis bancrofti*	MG593275	N/D	N/D	N/D
Hepatozoon sp.	*Oceanodroma Melania*	KF022102	N/D	N/D	N/D
Hepatozoon americanum	*Canis spp.*	AF176836	N/D	N/D	N/D
Hepatozoon silvestris	*Felis silvestris silvestris*	KX757032	N/D	N/D	N/D
Hepatozoon felis	*Panthera leo persica*	KX017290	N/D	N/D	N/D
Hepatozoon martis	*Martes foina*	MG136688	N/D	N/D	N/D
Hepatozoon sp.	*Podarcis hispanica*	JX531909	N/D	N/D	N/D
Hepatozoon sp.	*Podarcis bocagei*	JX531928	N/D	N/D	N/D
Hepatozoon canis	*Pseudalopex gymnocercus*	AY471615	N/D	N/D	N/D
Hepatozoon canis	*Canis lupus familiaris*	N/D (AY471615*)	MH615002	MK214282	MK214283, MK214284, MK214285
Haemogregarina sacaliae	*Sacalia quadriocellata*	KM887507	N/D	N/D	N/D
Adelina dimidiata	*Scolopendra cingulate*	DQ096835	N/D	N/D	N/D
Dactylosoma ranarum	*Rana esculenta*	HQ224957	N/D	N/D	N/D
Klossia helicina	*Cepaea nemoralis*	HQ224956	NC_058856	NC_058856	NC_058856
Klossia razorback	*Triodopsis hopetonensis*	MT094864	NC_058857	NC_058857	NC_058857
Klossiella equi	*Equus ferus caballus*	MH203050	MH211602	MH211602	MH211602

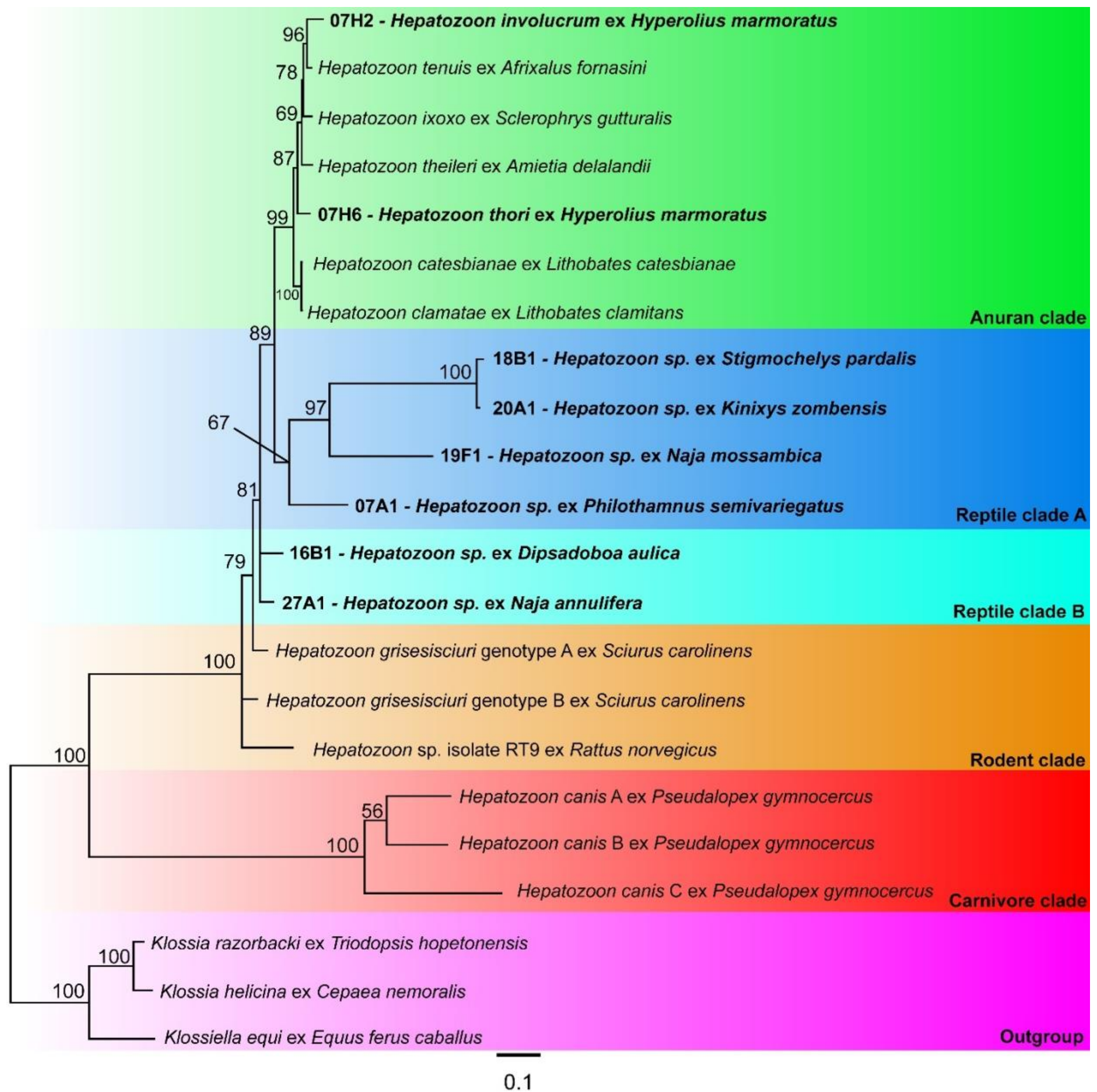
Description: The table lists the genetic sequences used for each taxon in the phylogenetic trees constructed.
N/D – No Data
SAMPLE – Sequences generated from samples in this study and from unpublished data.
*- The AY471615 18S sequence was used for phylogenetic analysis with COI, COIII, and CytB genes from a *H. canis* parasitizing *Canis lupus familiaris*.



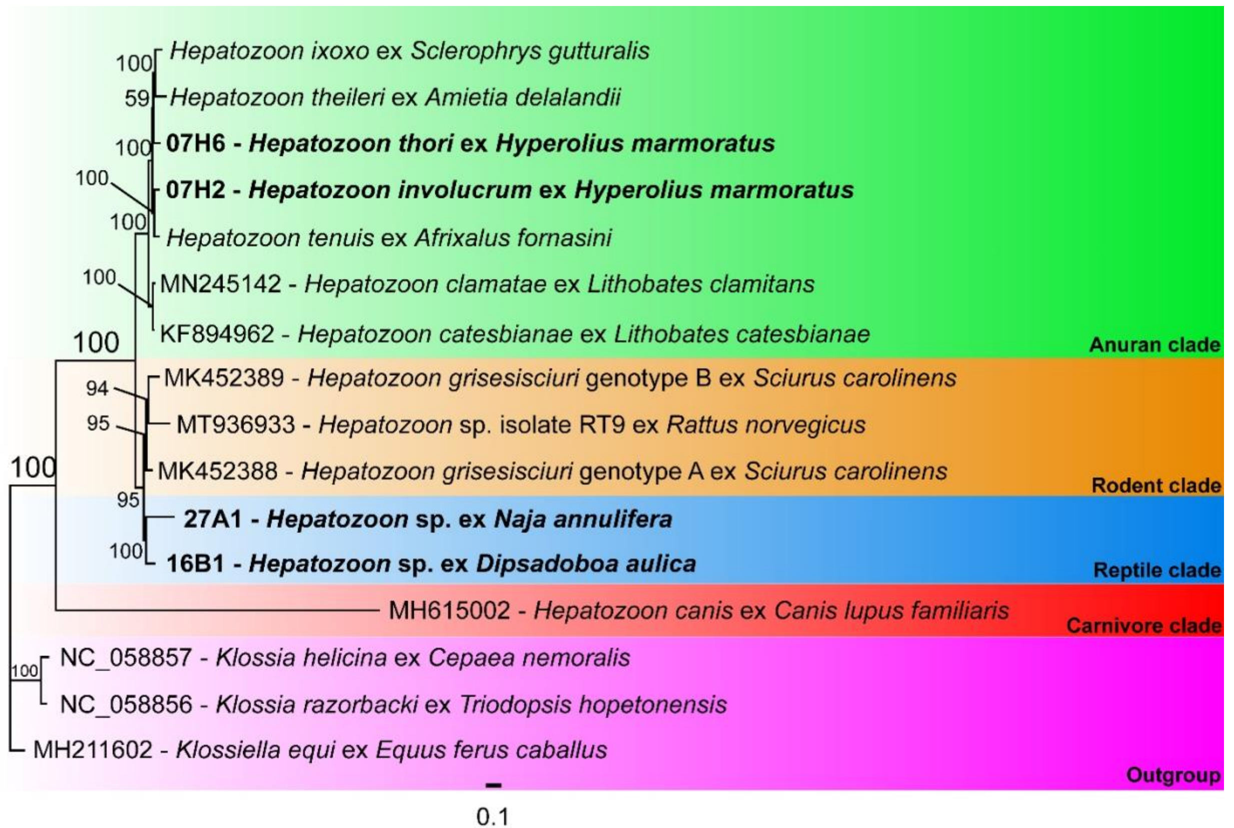
Supp. Figure 1: The COI phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade alongside *Hepatozoon* spp. parasitising rodent hosts. The marker places taxa in the correct order according to the 18S tree. The marker is able to resolve the monophyly of nodes deep within a phylogenetic tree of *Hepatozoon* spp. however posterior probabilities of nodes are low.



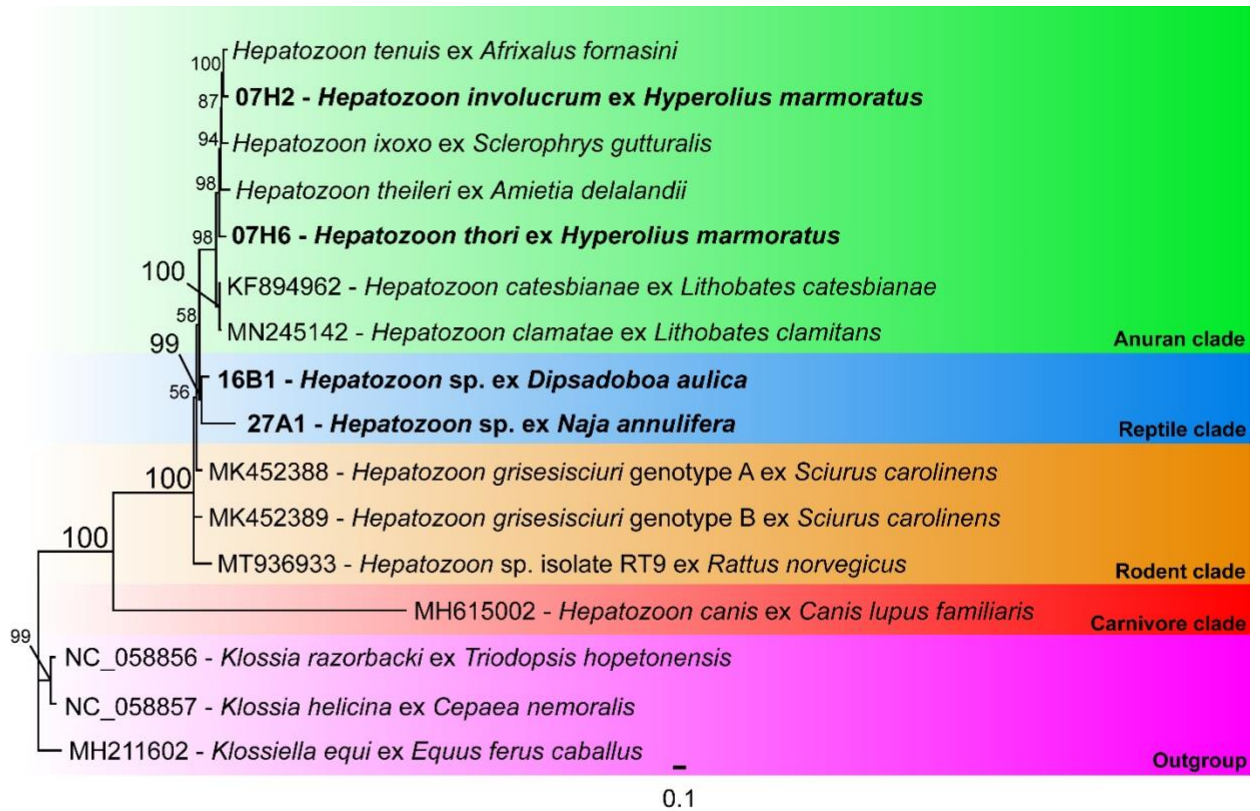
Supp. Figure 2: The COIII phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade alongside *Hepatozoon* spp. parasitising rodent hosts. The marker does not place taxa in the correct order according to the 18S tree. Two fragment sequences were generated for isolate 27A1 which are placed together with high posterior probability.



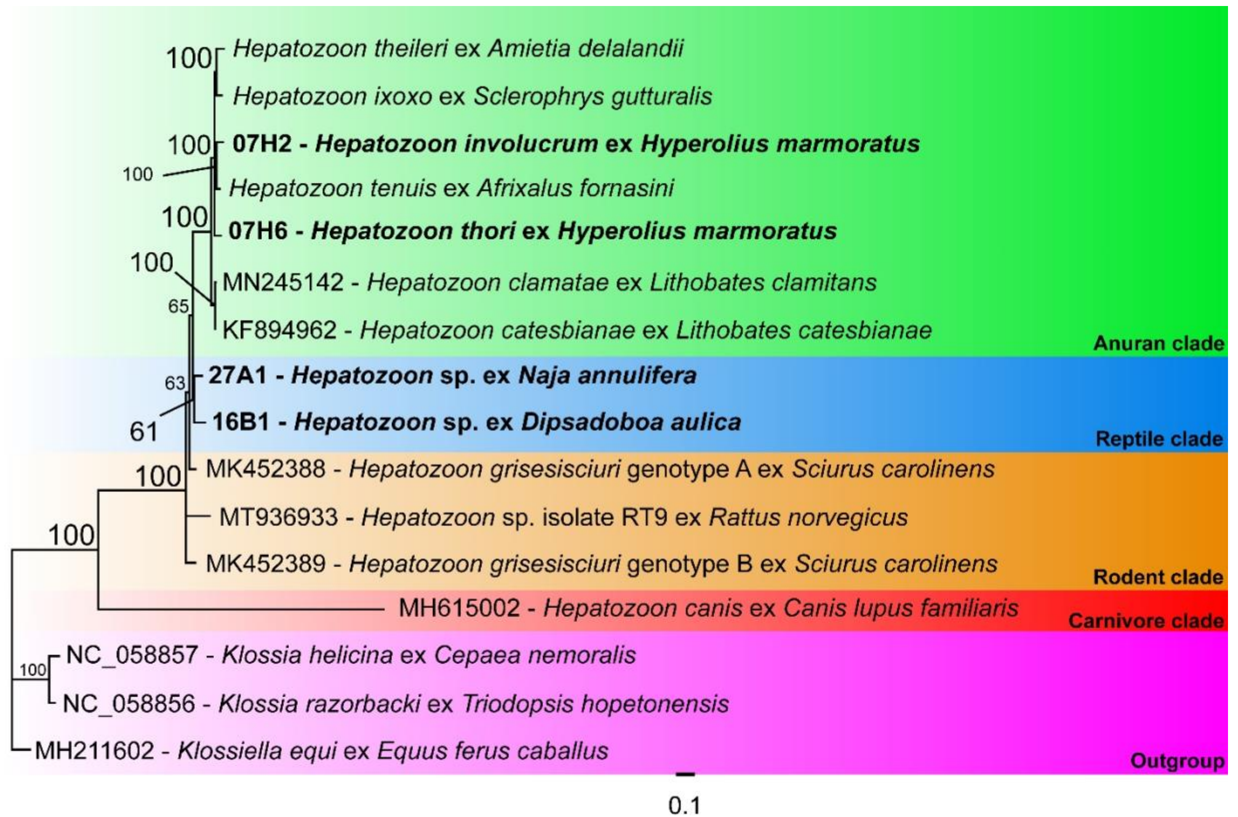
Supp. Figure 3: The CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 18B1, 20A1, 19F1 and 07A1 clade together as a monophyletic clade sister to the anuran clade. Isolates 16B1 and 27A1 clade together separate from other reptile taxa. The marker does not place taxa in the correct order according to the 18S tree. The marker is not able to reflect an accurate representation of recent speciation events.



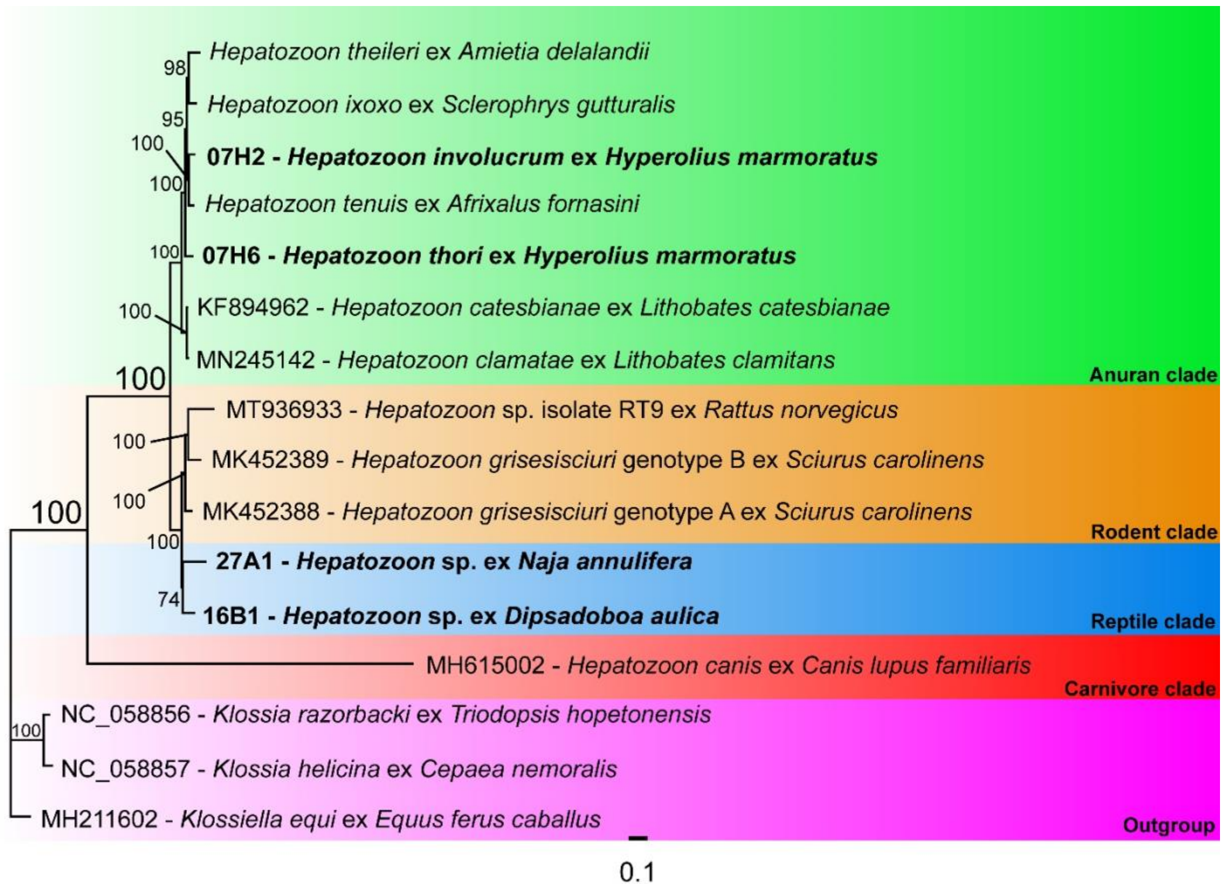
Supp. Figure 4: The COI-COIII phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade alongside *Hepatozoon* spp. parasitising rodent hosts. The marker combination places taxa in the correct order according to the 18S tree. The marker combination is able to resolve the monophyly of nodes deep within a phylogenetic tree of *Hepatozoon* spp. and posterior probabilities of nodes are high.



Supp. Figure 5: The COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic reptile clade close to *Hepatozoon* spp. parasitising anuran hosts. The marker combination does not place taxa in the correct order according to the 18S tree.



Supp. Figure 6: The COI-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade close to *Hepatozoon* spp. parasitising anuran hosts. The marker combination does not place taxa in the correct order according to the 18S tree.



Supp. Figure 7: The COI-COIII-CytB phylogenetic tree of species of *Hepatozoon* represented in this study. Isolates 07H2 and 07H6 is correctly placed alongside other *Hepatozoon* spp. parasitising anuran hosts. Isolates 27A1 and 16B1 clade together as a monophyletic clade alongside *Hepatozoon* spp. parasitising rodent hosts. The marker combination places taxa in the correct order according to the 18S tree. The marker combination is able to resolve the monophyly of nodes deep within a phylogenetic tree of *Hepatozoon* spp. and posterior probabilities of nodes are very high.