

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



Ciências
ULisboa

**Dating evolution on island ecosystems: a case-study with the
Cape Verde terrestrial biodiversity**

Ana Rita Leal Pena

Mestrado em Bioinformática e Biologia Computacional
Bioinformática

Dissertação orientada por:
Professora Doutora Maria Manuel Romeiras
Professor Doutor Octávio Salgueiro Paulo

ACKNOWLEDGEMENTS | AGRADECIMENTOS

First, I would like to express my sincere gratitude to both of my advisors Prof. Doutora Maria Romeiras and Prof. Doutor Octávio Paulo, for the continuous support, patience, motivation, and immense knowledge. The amount of work that went into this thesis wouldn't at all be possible if not for their help and intelligent advice.

I would also like to express my sincere thanks to Doutora Raquel Vasconcelos (CIBIO/Universidade do Porto), for her support in reviewing and correcting many of the results, and for the great suggestions to help improve them.

A special thanks to Tiago Menezes for his helpful advice on running and improving the analysis with *BEAST2.

And of course, a special thanks to my family and friends, in particular to my sister, for the continuous motivation and support to finish this massive project.

GENERAL CONTENTS

Resumo	iii
Abstract	vi
Acronyms and Abbreviations	vii
Figures List	viii
Tables List	ix
1. INTRODUCTION	
1.1. Bioinformatic tools for evolutionary biology.....	1
1.1.1. Main software packages.....	4
1.2. Macaronesia Islands as model systems in evolution.....	6
1.3. The study system – value of Cape Verde’s biodiversity for study evolution.....	7
1.4. Aims of the study.....	11
2. MATERIALS AND METHODS	
2.1. Data Collection.....	12
2.2. Data Analysis.....	13
2.2.1. Phylogenetic analyses.....	14
2.2.2. Divergence time estimation.....	15
2.2.3. Phylogenetic Networks.....	19
3. RESULTS	
3.1. Phylogenetic analyses.....	20
3.1.1 Separate mtDNA and nrDNA gene regions.....	20
3.1.2. Combined data.....	23
3.2. Divergence time estimation.....	29
3.3. Phylogenetic Networks.....	35
3.4. Reptiles richness and endemism.....	38
4. DISCUSSION	
4.1. Using Bioinformatic tools to study evolution on island species	40
4.2. Phylogenetic relationships among Cape Verde reptiles	41
4.3. Patterns of islands diversity and diversification hypotheses	43
5. FINAL REMARKS AND PERSPECTIVES	45
6. REFERENCES	46
Supporting Information	53

RESUMO

A área da bioinformática nasceu da necessidade de desenvolver e utilizar metodologias capazes de estudar grandes volumes de dados biológicos, em particular sequências de DNA ou proteínas. Nos últimos anos diversas áreas científicas beneficiaram da introdução de ferramentas computacionais especializadas em tarefas como a análise de dados e interpretação de resultados biológicos. Na área da Biologia Evolutiva o desenvolvimento do conhecimento foi substancial desde a introdução de métodos bioinformáticos, em especial nas análises filogenéticas e datação da origem e diversificação das espécies. Atualmente é possível analisar em larga escala sequências de DNA, recorrendo a vários métodos que permitem construir filogenias informativas e bem suportadas. Particular interesse é colocado na origem da grande diversidade de espécies endêmicas existentes nas ilhas oceânicas, e neste âmbito, refira-se que desde Darwin, que os ecossistemas insulares são considerados verdadeiros laboratórios naturais para estudos evolutivos.

A região da Macaronésia (i.e. Açores, Madeira e Selvagens, Canárias e Cabo Verde) constitui um hotspot de Biodiversidade, dentro da bacia do Mediterrâneo, sendo que, Cabo Verde é o único arquipélago situado geograficamente numa região tropical. As 10 ilhas do arquipélago de Cabo Verde apresentam uma grande diversidade de espécies terrestres e marinhas, muito delas endêmicas. De entre os principais grupos da fauna terrestre, refira-se os répteis, que dentro da Região da Macaronésia constituem a maior radiação nas ilhas de Cabo Verde. Neste arquipélago ocorrem três géneros que incluem 22 espécies endêmicas (13 das quais restritas a uma única ilha), nomeadamente: i) *Chioninia*, que inclui as espécies de escincídeos de Cabo Verde; ii) *Hemidactylus*, que apresenta uma grande diversidade e distribuição por diversas partes do Mundo; e iii) *Tarentola*, que para além de ser encontrado no arquipélago de Cabo Verde, ocorre também nas Canárias e Selvagens, dentro da Região da Macaronésia.

O elevado número de endemismos e habitats ricos com uma elevada diversidade de espécies, bem como existência de um grande número de sequências de DNA, pretendeu-se com este estudo utilizar diferentes ferramentas bioinformáticas de modo a explorar a relações evolutivas dos répteis nas ilhas de Cabo Verde. Também foram usados diferentes métodos de datação que permitiram de forma robusta e inovadora, relacionar o tempo de divergência entre diferentes espécies, contribuindo para compreender eventos evolutivos e padrões de colonização na Região da Macaronésia.

A metodologia escolhida teve por base a consulta exaustiva de bases de dados públicas, como o GenBank, que permitiu obter sequências de DNA para os répteis de Cabo Verde, bem como de possíveis espécies ancestrais. As sequências, de DNA mitocondrial e DNA nuclear, foram descarregadas em formato fasta, e utilizando o programa Geneious, foi possível efetuar revisões e correções nas mesmas. O alinhamento foi também feito com recurso a esta ferramenta, que disponibiliza vários pacotes acessórios, em concreto o programa de alinhamento MAFFT. Outros três programas de domínio geral, BioEdit, DNASp e MEGA foram utilizados para verificar a qualidade das sequências e dos alinhamentos. O programa DNASp foi também utilizado para correr análises de polimorfismo e haplótipos, que foram depois úteis para análises posteriores.

Utilizando o programa MEGA, construímos árvores filogenéticas utilizando métodos simples, como a Máxima Parcimónia, que permitiu verificar possíveis erros ou problemas nos dados. Foi encontrado um elevado número de sequências, em particular para as espécies endémicas em Cabo Verde, sendo que o número de sequências de genes de regiões do DNA mitocondrial, era muito superior em relação aos genes do DNA nuclear. No total foram recolhidas 1725 sequências de DNA (1447 originárias de Cabo Verde), provenientes de 15 genes diferentes e um total de 67 espécies de répteis. De modo a minimizar possíveis erros aquando a construção das árvores, foi realizada uma análise preliminar para selecionar as melhores sequências. Esta análise teve em conta a sua representatividade para cada espécie e gene (i.e., se o indivíduo a qual pertencia tinha sequências disponíveis para outros genes), o alinhamento obtido e por fim as árvores criadas com estes dados. Deste modo reduziu-se a redundância dos dados e o tempo de processamento das várias análises, bem como a qualidade dos resultados obtidos. Para tal foram realizadas várias análises filogenéticas para cada um dos genes, bem como criados ficheiros concatenados, utilizando o programa Concatenator, dos genes mitocondriais, nucleares e ambos, para analisar os resultados em conjunto. Para avaliar os modelos e partições associados aos nossos dados, foram feitas análises com o programa PartitionFinder. Os modelos e partições obtidos nestas análises foram depois introduzidos em três ferramentas de construção de árvores filogenéticas - RAxML, MrBayes e BEAST (incluindo a versão mais recente e o pacote *BEAST2) – que são atualmente os programas mais usados no estabelecimento das relações filogenéticas. O primeiro programa (i.e. RaxML), utiliza o método de máxima verosimilhança; complementarmente foi utilizando um outro programa – RAxMLGUI – que facilitou a conexão direta para os servidores da plataforma Cipres Gateway. Esta plataforma online, providencia servidores para diversas ferramentas de análise filogenética, tendo sido usada para correr as nossas análises com o RAxML, MrBayes e BEAST. Deste modo, foi possível reduzir substancialmente o tempo de processamento dos programas, bem como providenciar um local secundário para guardar alguns dos dados. Para testar os resultados obtidos com Inferência Bayesiana, utilizamos o programa MrBayes para as análises “clássicas” e a ferramenta BEAST na sua versão base e utilizando a mais recente versão do pacote *BEAST2, para fazer datações nas análises. Os resultados obtidos com estes programas foram verificados com a ferramenta Tracer, para assegurar a convergência dos vários parâmetros durante as várias corridas realizadas. Para além destas análises, pretendemos testar possíveis eventos de recombinação e hibridização genética, tendo sido para o efeito construídas redes filogenéticas para cada um dos géneros utilizando vários programas, como o SplitsTree4. Os resultados obtidos explorando diferentes ferramentas bioinformáticas, permitiram construir um grande número de árvores filogenéticas, reveladoras das relações evolutivas entre várias espécies de répteis das ilhas da Macaronésia. De uma forma geral a topologia obtida foi semelhante à obtida em estudos anteriores, com resultados ligeiramente melhores nas análises que usaram Inferência Bayesiana. As redes filogenéticas criadas apresentaram um reticulado que parece indicar que terão de facto ocorrido a presença de eventos de hibridização e recombinação nas linhagens de Cabo Verde nos três géneros analisados, sendo necessárias análises complementares para clarificar estes resultados.

Os resultados sugerem diferentes origens para cada género, e revelaram diferentes padrões e eventos de colonização nas ilhas de Cabo Verde. O género *Chioninia* terá tido origem em linhagens africanas, apresentando uma maior semelhança ao género *Trachylepis*, oriundo de África com exceção de um espécime do Brasil. As espécies de *Hemidactylus* endémicas em Cabo Verde apresentam uma relação com espécies de São Tomé e Príncipe (África), Trindade e Brasil (América Latina), o que põe em evidência a grande capacidade de dispersão deste género, inclusive através do Oceano Atlântico. Por fim, as espécies de *Tarentola* endémicas em Cabo Verde, parecem ter evoluído de ancestrais que colonizaram outras ilhas da Macaronésia (i.e. Canárias e Selvagens) e que diversificaram no Norte de África, com um possível ancestral na América Central. A aplicação do relógio molecular usando métodos Bayesianos (MrBayes, BEAST e *BEAST2) permitiu testar hipóteses evolutivas, sugerindo que o género *Chioninia* terá sido o primeiro a colonizar Cabo Verde, a partir de ancestrais da região Centro Africana, tendo colonizado este arquipélago há cerca de 16,7 milhões de anos. As espécies de *Hemidactylus* terão colonizado as ilhas de Cabo Verde a partir de espécies ancestrais africanas que diversificaram nas ilhas entre 2,4 – 11,5 milhões de anos. Por fim, o género *Tarentola* poderá ter tido origem no final da época do Mioceno (5,8 – 7,5 milhões de anos), com uma grande diversificação no arquipélago entre 5,1 – 5,8 milhões de anos, a partir de ancestrais das ilhas das Canárias e do Norte de África.

Os resultados obtidos no presente estudo oferecem uma nova compreensão sobre a história evolutiva de um dos maiores grupos da fauna terrestre das ilhas da Macaronésia – os répteis, que apresentam o maior centro de diversidade nas ilhas de Cabo Verde. Com o elevado número de métodos, programas e dados utilizados, foi possível construir árvores filogenéticas e aplicar técnicas de datação molecular inovadoras, como é o caso do *BEAST2, para os répteis de Cabo Verde. Contudo, para além dos tempos de divergência das diferentes espécies, podia-se ter explorado as áreas de distribuição ancestral e com base nos resultados dessas análises propostos cenários biogeográficos, para além dos evolutivos agora apresentados.

Por fim, refira-se que os ecossistemas insulares, reconhecidos como laboratórios naturais, mostraram mais uma vez serem são modelos únicos para testar métodos e programas bioinformáticos, evidenciando a grande utilidade da existência de programas e tutoriais disponíveis online e gratuitamente, permitindo ter uma base muito bem suportada para explorar qualquer análise na área da Biologia Evolutiva.

Palavras chave: Ferramentas Bioinformáticas, Evolução, Filogenética, Macaronésia, Répteis

ABSTRACT

The interdisciplinary field of bioinformatics was born from the need to develop and employ methodologies to study large amounts of biological data, namely proteins and DNA. Many science areas benefited greatly from the introduction of specialized and powerful tools that improved the implementation of databases, data analysis and biological interpretations. For evolutionary study fields such as phylogenetics, the knowledge jump has been substantial since the introduction of bioinformatic methods in their analysis. Making use of highly advanced software packages is now possible to perform high-throughput DNA sequencing, and thoroughly analyse this data with several methods of phylogenetic inference, to obtain informative and well supported phylogenies. Even though many important evolutionary discoveries were made in the last few years, understanding the evolutionary origin of biological diversity on island ecosystems is still a critical issue within the hotspot area of the Macaronesia Islands (i.e. Azores, Canary Islands, Madeira and Cape Verde). One of the least studied archipelagos is Cape Verde, which houses a huge diversity of endemic species whose evolution is still largely understudied. In these islands, the reptiles are one of the most diverse radiations among the terrestrial groups, hosting three genera with 22 endemic species. *Chioninia*, a group of lizards formerly included in the genus *Mabuya*; *Hemidactylus*, a widespread and diverse group of house geckos; and *Tarentola*, a genus of wall geckos that also occurs on other Macaronesia archipelagos. Some of these species are single islands endemics (SIE) and are threatened species, which puts them at high risk of extinction. This study aims to reconstruct a phylogeny for this group with large-scale taxon sampling using a wide sample of the most current bioinformatic tools available. DNA data was retrieved from the Genbank and subsequently analysed using several software packages for alignment, model fitting and finally phylogenetic inference. The collected data was composed of 1725 DNA sequences (1447 coming from Cape Verde) and 15 different genes, which originated from a total of 67 reptilian species. Both approaches – Maximum Likelihood and Bayesian Inference – were used to achieve more thorough results. Divergence time estimation was also performed using a classical BEAST and *BEAST2 across the three reptile groups to assess their evolution story in the Cape Verde Islands. Phylogenetic networks were also created to assess possible hybridization or recombination events in the Cape Verdean reptiles. This macroevolutionary perspective also combines data from ecology and distribution of the species to an integrated vision of Cape Verde evolution of terrestrial biodiversity. The results suggest different origins for each genus and different colonization patterns and events along the Cape Verde Islands, ranging from the Macaronesia region to Africa and South America: i) *Chioninia* originated 21.7 – 39.6 Mya, radiating along the islands 14.9 – 18.5 Mya, with a probable colonization from central African regions; ii) the *Hemidactylus* species having colonized the archipelago 2.4 – 11.5 Mya and originating from African ancestral species; and iii) the endemic Cape Verdean *Tarentola* are likely descendants of Northern African and Canary Islands lineages, that colonized the islands during the end of Miocene (5.8 – 7.5 Mya) and later diversified (5.1 – 5.8 Mya). In conclusion, presently the huge amount of DNA sequences available in international databases and the most up-to-date bioinformatic tools and methodologies, allow us to contribute to better understand the origin and colonization of insular ecosystems, using insular endemic species, as an excellent case-study in the field of Evolutionary Biology.

Key-words: Bioinformatic Tools, Evolution, Macaronesia, Phylogenetics, Reptiles

ACRONYMS AND ABBREVIATIONS

- ACM4** - Acetylcholinergic Receptor M4
- AICc** - corrected Akaike Information Criterion
- BEAST** - Bayesian Evolutionary Analysis Sampling Trees
- *BEAST** - StarBEAST
- BI** - Bayesian Inference
- C-mos** - Oocyte maturation factor Mos
- COI** - Cytochrome c oxidase subunit I
- cyt b** - Cytochrome B
- DDBJ** - DNA Data Bank of Japan
- DNA** - Deoxyribonucleic Acid
- EMBL** - European Molecular Biology Laboratory
- ESS** – Effective Sample Size
- GIGO** - Garbage In, Garbage Out
- HPD** - Highest Probability Density
- IUCN** - International Union for Conservation of Nature
- MC1R** - Melanocortin 1 Receptor
- ML** - Maximum-Likelihood
- mtDNA** - Mitochondrial DNA
- Mya** – Millions of years ago
- nDNA** - Nuclear DNA
- NGS** – Next Generation Sequencing
- NJ** - Neighbour Joining
- PDC** - Phosducin
- RAG2** - Recombination activating gene 2
- RAxML** - Randomized Axelerated Maximum Likelihood
- SIE** - Single Islands Endemics

FIGURES LIST

Figure 1.1: Map of the Macaronesian.

Figure 1.2: Map of the Cape Verde Islands.

Figure 2.1: Diagram of the workflow performed along this study.

Figure 3.1: Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Chioninia*.

Figure 3.2: Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Hemidactylus*.

Figure 3.3: Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Tarentola*.

Figure 3.4: Results from BEAST for the *Chioninia* genus with divergence times estimated using all the genes.

Figure 3.5: Results from BEAST for the *Hemidactylus* genus with divergence times estimated using all the genes.

Figure 3.6: Results from BEAST for the *Tarentola* genus with divergence times estimated using all the genes.

Figure 3.7: Results from *BEAST2 for the *Chioninia* genus with divergence times estimated using all the genes.

Figure 3.8: Results from *BEAST2 for the *Hemidactylus* genus with divergence times estimated using all the genes.

Figure 3.9: Results from *BEAST2 for the *Tarentola* genus with divergence times estimated using all the genes.

Figure 3.10: Phylogenetic networks created for *Hemidactylus*.

Figure 3.11: Phylogenetic networks created for *Chioninia*.

Figure 3.12: Phylogenetic networks created for *Tarentola* without the Azorean specimen.

Figure 3.13: Phylogenetic networks created for *Tarentola* with the Azorean specimen.

Figure 3.14: The reptile diversity in the Macaronesia region (excluding subspecies).

TABLES LIST

Table 1.1: Summary of the software packages used in this study, competing programs and the advantages we found in their use.

Table 1.2. List of the reptile species and subspecies found in the Cape Verde archipelago

Table 2.1: Summary of the partitions obtained in PartitionFinder in the final version of the analysis in both sets of analysis.

Table 2.2: Fossil calibrations used for estimating divergence dates in BEAST (adapted from Zheng & Wiens 2016).

Table 2.3 Calibrations points used in the *BEAST2 analysis for all three groups.

Table 3.1: Summary of the polymorphism data of the final set of sequences used in our analysis for each of the genera, organized by their source.

1. INTRODUCTION

1.1 Bioinformatic tools for evolutionary biology

Since the introduction of computational analysis, the number of available tools for science research has grown exponentially and as the complexity of the software's evolved, so did the necessity for specialised researchers to use them. The field of bioinformatics was born of the need for a field that focused in analysing a high volume of molecular data, such as DNA or proteins, using the more current tools available, and has since then evolved itself in more resourceful and extensive study field (Shenbagarathai 2007).

We can categorize bioinformatics in more specialized areas, specifically the development and implementation of databases; data analysis and data mining; and interpretation of biological data (Moore 2007). These days most biological studies have some type of bioinformatic analysis, and it's fairly common to make use of multiple options.

Starting in the collection and storage of data we have today numerous databases around the world available for this effect. One of such databanks is GenBank (Clark et al. 2016), which storages and organizes nucleotide sequences while also making them publicly accessible and ready to use. Other biology focused databases include EMBL (Kulikova et al. 2007), UniProt (EMBL et al. 2013), DDBJ (Kodama et al. 2012) and many others, all of them free and available online, allowing an ever growing global network of information that continues to grow, giving researchers an almost never-ending flow of new data.

One of the other issues that rose from the high output of data was its processing. Starting in sequencing, the past 30 years have been astonishing, with more than a million-fold improvement in the rate of sequence generation (Stratton et al. 2009), particularly with the introduction of Next generation sequencing (NGS). With such an extraordinary quantity of sequences being created, the alignment algorithms also grew and adapted.

A variety of alignment software's exist today, which with its algorithm and diverge functionalities proved invaluable for processing the amount of sequenced data. Some of these include MAFFT (Katoh & Standley 2013), BLAST (Altschul et al. 1990), MUSCLE (Edgar 2004) and Clustal (Larkin et al. 2007), which have undergone countless improvements over time and are publicly available online and through other software's as a built-in option.

One of the fields that took the most benefit of working with bioinformatic tools was phylogenetics, which uses molecular data, such as DNA sequences, to build evolutionary relationships between species. To create phylogenies, we start by retrieving and aligning molecular sequences using the resources already mentioned, before moving to the evaluation of the models associated to this molecular data. The two most commonly used software packages used for model calculation as of today are jModelTest2 (Posada 2008) and PartitionFinder2 (Lanfear et al. 2016). Both are capable of simple, yet resourceful analysis of molecular sequences, which is then employed as parameters for other biological analysis software's. PartitionFinder allows an extra layer of examination, as it also finds the most adequate partitions for the given sequences.

After carefully retrieving, aligning and examine the models in our molecular sequences, we can finally work on the construction and evaluation of possible phylogenetic trees. As in all the other parts of the analysis, there's an extraordinary amount of different solutions to this problem, which have also been changing and adapting along the years. The more simple algorithms for tree creation are UPGMA (Sokal & Michener 1958), Neighbour-Joining (Saitou & Nei 1987), and Maximum-Parsimony (Farris 1970; Fitch 1971), which are available in general purpose software such as MEGA (Kumar et al. 2016) and PAUP (Swofford 2002) as a quick and simple way to obtain a tree. However, there are today other, more sophisticated solutions, which have replaced these options almost completely in current scientific studies in this field.

Methods for Phylogenetic Inference

Different approaches and software's have been developed along the years to perform phylogeny estimation, with researchers arguing on which was the best option at the time. Nowadays, the two main opposite ideas challenge each other on how to analyse a multiple data set (i.e. with several different gene samples). The more conventional approach, commonly referred as the Concatenation method, involves combining the gene samples found for a given group of species, which strengthens the analyses due to the higher number of variable sites. The main setback with this method is of course, when the multiple loci don't share a similar topology. The alternative multi-species coalescent method, tries to solve this issue by estimating gene and species topologies simultaneously, which takes in account possible incongruences in the trees, such as agene duplication, horizontal deep transfer or deep coalescence (Edwards 2009).

With such opposite ideas, it can be challenging for a study to settle for one of these approaches, as the optimal solution could only be reached by knowing which was better suited to the data before it's even collected (Carstens 2013). To avoid such issues, it's not uncommon for studies to employ both solutions as it allows for a more thorough analysis, and an objective comparison between the methods without choosing a particular side. Since one of the main objectives of our study was to test a wide group of phylogenetic inference software's, the latter was the most natural and appropriate solution.

To perform "classical" concatenation analysis two main software packages are generally used: RAxML (Stamatakis 2014) and MrBayes (Huelsenbeck & Ronquist 2001), which also differentiate themselves on the method to obtain the best phylogenetic solution.

Randomized Axelerated Maximum Likelihood (RAxML) makes use of Maximum Likelihood estimation (ML), which was made popular by Fisher in the early-20th century (Aldrich 1997), to perform phylogenetic inference. RAxML adapts the algorithm developed by Felsenstein (Felsenstein 1981), which was the first to compute phylogenetic trees given a set of DNA data, and works by selecting the tree that has the highest probability of generating the observed data. Several optimizations and updates were built-in the program, which makes it extremely fast and accurate, and capable of analysing big volumes of data. Similar software's such as PHYML (Guindon et al. 2009) and GARLI (Zwickl 2006), also make use of ML calculations, and present alternative options which are also capable of fast and accurate analysis.

On the other hand, MrBayes estimates phylogenies making use of Bayesian Inference (BI), which has been around since Thomas Bayes' essay in 1763 as a special case of the Bayes Theorem (Fienberg 2006). This methodology raised in popularity in the late 1980's with the introduction of Markov Monte Carlo chains (MCMC), which reduced computational demands drastically, and finally allowed for a general use of BI methods in range of scientific applications. This approach differs from ML, by considering the parameters in the model to be random, instead of fixed variables, and so prior distribution need to be assigned to the data to obtain posterior distributions, which are used to perform the needed inferences (Yang & Rannala 2012). This approach can also be used to estimate molecular clocks (including relaxed clocks), by performing divergence time estimations with mutation rates, fossil and geological as calibration data, as seen in the software package BEAST (Drummond & Rambaut 2007).

Divergence Time Estimation – Molecular clocks

The selection of an appropriate clock model is another important step when performing a divergence time estimation in evolution. The simplest approach, a strict-clock, assumes that the evolution rates are fixed along the branches, and even though it performs well when low rate variations are expected, it's usually not adequate for most of the analysis. To correct this issue, particularly when using multilocus data, two other clock methods were introduced, namely correlated and uncorrelated relaxed clocks. Both approaches allow different substitution rates along the branches, but differ on their assumptions of the rate variation along them. As such, a correlated clock assume that neighbouring branches have similar evolution rates, while in an uncorrelated clock all the branches are independent from each other. In either case, a distribution is necessary to model the branch rates (Ho & Duchêne 2014). Several options are available for this effect, though three of them are more commonly used: lognormal, gamma and exponential.

With the introduction of relaxed clock methodologies, it was finally possible to make use of calibration techniques, which allow external dating information to be taken account when making the calculations, instead of using molecular data only. Calibration can be made on two different points: on the terminal or on the internal nodes. Since terminal calibration is only possible when incorporating fossil data or when working with some virus, internal calibration is more generally used, as it has less limitations (Ho & Phillips 2009). To calibrate internal nodes there's two main forms available, specifically one that makes use of fossil and biogeographical (which requires some a priori knowledge of the relation between the species), and another that uses substitution rates previously calculated. Is not uncommon for studies to work with both approaches, as it usually provides a better supported analysis. To perform our analysis, we used several internal nodes calibrations, which were taken from previous works that worked with different sources of data and information. As such, this approach should give a more thorough and reliable results.

Due to the complexity of the divergence estimations methods, the choice of a good software to perform these analysis is absolutely paramount. The most popular software used in this type of analysis is BEAST, which makes use of Bayesian multispecies coalescent methods to calculate and date phylogenies. However, working with high volumes of data, namely multilocus data, is a challenge to this program, as it exponentially slows the analysis. To solve this setback, several improvements were made in the second version of the program, BEAST2 (Bouckaert et al. 2013), which allow for the use of different packages that are suited for several particular tasks.

One of such packages, StarBEAST2 or *BEAST2 (Ogilvie et al. 2017), works very well with multilocus data by providing reliable trees, divergence times and substitution rates, while being computationally faster than previous versions and pipelines of BEAST and *BEAST.

Phylogenetic Networks and Tree Visualization

A different method of phylogenetic inference, commonly called phylogenetic networks, is usually used to test the existence of cryptic events such as hybridization, recombination, horizontal gene transfer, or the duplication or loss of genes, in the given molecular data set. This method of analysis is implemented in several software packages like TCS (Clement et al. 2000) and SplitsTree4 (Huson & Bryant 2001), the latter containing methods for recombination and hybridization networks.

To manipulate and visualize the trees created in both ML and BI phylogenetic analysis we have several programs at our disposal such as FigTree (Rambaut 2009), iTOL3 (Letunic & Bork 2016) and Dendroscope3 (Huson & Scornavacca 2012), which range in both complexity and offered tools.

In fact, the number of bioinformatic tools available today represents a huge framework for each question that arises, presenting several alternative solutions and ways to reach them. For phylogenetic analysis, this can prove to be an extraordinary time, as new knowledge and powerful software and hardware utilities continue to improve every day.

1.1.1 Main Software Packages

As discussed above, there's a wide range of software programs available for the several steps in a phylogenetic study. Throughout the course of this study we tested and applied many of these methods and programs, some being chosen over others currently available, as its summarized in Table 1.1. across the analysis we used the most current version available at the time, and tested the given options to assess which gave us the best workflow and results.

Table 1.1. Summary of the software packages used in this study, similarly focused programs and the provided advantages we found in the several steps of the phylogenetic analysis. Most of these programs are open-source, with the exception of Geneious which offers a paid version with additional functionalities.

Software	Version	Specification	Similar Software's	Advantages
BEAST (Drummond & Rambaut 2007)	1.8.4.	Bayesian inference; molecular clock	r8S (Sanderson 2003)	- More consistent estimates given limited sampling of loci (Mulcahy et al. 2012); -More commonly used in current studies;
BEAUti (Drummond & Rambaut 2007)	1.8.4.	Associated program to BEAST	NA	-Necessary to work with Beast;
BioEdit (Hall 1999)	7.2.5	Biological sequence alignment editor and analysis	Geneious (Kearse et al. 2012)	- Simple to use; - Range of tools for analysis; - Helpful to review data;
Concatenator (Pina-Martins & Paulo 2008)	1.1.0	Concatenation and conversion for nexus and fasta files	Various multitasked programs have this functionality built-in	- Simple to use; - Better control of the concatenation of the files as it has an option to format the output to work with other programs (e.g. MrBayes and PAUP);
DNASp5 (Librado & Rozas 2009)	5.10	Analysis of DNA polymorphism data; Haplotype analysis and phasing; Other similar tasks	MEGA (Kumar et al. 2016); Mesquite (Maddison & Maddison 2008)	- Contains a variety of tools easy to use; - Simple way to calculate and verify haplotypes; - Quick analysis of polymorphism information and review;
FigTree (Rambaut 2009)	1.4.3	Tree plotting/drawing	Dendroscope3 (Huson & Scornavacca 2012); iTOL3 (Letunic & Bork 2016)	- Easy and intuitive interface; - Range of tools for manipulation of trees;
Geneious (Kearse et al. 2012)	10.2.3	General purpose program	MEGA; Mesquite	- Wide range of tools and available plugins for analysis. - Easy to use and manipulate
LogCombiner (Drummond & Rambaut 2007)	1.8.4	Associated program to BEAST	NA	-Necessary to work with Beast
MAFFT (Kato & Standley 2013)	1.3.6	Alignment algorithm	MUSCLE; Clustal; Blast	-Widely used and easy to work it; -Made available in Geneious
MEGA (Kumar et al. 2016)	7.0.18	Several tasks for sequence processing and reviewing	DNASp5; Mesquite	- Easy to work with; - Several of tools available for simple and quick analysis
MrBayes (Huelsenbeck & Ronquist 2001)	3.2.6	Bayesian inference	Various multitasked programs have built-in options for BI	- Most commonly used in current studies; - Easy and simple to use.
PartitionFinder (Lanfear et al. 2016)	2.1.1	Model and partition scheme calculation in for DNA and protein alignments	jModelTest2 (Posada 2008)	-Allows model analysis and partition analysis as well; - Easy to use
PAUP (Swofford 2002)	4.0a152	Phylogenetic analysis using parsimony and other methods	MEGA	- Allows a simple way to perform ILD tests;
RAxMLGUI (Silvestro & Michalak 2012) / RAxML (Stamatakis 2014)	1.5.b / 8.1.21	Maximum likelihood, simple Maximum parsimony	PHYML; PAML (Yang 2007)	- Most commonly used on current studies; - Easy to use and interact with
SplitsTree4 (Huson & Bryant 2006)	4.14.6	Phylogenetic network construction	TCS (Clement et al. 2000)	- Allows recombination and hybridization networks; - Easy to use;
Tracer (Rambaut & Drummond 2013)	1.6	Analysing the trace files generated BI Bayesian MCMC runs	NA	- Invaluable to assess the quality of BI results in BEAST and MrBayes;
TreeAnnotator (Drummond & Rambaut 2007)	1.8.4	Associated program to BEAST	NA	-Necessary to work with Beast
*BEAST2 (Ogilvie et al. 2017)	2.4.7	Software package for BEAST2 (Bouckaert et al. 2013)	Classical BEAST; BEAST2; *BEAST	- Computationally faster; - Allows the use of several loci

1.2. Macaronesia Islands as model systems in evolution

The importance of evolutionary processes on islands was first recognized in Charles Darwin's seminal studies on the Galapagos islands, which gave rise in 1859 to the theory of evolution by means of natural selection (Darwin 2009). The pattern of species accumulation on oceanic islands depends on the rate of evolution relative to the frequency of island colonization, the extreme cases leading to extensive adaptive radiation as a result of *in situ* evolution with associated adaptation to occupy the available ecological space (Gillespie & Clague 2009). Over the last decades, research has focused on multiple facets of islands and they had a relevant role in establishment of Ecological and Evolutionary theories (Bellard et al. 2014).

Islands can be broadly classified as one of the two types, continental or oceanic. Continental islands are fragments of a larger landmass, whereas oceanic islands result from volcanoes rising above the water on or near a mid-ocean ridge. Particularly, oceanic islands were the focus of studies that had contributed to our understanding of speciation and the origins of biodiversity, as is the case of Macaronesian Islands, which comprises five archipelagos: Azores, Madeira, Selvagens, Canaries and Cape Verde Islands (see Fig. 1.1).

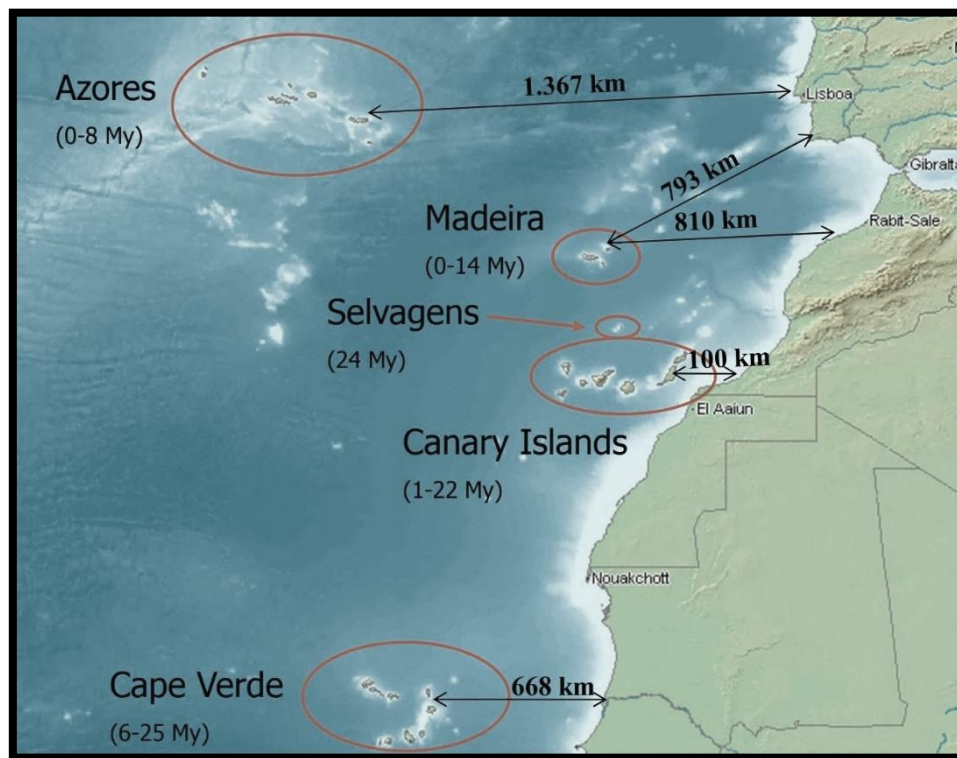


Figure 1.1. The five archipelagos of the Macaronesian region - Azores, Madeira, Selvagens, Cape Verde and Canary Islands – with the corresponding age and distance to the closet mainland point. (Adapted from <http://www.juancarlosillera.es>)

The Macaronesian Region, belongs to the Mediterranean Basin biodiversity hotspot, which is characterized by high levels of endemism. The relatively short distance to the African and European continents would have allowed for the colonization from this area with the influence of oceanic currents (North Atlantic and Canary streams) and north-easterly trade winds likely to have fostered the dissemination of ancestral species from the continent. For this reason, it has been proposed that frequent colonisations from the continent have contributed to the high number of species endemic to the Macaronesian Islands (Francisco-Ortega et al. 2000).

In general, terrestrial endemic lineages are often characterized by occurrence in different habitats; striking morphological differences among species; and their frequent rarity, being present in a few small populations (Takayama et al. 2015). The conservation of endemic island species is a complex, multifaceted topic, including the preservation of native habitats, control of alien plants and animals, and minimizing the impact of human activities (Caujapé-Castells et al. 2010; Romeiras et al. 2016). While these factors are very important from a conservation perspective, there is still a gap in our knowledge on how this huge endemic diversity has evolved in each archipelago, with lacking details for entire taxonomic groups. This information is needed to understand the evolution and diversification patterns of endemic diversity on the Macaronesian hotspot area, which begs the necessity for more complete studies that encompass these endemic lineages, while taking in account their taxonomic and regional relatives.

1.3. The study system – value of the Cape Verde biodiversity for study evolution

The Cape Verde Islands are the most southern archipelago of the Macaronesian Region. The archipelago is a group of 10 volcanic islands located 1500 km southwest of the Canary Islands and 668 km west of the African mainland. Within the archipelago, the islands form three clusters: (i) Northern group (Santo Antão, São Vicente, Santa Luzia and São Nicolau); (ii) Southern group (Santiago, Fogo and Brava) and (iii) Eastern group (Sal, Boavista and Maio) (see Fig. 1). The northern and the southern islands are characterized by high mountains [e.g. Monte Gordo (1304 m) in São Nicolau; Pico da Antónia (1392 m) in Santiago; Tope de Coroa (1979 m) in Santo Antão and Pico do Fogo (2829 m) in Fogo], offering a wide range of habitats over relatively short distances (Duarte & Romeiras 2009). The eastern islands are lower, drier and more homogeneous in their ecology. The islands' ages range from ~25.6 to ~21.1 Ma (for Sal and Maio, respectively) to <6 Ma (for Brava), with ages decreasing from east to west (Doucelance et al. 2003). Only Fogo Island currently has volcanic activity, with the most recent eruptions occurring in 1995 and 2014-2015.

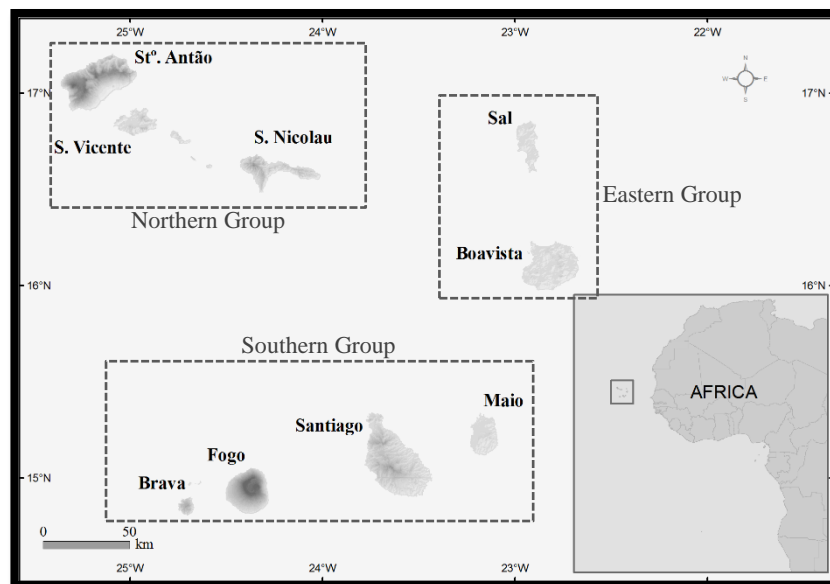


Figure 1.2. The Cape Verde Islands archipelago in detail, showing the geographical position and size of the islands. All nine inhabited islands are marked on the map, while the uninhabited Santa Luzia Island (not named in the map), is seen as the large land mass south of São Vicente and north of São Nicolau. The Branco and Raso islets are found south of Santa Luzia, while the Rombo islets are neighbours to island of Brava. Adapted from (Romeiras et al. 2015).

Cape Verde's biodiversity is of great scientific value, with several important marine radiations, as well as important terrestrial endemic groups of arthropods, plants and reptiles (Arechavaleta et al. 2005; Duda & Rolán 2005; Cunha et al. 2014). Among these groups its highlighted that several endemic species, some of which economically valuable, are at risk of extinction and make conserving the archipelago's biodiversity a world concern (Romeiras et al. 2016). The reptiles have three vast reptile genera – *Chioninia*, *Hemidactylus*, *Tarentola* – with 22 endemic species, which represent the huge diversity of this group in the Macaronesian region according to Vasconcelos et al. (2013).

To date, molecular analyses of the Cape Verde flora and fauna have been limited and an almost complete gap exists in several groups, even among most arthropods, that have the highest number of endemic species. The majority of the analyses made to study the relationships of Macaronesian lineages have focussed largely on Azores, Madeira and the Canary Islands (Roux 2004; Santos et al. 2010; Machín-Sánchez et al. 2014; Orellana et al. 2016), and where Cape Verdean taxa have been included, the sampling has often been limited (Pyron et al. 2013; Karin et al. 2016; Metallinou et al. 2016; Zheng & Wiens 2016). Nevertheless, taxonomic and molecular relationships for the Cape Verdean reptiles were recently well documented in several studies (Arnold et al. 2008; Vasconcelos 2010; Miralles et al. 2011), and provided a good starting point to a combined study such as ours– which encompasses all three endemic genera – using the existent sequences produced in previous studies and made available in GenBank, as results of these works. As such, these three reptile lineages – *Chioninia*, *Hemidactylus* and *Tarentola* – provide a good model to test evolution and to apply different bioinformatic tools to understand the origin and diversification patterns on islands, particularly in the Cape Verde archipelago.

***Chioninia* (Scincidae)**

The Scincidae family, which includes the genus *Chioninia*, has undergone complex taxonomic changes along the years, with several groups being reevaluated and renamed as new phylogenetic studies provided more accurate information. One of the affected groups was *Mabuya*, which was divided in four different genera accordingly to molecular results and their ecological distributions (Mausfeld et al. 2002; Carranza & Arnold 2003), namely: *Chioninia*, which now holds the species endemic to Cape Verde (7 species and 7 subspecies, see Table 1.2 and Appendix I); *Trachylepis* which is mostly composed of Afro-Malagasy originated species, with the exclusion of *T. atlantica* found in the Fernando de Noronha island, Brazil; the Asian clade now designated as the genus *Eutropis*; and the “true” *Mabuya* species which are nowadays restricted to Neotropic regions.

***Hemidactylus* (Gekkonidae)**

Within the Gekkonidae, *Hemidactylus* is one of the most diverse reptile groups, with 144 taxa in the Reptile Database (see <http://www.reptile-database.org>). These species can be found across continents and islands in a wide range of habitats, as both endemic and introduced species (Carranza & Arnold 2006). Their remarkable extensive distribution was related to fascinating adaptations to travel huge masses of land and sea, including calcareous and adhesive egg shells, resistance to sea water, and fat reserves which allow for long periods of self-sustenance (Bansal & Karanth 2010). At least one transmarine cross is described, with reptiles traversing from Africa to South-America (Kluge 1969), possibly crossing several Atlantic islands along the way. This specimens are also commensal, and are often found living and traveling alongside humans, which adds yet another layer to their already complex phylogenetic reconstruction (Rocha et al. 2005). In Cape Verde, there are 3 described species and 2 subspecies of these reptiles, several of which endangered or restricted to a single island (see more details in Table 2.1 and Appendix I).

Table 1.2. List of the reptile species and subspecies found in the Cape Verde archipelago. More detailed information can be found in Appendix I, which includes the other reptiles found in the Macaronesia region. A species of *Hemidactylus*, currently classified as *Hemidactylus sp.* is currently being surveyed, and its likely a single island endemic in São Nicolau.

Species/Subspecies	Distribution	Single Island Endemic (SIE)	Red List Category
<i>Chioninia coctei</i>	São Vicente, Santa Luzia, Raso, Branco		Extinct
<i>Chioninia delalandii</i>	Brava, Santiago, Fogo, Rombos		Least Concern
<i>Chioninia fogoensis</i>	Santo Antão	•	Least Concern
<i>Chioninia nicolauensis</i>	São Nicolau	•	Least Concern
<i>Chioninia spinalis</i> ssp. <i>boavistensis</i>	Boavista	•	Least Concern
<i>Chioninia spinalis</i> ssp. <i>maioensis</i>	Maio	•	Least Concern
<i>Chioninia spinalis</i> ssp. <i>salensis</i>	Sal	•	Least Concern
<i>Chioninia spinalis</i> ssp. <i>santiagoensis</i>	Santiago	•	Least Concern
<i>Chioninia spinalis</i> ssp. <i>spinalis</i>	Fogo	•	Least Concern
<i>Chioninia stangeri</i>	São Vicente, Santa Luzia, Raso, Branco		Near Threatened
<i>Chioninia vaillanti</i> ssp. <i>vaillanti</i>	Santiago	•	Endangered
<i>Chioninia vaillanti</i> ssp. <i>xanthotis</i>	Fogo	•	Endangered
<i>Hemidactylus boavistensis</i>	Boavista, Sal		Near Threatened
<i>Hemidactylus bouvieri</i>	São Nicolau, Santo Antão, São Vicente		Critically Endangered
<i>Hemidactylus bouvieri</i> ssp. <i>razoensis</i>	Santa Luzia, Raso		Critically Endangered
<i>Hemidactylus lopezjuradoi</i>	Fogo	•	Data Deficient
<i>Hemidactylus sp.</i>	São Nicolau	•	NA
<i>Tarentola boavistensis</i>	Boavista	•	Vulnerable
<i>Tarentola bocagei</i>	São Nicolau	•	Least Concern
<i>Tarentola caboverdiana</i>	Santo Antão	•	Least Concern
<i>Tarentola darwini</i>	Santo Antão	•	Least Concern
<i>Tarentola fogoensis</i>	Fogo	•	Least Concern
<i>Tarentola gigas</i> ssp. <i>brancoensis</i>	Branco	•	Endangered
<i>Tarentola gigas</i> ssp. <i>gigas</i>	Raso	•	Endangered
<i>Tarentola hartogi</i>	Brava, Rombos	•	Data Deficient
<i>Tarentola maioensis</i>	Maio	•	Least Concern
<i>Tarentola nicolauensis</i>	São Nicolau	•	Least Concern
<i>Tarentola protogigas</i>	Fogo	•	Least Concern
<i>Tarentola raziana</i>	Santa Luzia, Raso, Branco		Near Threatened
<i>Tarentola rudis</i>	Santiago	•	Data Deficient
<i>Tarentola substituta</i>	São Vicente	•	Least Concern

***Tarentola* (Phyllodactylidae)**

The last endemic genus found in the Cape Verde is the Phyllodactylidae genus *Tarentola* with 13 species and 2 subspecies in this archipelago. This group, commonly called wall-geckos, are found in the Mediterranean area, including the Macaronesia regions, and across the Atlantic Ocean in several neotropical islands (Carranza et al. 2000; Gamble et al. 2008a; Gamble et al. 2008b). Previous studies suggest possible cryptic and complex relationships (Vasconcelos et al. 2010), with a likely radiation in a Cape Verde lineage after a single colonization event from the neighbour Canary Islands (Vasconcelos et al. 2012). An endemic Cape Verdean species, *T. substituta*, was reportedly introduced in the island of São Miguel, Azores (Rato et al. 2015), and was also included in our study to review its native status within the Macaronesia clade.

Other Reptiles in Macaronesia

Aside to the mentioned endemic taxa from the genera *Chioninia*, *Hemidactylus* and *Tarentola*, three other reptilian genera can be found across the Macaronesian region, namely *Teira*, *Chalcides* and *Gallotia*.

A single species belonging to the lacertid *Teira*, *T. dugesii*, is found in Madeira archipelago and the Selvagens, with some introduced specimens in the Azores and some Portuguese mainland areas (Arnold & Ovenden 2002).

The Canaries hold the other two endemic reptile genera – *Chalcides* a Scincidae skink, with four described species (Austin & Arnold 2006); and *Gallotia*, a lacertid which can only be found in these islands and has seven recognized species, with several subspecies (Cox et al. 2010).

The conservation status for several Macaronesian endemic reptiles were consulted in the IUCN Red List (IUCN 2017), as its seen in Table 1.2. and in more detail in Appendix I. This information reveals that four from Cape Verde species are in threatened categories [i.e. Critically Endangered (CR), Endangered (EN) and Vulnerable (VU)]: *Hemidactylus bouvieri* (CR), *Chioninia vaillantii* (EN), *Tarentola gigas* (EN), and *Tarentola boavistensis* (VU). The IUCN Red List also includes two possible extinct reptiles – *Chioninia coctei*, from Cape Verde; and *Gallotia auaritae* from the Canary Islands – that were categorized as CR/ Possibly Extinct. Moreover, *Hemidactylus lopezjuradoi* and *Tarentola rudis* were not evaluated because of insufficient information (i.e. Data Deficient); and *Chioninia stangeri*, *Hemidactylus boavistensis* and *Tarentola raziana* are close to meeting the threatened thresholds (i.e. Near Threatened). The remaining 13 species have been classified as Least Concern (LC). This information is also important to better understand how this species have evolved, and how they environmental factors might have influenced their current observed status.

1.4. Aims of the study

We aimed to explore the usefulness and performance of several bioinformatic methods, currently applied to phylogenetic studies, to thoroughly analyse recent radiated insular lineages. The Cape Verde Islands were selected as a model system to study evolution, as the origin of the biodiversity in this archipelago is still not completely understood. All the endemic reptiles taxa, which are distributed in three genera (i.e.: *Chioninia*, *Hemidactylus* and *Tarentola*) were selected as a case-study, since there was a massive amount of molecular data available for this highly diversified group. These three genera were featured in previous studies that offered possible molecular, taxonomic and phylogenetic relationships within each group. Few however, had ever taken in account the three genera simultaneously and accounted for all the Cape Verde endemic reptile diversity. Phylogenetic procedures have also change along the years, and as such most of the software packages used in previous studies have been updated or replaced in the meantime. Simple phylogenetic networks were also performed to provide a look into possible hybridization and recombination processes that may have occurred in these lineages. The collected data sequences were publicly available, as were the several references and software packages used throughout this study. As such, there was a great opportunity to review and combine all the previous finds regarding the Cape Verdean reptiles, using the most up-to-date bioinformatic tools and knowledge, to provide new insights to study evolution on islands. The application of several methodologies explored in different software's was also a goal of this study, namely the use of different methods of phylogenetic inference, in which we worked with different parameters, to take a look on how these alterations influence the final results.

2. MATERIALS AND METHODS

2.1. Data Collection

We compiled a list of the Macaronesian endemic reptiles, with a particular focus on the Cape Verde endemic taxa, including the recently described species, which were featured in previous studies and had available DNA sequences. After sorting and organizing this information, we then ran a search in the GenBank database (<https://www.ncbi.nlm.nih.gov/genbank>) and downloaded these sequences, directly as fasta files or through other tools, such as Geneious. This tool was also used for checking the sequences for possible errors and to apply corrections when applicable.

The first group of sequences found across the multiple references was composed of 1725 DNA sequences for 15 different genes, namely 916 sequences for the genus *Tarentola* for 6 regions, 588 sequences for *Chioninia* for 13 regions, and 221 sequences for *Hemidactylus* for 3 regions (see more details in Table 2.1). Most of these sequences belonged to Cape Verde endemics (1447 in total), the rest coming from other species used to assess the speciation process of each group, as well as three sister taxa species to be used as outgroups in one of the analysis sets. The data undergone a thorough preliminary analysis to carefully evaluate the best possible results. These tests comprised the assessment of the quality of the sequences, the ensued alignment obtained with them, and finally the performance in both Maximum-Likelihood and Bayesian Inference methods of tree calculation.

The final version of the analysis was composed of a smaller set of sequences (see Appendix II), which reduced both redundancy and runtime, while improving the overall results. Two main data groups were created: one using three sister taxa to function as outgroups in the following phylogenetic analysis; and another one in which the outgroups were from within the genus (in the case of *Hemidactylus* and *Tarentola*) or closer to the group (in the case of *Chioninia*, as it's a genus endemic to Cape Verde). In both cases the outgroups were retrieved from previous studies, and whenever possible from trees containing Cape Verde species. This approach allowed us to test these two methodologies and their performance, and provided two different perspectives into the relations of the Cape Verdean reptiles between themselves and with their relatives. All the relevant data regarding the molecular sequences used in this study, such as their source and accession numbers, is shown in full in Appendix III.

Regarding the loci used for each of the reptile groups, their sorting was made by source, specifically, if they were of mitochondrial or nuclear in origin: A) the *Chioninia* analysis set had the biggest gene diversity: four mitochondrial genes, 12S, 16S, cytochrome b (cyt b) and cytochrome c oxidase subunit I (COI); and nine nuclear, oocyte maturation factor Mos (C-mos), recombination activating gene 2 (RAG2), brain-derived neurotrophic factor (BDNF), breast cancer 1 (BRCA1), breast cancer 2 (BRCA2), exophilin 5 (EXPH5), kinesin family member 24 (KIF24), melanocortin 1 receptor (MC1R) and recombination activating gene 1 (RAG1); B) the *Hemidactylus* genus had the lowest number of available genes, with just two mtDNA genes (12S and cyt b) and one nDNA loci (RAG2); C) finally, we used six different genes for the genus *Tarentola*, two of mitochondrial origin, 12S and cyt b; and four retrieved from nuclear DNA, C-mos, MC1R, acetylcholinergic receptor M4 (ACM4) and phosphodiesterase 4 (PDE4). In some genes, namely cytb, there were sequences that were composed of first and second halves' of the locus, while other contained the complete region. To solve this, we first joined the split sequences for each of the individuals, and then added the complete sequences to the data matrix. After completing the alignment, the separated sequences were suitably mapped to the complete data sequences.

2.2. Data Analysis

The first step after acquiring the sequences was to verify and correct those using Geneious, which also allowed a simpler way to download them as previously described. BioEdit v.7.2.5 was also used for some corrections and verifications between analysis, and for some conversions between file formats.

While surveying the several data references, there were some errors when matching their information with the one found in GenBank. Some of the sequences had different names, and in some cases wrong accession numbers. In order to correctly compile and summarize the information we crossed referenced through the several studies (mainly Arnold et al. 2008; Vasconcelos et al. 2010; Miralles et al. 2011; Vasconcelos et al. 2012) and also through the Reptile Database (Uetz 2010) to better assess the most current taxonomic description of the used specimens.

After the aforementioned processing, we ended with a fasta file for each gene, containing all its sequences. The alignment was done through Geneious using MAFFT v.1.3.6, with the default parameters, and then exported as a nexus file.

The next step was using the software DNASp to analyse the haplotypes, and to run fastPhase when ambiguities were found. In some cases, it wasn't possible to use this last tool due to the high number of ambiguities, and so those sites were removed using the Mask Sites tool of Geneious, which marked and removed those sites. After making sure the sequences were free of any faults, they were exported as fasta and phylip files using Geneious.

The tool Concatenator v.1.1.0 (Pina-Martins & Paulo 2008) was used to compile the genes in three separate groups: mitochondrial, composed of all the available mitochondrial genes; nuclear, which had the found nuclear genes; and with both of them. To assure that there were no conflicts between them, an ILD test was done in PAUP v4.0a152 (Swofford 2002), and no incongruences were found in any of the combined genes for *Chioninia* (P=0.974 for the nuclear set, P=1 for the mitochondrial and complete set), *Hemidactylus* (P=0.305 for the mitochondrial set and P=0.571 for the complete set), and *Tarentola* (P=0.837 for the mitochondrial set, P=0.195 for the nuclear set and P=0.988 for the complete set). This test is invaluable to measure possible conflicts between different data sources (in our case nuclear and mitochondrial DNA) to ensure that they converge toward the same phylogenetic tree (Planet 2006).

A polymorphism data analysis was performed for each reptile dataset using a combination of the tools provided in DNASP5 and MEGA7. This evaluation was made for the concatenated sets of mitochondrial and nuclear sequences (see Table 3.1) and for each single gene, which allowed us to assess how each locus contributed to the combined results (Appendix IV). We also created simple phylogenetic trees in MEGA7, using the Neighbour Joining (NJ) methodology to test for possible errors and misinterpretations in our data sequences. These tests provided a simple and yet important way to survey the quality of the data used in the following analysis. This is a vital step for any research study to prevent poor results due to the GIGO (Garbage In, Garbage Out) phenomenon, which can occur when low quality input data is provided (Gelman 2011).

2.2.1. Phylogenetic analyses

After completing the data processing, we analysed the substitution models and partitions across our files using the software PartitionFinder v.2.1.1 (Lanfear et al. 2016). For each datafile, three separate analyses were done, each one using the models available in the following software packages: RAxML, MrBayes and BEAST. Across all analysis, we used the AICc (corrected Akaike Information Criterion) as the selection model, and the greedy scheme, since they're the recommended settings given by the authors of PartitionFinder. The search was made by codon whenever possible, i.e. when using coding genes, since it gave a more thorough analysis of the sequenced data. The results of this analysis are summarized in Table 2.1., and fully detailed in Appendix V.

To perform Maximum-Likelihood analysis (ML) we used RAxMLGUI v.1.5.b, a program that simplifies the use of RAxML, by providing a window like interface and connection to the Cipres Gateway (Miller et al. 2010). This way the input files were imported into the program, and after setting the adequate parameters, directly uploaded to Cipres Gateway, where the analysis was then run. This web based tool allows for storage, management and analysis of data in their cloud servers. Besides the obvious advantages of having a backup storage place, this tool is particularly useful for researchers worldwide, by providing a powerful system that can run the most broadly used programs for phylogenetic inference. This saves a considerable time in the analysis runtime, as the software packages perform a lot faster than in any common personal computer.

For the Bayesian Inference (BI) analysis we used the software MrBayes v.3.2.6. All the runs were performed in Cipres Gateway by uploading a nexus file with the appropriate commands inserted at its bottom. The results were then analysed using Tracer v.1.6. to assess the convergence for the parameters obtained for each run.

To build the created phylogenetic trees we used FigTree v.1.4.3, which allowed for a quick visualization and manipulation of the results. We made use of the annotation feature, which allowed a straightforward way to assign new tip labels and colour schemes by certain parameters. This workflow is graphically represented in Figure 2.1.

Table 2.1. Summary of the partitions obtained in PartitionFinder in the final version of the analysis in both sets of analysis (with and without external outgroups).

Group	Set	No. Genes	No. Partitions RAxML (With/Without External Outgroups)	No. Partitions MrBayes (With/Without External Outgroups)	No. Partitions BEAST (With/Without External Outgroups)
<i>Chioninia</i>	Mitochondrial	4	7/7	8/7	-
	Nuclear	9	10/12	14/17	-
	All	13	15/17	18/23	21/23
<i>Hemidactylus</i>	Mitochondrial	2	4/4	4/4	-
	Nuclear	1	3/3	3/3	-
	All	3	7/7	7/7	7/7
<i>Tarentola</i>	Mitochondrial	2	4/4	4/4	-
	Nuclear	4	6/6	8/8	-
	All	6	10/10	13/12	13/13

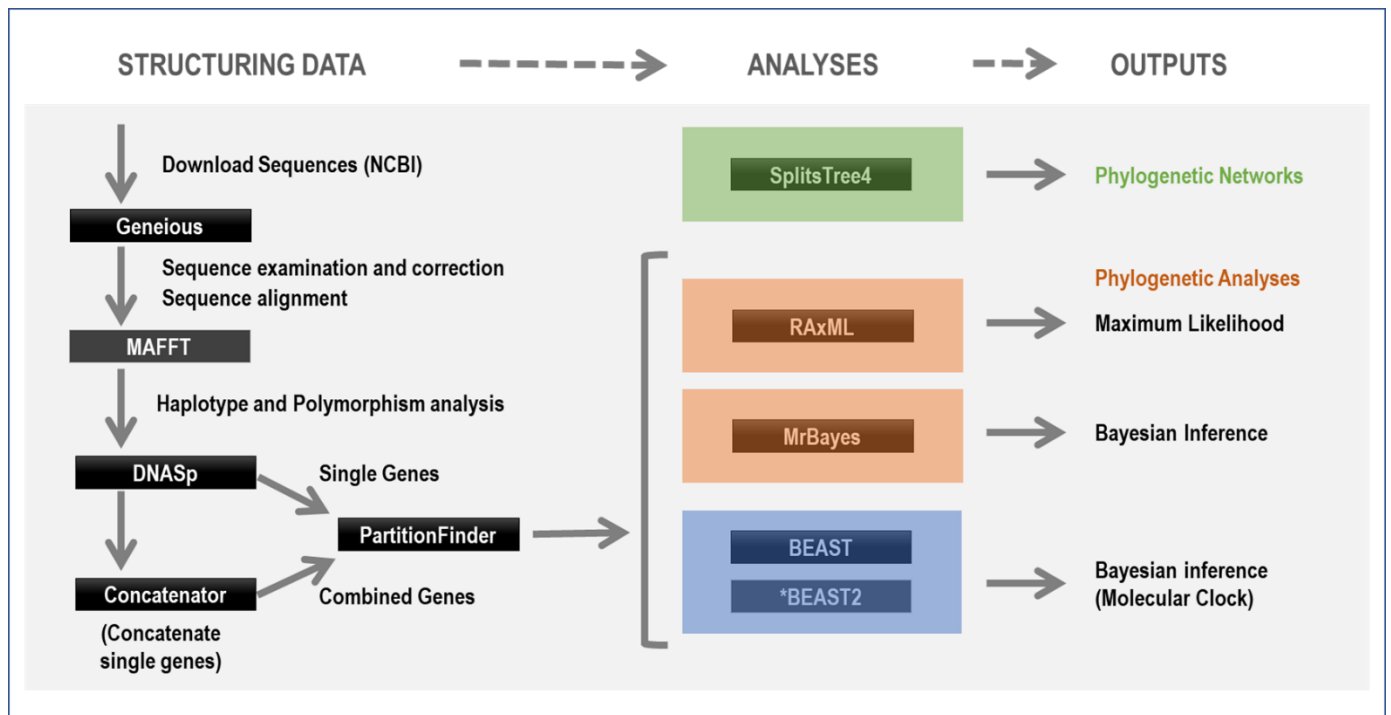


Figure 2.1. Diagram of the workflow performed along this study, with detail on some of the significant stages, software packages and final outputs. (For more details on the software packages used see Table 1.1)

2.2.2. Divergence time estimation

To estimate the divergence time of the three-reptile genera we performed two separate analysis: one featuring the more “classical” approach using BEAST (version 1.8.4); and another using the most recent multispecies coalescent method provided in *BEAST2 (version 2.4.7 of BEAST2). In both cases we made use of the main program and its associate tools, BEAUTI, LogCombiner and TreeAnnotator all in the adequate version for each case. BEAUTI was used to create xml files with the desired parameters, that were then used as an input for each BEAST/*BEAST2 analysis. On the other hand, the other two software packages were used on the attained outputs – LogCombiner, to combine log and tree files from independent runs; and Tree Annotator, which summarizes the information from a sample of trees produced by BEAST/*BEAST2 into a single “target” tree.

To attain the most comprehensive analysis possible in BEAST we used the concatenated sets with all the genes in each of the groups (as described in the results section). After performing the model calculations and partition scheming with PartitionFinder, we ended up with 21 partitions for *Chioninia*, 7 for *Hemidactylus* and 13 for *Tarentola*. A second analysis was made for the *Chioninia* dataset, as some partitions had low ESS (effective sample size) values and were subsequently removed, which divided the data in 10 partitions instead of the previous 21.

To calibrate our trees in this analysis, we opted for the use of gene mutation rates, and the introduction of fossil calibration points indirectly. As the number of reptile fossils for the study reptile genera is rather limited (Table 2.2), most of the time estimate was made using the substitution rate of one or two genes, that had been calculated in previous studies. However, to better date the tree roots, we made use of the dates provided in (Zheng & Wiens 2016), which were calculated using 13 fossil calibrations and previously assess gene mutation rates (Mulcahy et al. 2012; Pyron et al. 2013). Regarding the genes used to estimate divergence time for each of the reptile groups, we used 16S for *Chioninia* (Karin et al. 2016), cytb and 12S for the *Hemidactylus* (Šmíd et al. 2013), and 12S for *Tarentola* (Rato et al. 2012).

All three analyses were performed using the models and partition schemes defined by PartitionFinder, a lognormal relaxed clock and the parameters used in each of the based studies. In all cases we opted to start with a Yule Model of speciation, as it's the most simple model available and is generally more appropriate when considering sequences from different species (Ho & Duchêne 2014), namely: A) for *Chioninia*, we used a normal prior on the lognormal clock mean (16S rate = 0.0080, stdev = 0.0020), as its described in (Karin et al. 2016); B) in the *Hemidactylus* analysis we set the parameters as its done in Šmíd et al. (2013): Yule process of speciation; random starting tree; alpha Uniform (0, 10); yule.birthRate (0, 1000); ucl.d.mean of 12S Normal (initial value:0.00755, mean: 0.00755, Stdev: 0.00247); ucl.d.mean of cytb Normal (initial value: 0.0228, mean: 0.0228, Stdev: 0.00806; C) finally, for the *Tarentola* genus we used a relaxed Uncorrelated Lognormal Clock; Yule process of speciation; random starting tree; alpha Uniform (0, 10); yule.birthRate (0, 1000); ucl.d.mean of 12S Normal (initial value: 0.00827, mean:0.00827, Stdev: 0.00162), as seen in (Rato et al. 2012).

The MCMC chain length and run number was different for each of the analysis, but in all cases the final results were examined in Tracer 1.4. to confirm the existence of convergence and a sufficient ESS (above 200) across all parameters, namely: A) for *Hemidactylus* a single run was needed achieve this quota, which ran for 5×10^8 generations and sampled at each 5000; B) for *Chioninia* we combined two independent runs of 1×10^9 with a sample of 10000, which then composed a single analysis with 1×10^9 with a resample of 20000; C) finally in the *Tarentola* genus we combined two independent runs with 5×10^8 generations and a sample of every 5000 trees, in a single file with 1×10^9 generations and a sampling every 10000. After this processing, we used TreeAnnotator to discard 25% of the each run as burn-in and to produce the maximum credibility trees, using the divergence time mean and 95% highest probability density (HPD). Similarly, to what we did with the obtained results of RAxML and MrBayes, we finalized this analysis by visualizing and manipulating the trees in FigTree v.1.4.3, as we show in Figure 2.1.

For the *BEAST2 analysis the defined parameters were set accordingly to the initial template provided by the program. After obtaining the best substitution models and partitions for each group using PartitionFinder, the sequences were imported into BEAUti with the *BEAST2 template in place. A selection of taxa was made as to reduce sequence redundancy, while maximizing number of available sequences for the selected individuals. In this analysis, we also introduced three new taxa that were considered outgroups, the same that were introduced in the more thorough phylogenetic analyses previously discussed. Likewise, this allowed us to compare the results when looking at divergence times when looking inside the Cape Verdean group and when using external outgroups.

Chioninia (Scincidae)

For *Chioninia* 23 partitions were defined, model sites were all unlinked, while five clock models were defined: 12S clock, 16S clock, a clock containing part of a mitochondrial and two nuclear genes, and finally two others for the remainder of mitochondrial and nuclear genes. The trees were linked in a similar fashion, with a single tree for all the mitochondrial genes, one for the nuclear genes and one with the partition which had a sample of both (the one with the unique clock previously mentioned). The gene ploidy was defined accordingly to each gene tree, namely a value of 0.5 for the mitochondrial group, 2.0 for the nuclear tree and 2.0 for the mixed tree, since the nuclear samples were in higher number. The population model used was the default (Analytical Population Size Integration), the site models were defined accordingly to the results attained in PartitionFinder, and all the clocks were defined as uncorrelated lognormal ones. To obtain the best results all tree species models were tested (Yule, Yule Calibrated and Birth and Death models), and so for each case three independent runs were made and later combined using LogCombiner with a 10% burnin. The results were then verified using Tracer, which showed Yule as the one with higher values of ESS along all the parameters. The selected combined set was then processed using TreeAnnotator using Maximum Clade Probability and Median Heights to produce the final tree. Instead of using a single point of calibration in the root like in the BEAST analysis, we made use of multiple node calibrations using previous calculated times in other studies, and prioritized the ones that made use of fossils and/or mutation rates whenever possible (see Table 2.3 for more details on the used calibration points). As recommended when using secondary calibration values, i.e. imported from other studies, we used normal distributions to set node ages across all analysis, to take in account possible uncertainties (Ho 2007; Ho & Phillips 2009). Aside from the introduced calibration points, all the priors were left as default, apart from the 16S clock which followed the same values as the one in (Karin et al. 2016). The MCMC chain was run for 100000000 generations with a sample every 1000, which showed convergence when verified with Tracer.

Table 2.2. Fossil calibrations used for estimating divergence dates in (Zheng & Wiens 2016) for the BEAST analysis. Nodes with a range of age were constrained by both minimum and maximum ages. The others were limited by minimum ages. No parametric distributions were applied to these age priors. Adapted from Table 6 in the Appendix of Zheng & Wiens 2016.

Fossil calibration	Date (Ma)	Reference
Gekkotan Yantarogekko	54 (min)	(Bauer et al. 2005)
Konkasaurus	65.2 (min)	(Krause et al. 2003)
Chamops, Haptosphenus, Letpochamops, Meniscognathus	70.0 (min)	(Estes 1964), (Bryant 1989) and (Denton & O'Neill 1995)
Palaeosaniwa, Telmasaurus, Cherminotus	70.0 (min)	(Bryant 1989), (Borsuk-Bialynicka 1984) and (Keqin & Norell 2000)
Odaxosaurus	70.0 (min)	(Bryant 1989), (Wiens et al. 2006), (Hugall et al. 2007) and (Conrad 2008)
Priscagamines, iguanines, Isodontosaurus	70.0 (min)	(Keqin & Norell 2000) and (Conrad 2008)
Contogenys, Sauriscus	70.0 (min)	(Estes 1969) and (Bryant 1989)
Coniophis	92.7 (min)	(Marsh 1892)
Primaderma	99.6 (min)	(Nydham 2000), (Vidal & Hedges 2005), (Wiens et al. 2006), (Hugall et al. 2007) and (Conrad 2008)
Hodzhakulia	111 (min)	(Evans 2003) and (Wiens et al. 2006)
Oldest rhynchocephalian	222.8 (min)	(Sues & Olsen 1990) and (Evans 2003)
Bird–crocodile	239–250.4	(Benton et al. 2009)
Lepidosauria–Archosauria	255.9–299.8	(Donoghue & Benton 2007) and (Benton et al. 2009)

Hemidactylus (Gekkonidae)

For *Hemidactylus* we followed a similar workflow, only differing on the defined values of some parameters. This group had seven partitions identified by PartitionFinder, and as such these were defined in BEAUti. The site models were unlinked, and three clocks were defined, one for each of the genes (12S, cyt b and RAG2). The trees were once again group according to the cellular source, mitochondrial and nuclear, and the ploidies were defined as 0.5 and 2.0, respectively. The population model was left as default, the clocks were all set as uncorrelated lognormal and the site model values were altered as seen in the PartitionFinder results. Comparably to what was done in the *Chioninia* analysis, we ran three independent analysis for each tree models (MCMC chain of 100000000 and 1000 sample), and later combined them using LogCombiner to better assess the most appropriate model for the data. The Yule model showed the better results when inspected with Tracer, with higher ESS values across all parameters. All priors were left as default except for the clock values, namely the 12S and cyt b clocks which were defined as seen in (Smîd et al. 2013). For this analysis set we used four calibration points, which were obtained from previous studies (Table 2.3) and followed the same protocol as in the previous analysis group. The final tree was obtained using the combined file with TreeAnnotator using Maximum Clade Probability and Median Heights.

Table 2.3. Calibration nodes used for estimating divergence dates in the *BEAST2 analysis. Each one is marked with an annotation of “C” and a number, which is matched in the corresponding figure of each analysis. A normal distribution was used as the model for all the calibrated, and the mean and SD values were chosen so that 95% of the distribution lied between the two bounds of each node’s age. In the cases where a single age was assigned to a node, the SD value was left as default (1.0).

Group	Calibration Points	Distribution Parameters (Normal)	
		Mean	SD
<i>Chioninia</i>	C1: 64.6 – 67.4 Mya (Skinner et al. 2011; Zheng & Wiens 2016);	66.0	0.7
	C2: 48.9 Mya (Zheng & Wiens 2016)	48.9	1.0
	C3: 43.7 Mya (Zheng & Wiens 2016)	43.7	1.0
	C4: 42.3 Mya (Zheng & Wiens 2016)	42.3	1.0
	C5: 40.7Mya (Zheng & Wiens 2016)	40.7	1.0
	C6: 16.4 Mya (Zheng & Wiens 2016)	16.4	1.0
<i>Hemidactylus</i>	C1: 58.1 – 68.4 Mya (Metallinou et al. 2012; Grismer et al. 2015)	63.25	2.65
	C2: 51.3 – 61.7 Mya (Metallinou et al. 2012; Grismer et al. 2015)	56.5	2.65
	C3: 38.5 – 42.6 (Gamble et al. 2011; Bansal et al. 2013; Grismer et al. 2015)	40.6	1.0
	C4: 22.9 Mya (Gamble et al. 2011)	22.9	1.0
	C5: 20.4 Mya (Zheng & Wiens 2016)	20.4	1.0
	C6: 18.5 Mya (Gamble et al. 2011)	18.5	1.0
<i>Tarentola</i>	C1: 67.5 – 76.8 Mya (Gamble et al. 2011; Bansal & Karanth 2013)	72.15	2.35
	C2: 14.7 – 15.8 Mya (Carranza et al. 2002; Gamble et al. 2011)	15.25	0.3
	C3: 13.6 – 15.4 Mya (Carranza et al. 2002; Gamble et al. 2011; Metallinou et al. 2012; Rato et al. 2012)	14.58	0.45
	C4: 11.4 – 13.6 Mya (Rato et al. 2012 and Carranza et al. 2002)	12.5	0.55
	C5: 9.4 – 10.1 Mya (Crottini et al. 2012; Carranza et al. 2002; Metallinou et al. 2012)	9.8	0.2
	C6: 5.3 – 6.0 Mya (Carranza et al. 2002; Vasconcelos et al. 2010)	5.65	0.18

***Tarentola* (Phyllodactylidae)**

To perform the *BEAST2 with the *Tarentola* genus we used the same pipeline as stated above: PartitionFinder was used to find the correct partitions (13 in total) and their substitution models; three clocks were assigned, one for 12S and another for cyt b, and one for all the nuclear loci; two trees were defined as before with a mitochondrial and nuclear one each with the correct ploidy (0.5 and 2.0 respectively); the population model was left as default and the clocks were all set as uncorrelated lognormal. Once again, we performed three independent runs (MCMC of 100000000 generations with a sample each 1000 trees) which differed on the tree model used and analysed the combined result, with the Yule analysis once again attaining the best ESS values. The priors were kept as default except for the 12S clock that was altered accordingly to the analysis in (Rato et al. 2012) and the six introduced calibrated points taken from previous studies (see Table 2.3 for details) using normal distributions as seen in the other two analysis sets. The combined runs were combined using LogCombiner with a 10% burnin, and then processed using TreeAnnotator with Maximum Clade Probability and Median Heights to obtain the final tree.

2.2.3. Phylogenetic Networks

As we discussed earlier, several research studies make use of phylogenetic networks to find possible hybridization or recombination events during the course of evolution, in a given group of species. Since the other analysis in our study were mainly used to clarify the origin and divergence history of the Cape Verdean endemic reptiles, we also explored this methodology to take a closer look at the relationships within the clade itself.

To do so we made use of the software SplitsTree4 v.4.14.6 (Huson & Bryant 2006), and worked with the concatenated files with all available loci we had previously created, in which only Cape Verde taxa were used, as they were the target for this analysis. This software provides several network, tree and distance options, but we choose to apply the default settings, which employ the Neighbour-Net network method (Bryant & Moulton 2004) and the uncorrected distance method. We created four separate networks, i.e.: one for *Chioninia* and *Hemidactylus*; and two for *Tarentola* – one with only Cape Verde taxa, and one that included the introduced Azorean species, *Tarentola substituta*, which is native to Cape Verde.

3. RESULTS

3.1. Phylogenetic analyses

3.1.1 Separate mtDNA and nDNA gene regions

Chioninia (Scincidae)

Previous studies of this genus had issues with their analysis (see Karin 2016), where conflicting results and poorly support trees were common. To accommodate this fact several analyses were run to test which created better results, as previously described in the methods section.

This genus had also the biggest gene pool sample of the three genera, comprising itself of thirteen genes in total. As such four mtDNA genes were used i.e., 12S, 16S, cyt b and COI; and nine nDNA genes i.e., C-mos, RAG2, BDNF, BRCA1, BRCA2, EXPH5, KIF24, MC1R and RAG1. The genes COI and RAG2 couldn't be analysed on their own in the first analysis set (without external outgroups) as there weren't outgroup sequences available, and on the second set (with external outgroups) these genes, along with BRCA1, BRCA2, EXPH5 and MC1R, couldn't be analysed due to the same limitation.

From the 588 sequences found in the multiple references used, a smaller set was selected to cut on redundancy and analysis processing. Of the 202 selected DNA sequences, 133 came from Cape Verde species, the remainder 69 being composed species from outside of the archipelago, 13 coming from sister taxa which were only used in the applicable set. The most external outgroup wasn't always the same across analysis as there weren't sequences available in all cases and due to the two methodologies used with the outgroups.

In the first analysis set (without external outgroups) performed for the 12S and cyt b genes we used 41 sequences each, and found 30 haplotypes in the first and 36 in the latter. In the second set (using sister taxa as outgroups) 44 and 43 sequences were used, where 33 and 39 haplotypes were found for 12S and cyt b, respectively. The trees obtained in the 12S analysis were generally well supported (see Appendixes VI and IX), and indicated a division between *Chioninia delalandii*, (found in the islands of Santiago, Brava, Fogo and Rombo) and the remainder species of the archipelago. For the cyt b trees several options were tested, the better tree resulting from the analysis ran without ambiguities in their sequences in the first set of analysis. With this gene, the trees were less supported and the division found in 12S trees wasn't found in most cases, and instead there were other groups of specimens (Appendix VI and Appendix IX).

For the 16S gene analysis we used 11 DNA sequences, only one coming from the Cape Verde islands and with 11 haplotypes across the individuals. The trees found for this gene weren't well supported, result of such a limited pool of sequences, but still showed a similar structure found in the previous genes.

The remainder single gene analysis were very limited due to the limited sample size, but were mostly well supported and structured. All the created trees for this genus can be found in the supplementary data section (Appendix VI and IX).

***Hemidactylus* (Gekkonidae)**

For *Hemidactylus*, three separate genes were analysed: 12S, cytochrome B (cyt b) and RAG2. The first two are of mitochondrial origin and the last one is a nuclear locus. In this genus, all the available sequences found from Cape Verde specimens were used, as their number was already restricted.

For the 12S gene we used 28 sequences in total in the first analysis and 31 in the second, 20 belonging to Cape Verde individuals. On the first analysis set *H. festivus* (native from Oman) was used as the most exterior outgroup, with the remaining species coming from other Africa regions and the Neotropics, while on the second, the most exterior taxa was *Tenuidactylus longipes*, a sister taxa to *Hemidactylus* from Iran. The trees obtained were overall well supported, showing a clear division between Cape Verde specimens and their exterior relatives, and within Cape Verde endemics themselves. Once again, the trees were better supported for the BI analysis (Appendix VII and X).

The sequences used for the cytochrome B analysis were from the same individuals as those used on the 12S analysis. Some of the sequences were split in two (cytb1 and cytb2) and others comprised the two parts in a single sequence. To solve this problem cytb1 and cytb2 parts were concatenated and then the alignment was performed, as it's described in the methods section. The results obtained were in all similar to the ones found in the 12S analysis, displaying slightly better bootstrap values on some of the nodes (see Appendixes VII and X).

The RAG2 gene had the less number of sequences of the three (14 from Cape Verde, 20 in total with 2 coming from sister taxa), since most of the studies with Cape Verdean specimens used mitochondrial loci in their assessments. We found five haplotypes from Cape Verde specimens, some across different species, which clear shows that this gene has undergo less evolution then the other two, as is to be expected when comparing nuclear and mitochondrial regions. Both trees from this genus show a good support, particularly between the outgroups, which proved to be invaluable in the concatenated analysis (Appendix VII and Appendix X).

The three analyses seem to indicate a common colonization of the Cape Verde Islands tropical Atlantic islands of São Tomé and Príncipe, with some indicating a proximity from Brazil and others from Kenya.

The specimens *Hemidactylus platycephalus* and *H. longicephalus* seemed to cause some disruption in some of the single gene trees, possibly indicating that some of these genes still hold a thread connecting them to the Cape Verde endemics. This issue wasn't present in the combined analysis, so perhaps the gaps present in the single gene analysis were a factor that was addressed when looking to multiple genes.

All the results obtained for this set of analysis can be found in the supplementary data section, in Appendix VII and X.

***Tarentola* (Phyllodactylidae)**

For the *Tarentola* group six genes were analysed: cyt b and 12S, both mtDNA; C-mos, MC1R, acetylcholinergic receptor M4 (ACM4) and phosphocin (PDC), all nDNA. The latter could not be analysed on his own in the first set (without external outgroups), as it didn't have any outgroups available, i.e. sequences outside the Cape Verde archipelago.

A large group of sequences was found in the several references used, and as such a selection was made to remove redundancy, which reduced the number from 916 to 157 DNA sequences. In this analysis 200 sequences were from Cape Verde species, the remainder belonging to the other groups, some from other Macaronesia regions such as Azores and the Canary Islands. The most exterior outgroup was different between analysis, as some specimens lacked sequences for all the genes and due to the two sets of analysis.

For the cyt b, 88 DNA sequences were selected in the first analysis, and 90 in the second set. Some sequences contained first and/or second halves of the gene, and others were composed of the complete locus. To correct this issue, we combined the disjointed sequences and later added the complete sequences to the matrix, and after performing the alignments the separated were suitably mapped to the complete data sequences. In the end, 70 haplotypes were found in this group in the first and 73 in the second analysis, all of the shared ones being found in Cape Verde specimens and generally across the same species. In the first analysis set (without external outgroups) the most exterior species used was *T. americana* and in the second it was *Ptyodactylus ruusaljibalicus*, the closest relatives being from the Canary Islands and Northern African regions. There's also a clear distinction between Cape Verde individuals: first between *Tarentola nicolauensis* and the rest; and between the latter group, which show two clear groups. The resulted BI tree was more robust then its ML counterpart, showing a better distinction between the outgroups (Appendix VII and XI).

The other mitochondrial gene, 12S, had 56 DNA sequences available, in which 39 and 42 haplotypes were found, in the first and second analysis set, respectively. The results obtained in this gene were slightly worse than the previous one, the BI one once again being better supported.

From all the other genes, C-mos had the least DNA sequences available (19 sequences in total), and only 5 belonging to Cape Verde individuals. As such, the trees created for this analysis were not very supported or robust, but still showed a resemblance between the Cape Verde specimens and a distinction between them and the other species.

The remainder two genes that were analysed individually, ACM4 and MC1R, had 33 and 31 sequences available, respectively. In both cases, haplotypes were commonly found across different species of Cape Verde origin, showing again how differently nuclear genes behave from mitochondrial ones evolutionarily.

In overall, cyt b showed the best results of all the single gene analysis, which seemed to indicate a similar colonization origin as 12S: the Canaries and the Northern Africa region. The complete set of trees created for this genus are shown in full in the supplementary data section (Appendix VIII and XI).

3.1.2. Combined data

To test how groups of genes performed together three analyses were done across the three groups: one with nDNA, one with mtDNA and combined the two. To assess the congruence in each group of genes an ILD test was ran in PAUP. As no incongruences were found in this analysis, all the genes were concatenated as planned for the analysis. The three sets of matrices had a high percentage of missing data (see Table 3.1) due to different found in the number of genes available for Cape Verdean and outgroups, which had higher amount of mtDNA and nDNA respectively. This analysis will also allow for a comparison between the two main approaches in phylogenetic analysis nowadays by providing the results obtained using the concatenation approach.

Table 3.1. Summary of the polymorphism data of the final set of sequences used in our analysis for each of the genera, organized by their source. The “All” column includes both mitochondrial and nuclear genes combined. There was only one nuclear gene for the genus *Hemidactylus*, therefore there is no "combined" file for nuclear genes. Data obtained using DNASp and MEGA.

	<i>Chioninia</i>			<i>Hemidactylus</i>		<i>Tarentola</i>		
	Mitochondrial	Nuclear	All	Mitochondrial	All	Mitochondrial	Nuclear	All
Number of genes	4	9	13	2	3	2	4	6
Number of sequences	46	34	46	26	25	87	46	88
Alignment length (base pairs)	2138	6847	8985	1528	2397	1133	1865	2998
Number of Haplotypes	41	14	42	23	23	74	41	88
Number of Sites excluding gaps/missing data	1412	1384	2408	580	583	732	1050	1272
Number of Sites with gaps or missing data	726	5463	6577	948	1814	401	815	1726
% of gaps and missing data in the matrix	33.9	79.8	73.2	62.0	75.6	35.4	43.7	57.6
Conserved sites with all ingroups	1463	5958	7421	618	1416	581	1790	2371
Conserved sites with Cape Verde species only	1191	789	1980	826	1624	851	1833	2684
Variable (polymorphic) sites with all ingroups	666	772	1438	454	455	524	70	594
Variable (polymorphic) sites with Cape Verde species only	411	10	421	220	223	251	25	276
Parsimony informative sites with all ingroups	527	96	623	316	267	388	27	415
Parsimony informative sites with Cape Verde species only	389	5	394	200	175	212	14	226
Singleton variable sites with all ingroups	137	631	768	142	182	135	43	178
Singleton variable sites with Cape Verde species only	22	5	27	19	48	39	30	50
GC Content	0.443	0.441	0.442	0.477	0.458	0.492	0.482	0.488

***Chioninia* (Scincidae)**

To perform the analysis of mtDNA we combined the mitochondrial genes mentioned in the single gene analysis: 12S, 16S, cyt b and COI. To test which parameters worked better for this group, several trials were made, which included comparisons between files with and without ambiguities on their sequences, and with or without haplotype analysis. The group with discriminated haplotypes and without ambiguities achieved better results than the others, and as such they were used as the final inputs for the analysis with RAxML and MrBayes. The outgroup *Eumecia anchietae* was incorrectly clustered with Cape Verde species in some of the analysis, and as such separate runs were made without this outgroup in some cases.

Other options were also verified, particularly how the results varied when using multiple outgroups or a midpoint root. As it was done in the single gene analysis, two different sets were performed, one in which the outgroups were within the genus, and another in which external taxa were used. Overall, the trees found in both ML and BI analysis were poorly supported externally, the better bootstrap values being found further in the tree. Interestingly, the results obtained using multiple outgroups and midpoint roots were also a lot better than the ones without this parameter, the most exterior nodes showing the biggest improvement.

The concatenated file with the nine nDNA genes had a high percentage of missing data (see Table 3.1), consequence of the lack of nuclear data available for Cape Verde specimens, which only had RAG2 sequences. On the other hand, outgroup species lacked this gene, but had most of the others, so the analysis could still be performed. We also constructed a tree for both ML and BI without RAG2 (i.e. without any ingroup species), which gave us an idea of the relationship between the several outgroups. Overall, the results of this analysis weren't very well supported, derived of the issues described above or perhaps because of the slower evolution rate of nuclear genes, which could result in incomplete speciation.

Finally, we combined all the possible genes in a single file to access the complementarity of mtDNA and nDNA sequences, and a more robust examination for this group. As it was in the case of the single genes studies, the trees created using MrBayes were better supported than the others obtained in RAxML, with nodes better supported within the trees. In the ML analysis, trees using multiple outgroups or midpoint roots, also showed better bootstrap values when compared to the general ML and BI created trees.

In all the analysis, there seems to be an indication of two groups within Cape Verde: one with *C. delalandii* and *C. vaillanti* (both subspecies); and the other with the remaining specimens. Since the outgroups used for this examination were farther apart than the ones used in *Hemidactylus* and *Tarentola*, is harder to point a possible colonization point for this genus.

Nevertheless, there seems to be a closer relation to the *Trachylepis* group and the *Heremites* groups, which are mostly found in Africa and the Middle-East. Both genus seems to be equally related to the Cape Verde specimens, none of the species showing a closer resemblance than another, a possible indicator that the original colonizer species of the islands have become extinct since then (Figure 3.1), or that they weren't included in the analysis. Unlike the other two analyses with *Hemidactylus* and *Tarentola*, in this group the introduction of three more external taxa to the analysis didn't improve the overall quality of the trees, even though higher values of bootstrap were achieved, it still wasn't possible to clear the relationship between the Cape Verdean clade and the genera *Trachylepis* and *Heremites*.

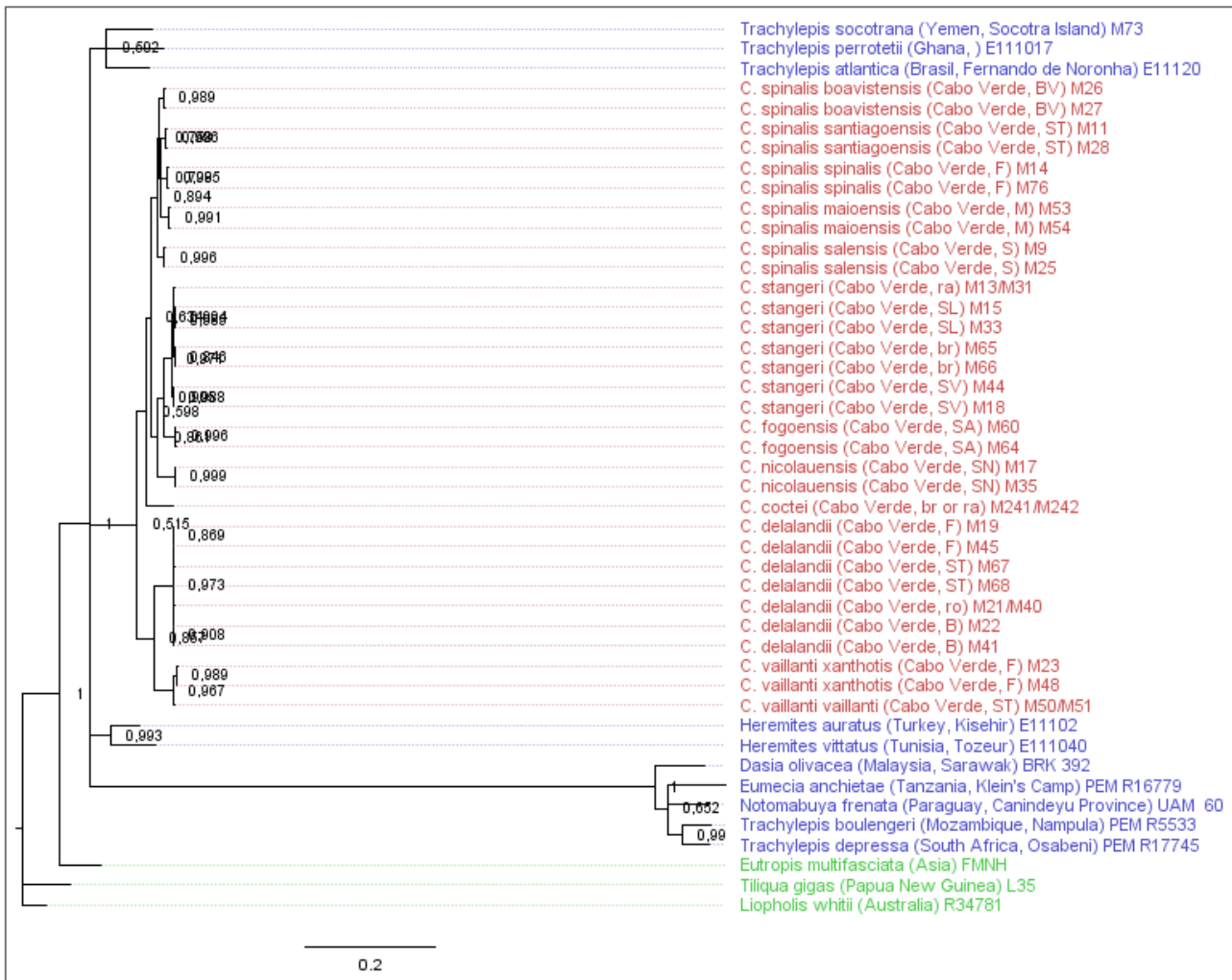


Figure 3.1. Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Chioninia*. Cape Verdean species are marked in red, ingroups are in blue and the outgroups in green. Each taxon has its country and region, followed by the code given by the study that collected the sample in parenthesis (see Appendix II). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio.

Hemidactylus (Gekkonidae)

Hemidactylus had the less available genes of the three genera, 12S and cyt b were concatenated and examined as part of the mtDNA analysis, and posteriorly combined with RAG2 for a complete analysis. A combined nDNA couldn't be done as only a single gene was available. ILD tests were performed with PAUP and no incongruences were found between any genes.

The mtDNA results showed well supported trees, the ones obtained with MrBayes being slightly better than the ones created with RAxML. The same pattern of divergence within Cape Verde shown in the single gene analysis, was also visible in this combined set (see Appendix VII and X).

When using the three available genes in the analysis, the obtained trees had high bootstrap values and once again two separate groups within Cape Verde. The closest species were the same as the ones seen in the previous analysis, indicating a probable common ancestral with species from São Tomé and Príncipe, *H. greeffi* and *H. longicephalus* (Figure 3.2).

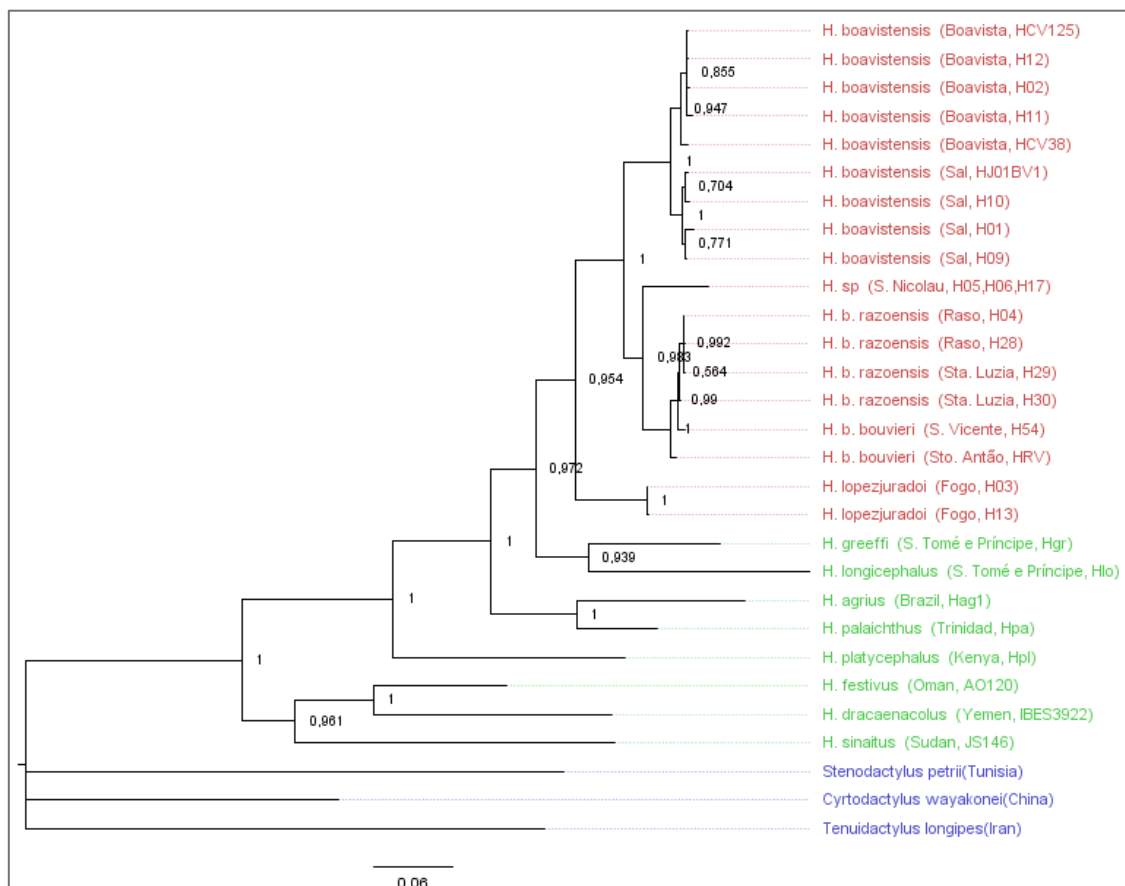


Figure 3.2. Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Hemidactylus*. Cape Verdean species are marked in red, ingroups are in blue and the outgroups in green. Each taxon has its country and region, followed by the code given by the study that collected the sample in parenthesis (see Appendix II). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio.

***Tarentola* (Phyllodactylidae)**

For *Tarentola*, our analysis had two mitochondrial and four nuclear genes, allowing analysis for mtDNA, nDNA and the two sets combined. All the genes were found to be congruent when tested in the ILD through PAUP.

This genus had the biggest sample pool of the three genera, with a total of 141 sequences selected for the mtDNA analysis, 64 haplotypes being found when analysing through DNASp. The results created through this combined set showed a resemblance to the ones seen in the single gene analysis. Both ML and BI trees showed good bootstraps values and the same pattern of clustering, with species from the Canary Islands and northern Africa showing a higher resemblance to the Cape Verde species (Appendix VIII and Appendix XI).

On the other hand, the results obtained with the nDNA were not as supported as the previous group, likely resulting from the lower rate of evolution. The number of haplotypes was also quite inferior than the previous set, with only 41 found in the 102 used sequences. The created BI tree still managed to group similar species together, but with no distinction being shown between them (Appendixes VIII and XI).

When we combined the two sets of genes the total of sequences analysed was 243, with 88 haplotypes being found. The two trees obtained with this group were of an identical topology as the ones found in the single genes and the mtDNA analysis, namely: A) both trees showed higher bootstrap values of all the runs, showing a good complementarity between the genes; B) and a closer proximity of the Cape Verde species to the ones coming from the Canary Islands, (*T. gomerensis* and *T. delalandii*) and Morocco (*T. chazaliae*) was observed (Figure 3.3).

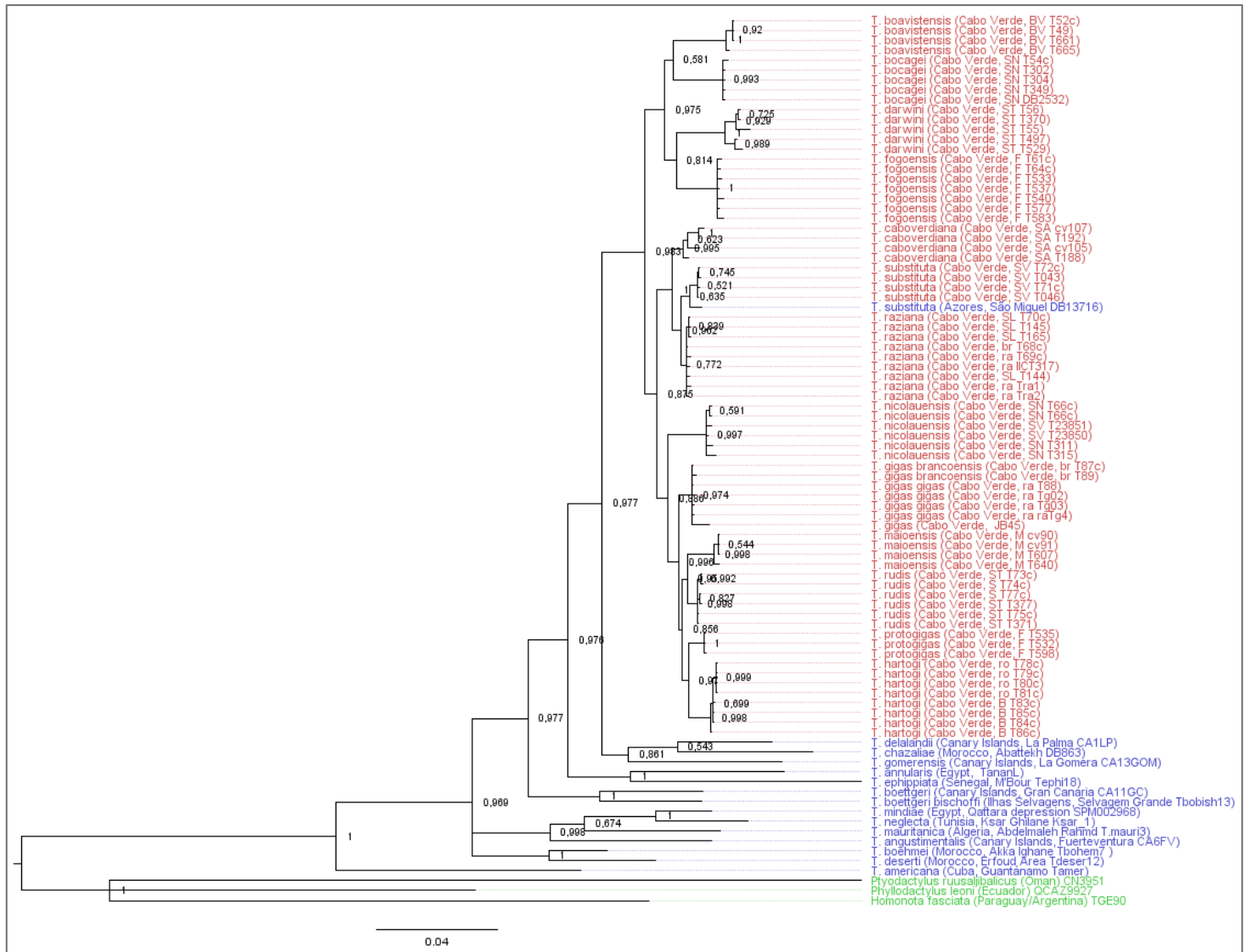


Figure 3.3. Phylogenetic tree obtained in the BI analysis using external outgroups for the combined set with all the genes for *Tarentola*. Cape Verdean species are marked in red, ingroups are in blue and the outgroups in green. Each taxon has its country and region, followed by the code given by the study that collected the sample in parenthesis (see Appendix II). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio.

3.2. Divergence time estimation

In order to better understand how and when the three analysed genera colonized and diverged in the Cape Verde Islands, we employed the use of the software packages BEAST and *BEAST2, and its associated tools for the three genera, as it's described in the methods section.

As such we collected all the available sequences from all Cape Verdean species and several of the previously used outgroups, and the most commonly used software program, using two different approaches, to estimate the divergence of these lineages, namely one using the first version of BEAST and another using the package *BEAST2, provided in the second version of the software program.

For BEAST we ran three separate analysis, one for each of the genera, using concatenated inputs with all the available genes, similarly to what was done in the previous phylogenetic analysis with RAxML and MrBayes. As such we performed model selection analysis in PartitionFinder, and then used BEAUti to assign the adequate parameters (see method section for details). The calibration of each tree was made in the root, using the calculated times from (Zheng & Wiens 2016): *Chioninia* – 43.7 Mya; *Hemidactylus* – 73.3 Mya; *Tarentola* – 40.5 Mya.

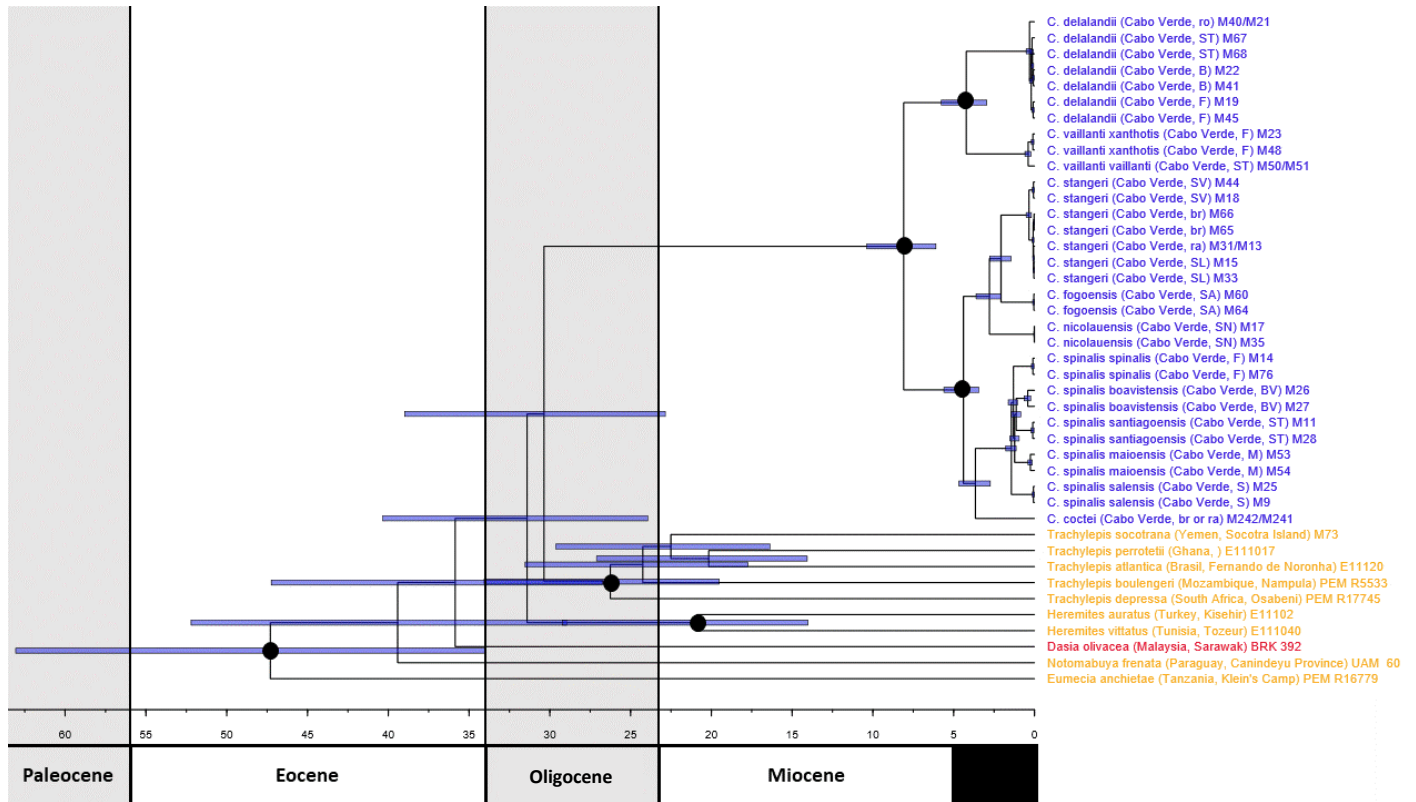


Figure 3.4. Results from BEAST for the *Chioninia* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. Cape Verdean species are marked in blue, the outgroups are in yellow and the taxon in red, *Dasia olivacea*, was the most exterior outgroup in the other two analysis (RAxML and MrBayes). Each taxon has its country and region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

The results obtained for the *Chioninia* genus, were similar to the ones seen in the previous analysis, with a clear Cape Verdean clade, with the closest relatives belonging to the *Trachylepis* genus. Within the Cape Verde lineages, two groups are visible: one containing *C. delalandii* and the two subspecies of *C. vaillanti* that diverges ~5 Mya ago; and the other containing the remainder species with a similar if yet earlier divergence time. One of the possible problems in this tree, is the position of the species *Eumecia anchiatae*, which was previously found closer to the Cape Verdean clade.

The obtained divergence times for *Hemidactylus* were earlier to the ones obtained in *Chioninia*, which is a possible result of the high number of taxa and diversity found in this genus. The results were well supported, with most of the nodes reaching values over 0.95 of BI posterior probability.

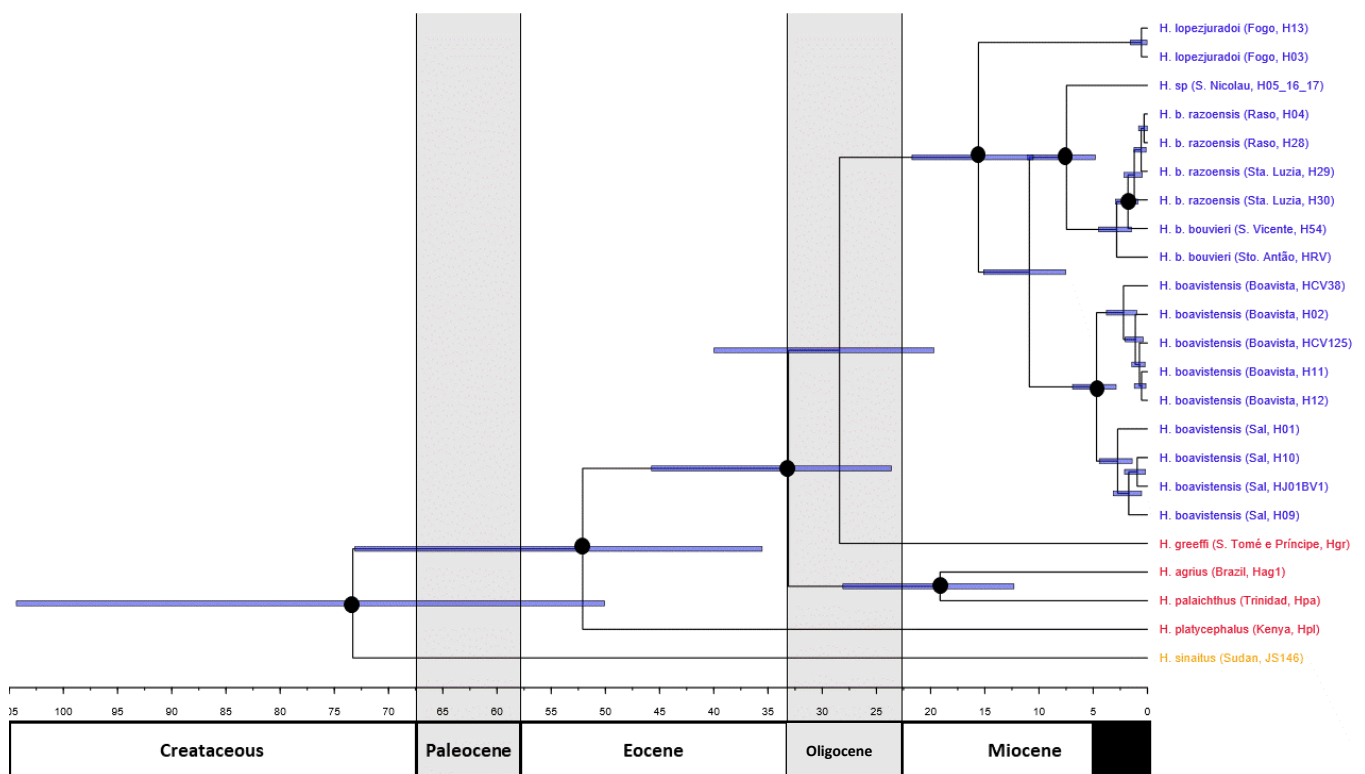


Figure 3.5. Results from BEAST for the *Hemidactylus* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. Cape Verdean species are marked in blue, the outgroups are in yellow and the taxon in red, *Hemidactylus sinaitus*, was the most exterior outgroup in the other two analyses (RAxML and MrBayes). Each taxon has its country/region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

As seen in the previous ML and BI results, this genus seems to be closer to the specimens of São Tomé e Príncipe, from whom they diverged during the Oligocene epoch, around 25 Mya. The Cape Verdean clade has diverged later on, during the Miocene.

Finally for *Tarentola*, we found a similar topology similar to the one seen with RAxML and MrBayes. This genus shows a clear split between the Cape Verdean clade and the outgroups, with a closer resemblance to the specimens found in the Canary Islands and Morocco (around 8 millions years ago). The obtained results show that this is the most recent genus of the three endemics, with most of the divergence happening after the Eocene epoch.

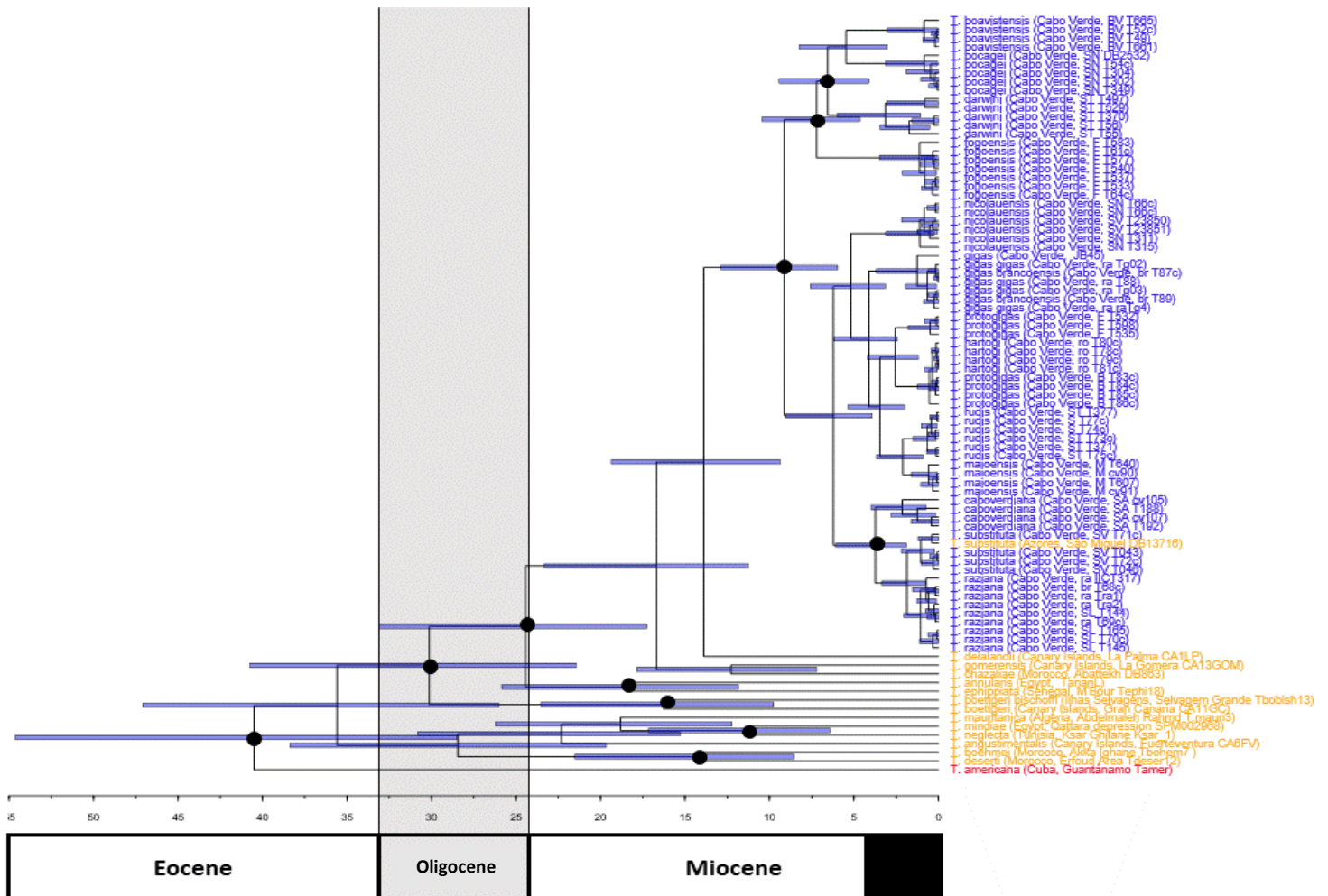


Figure 3.6. Results from BEAST for the *Tarentola* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. Cape Verdean species are marked in blue, the outgroups are in yellow and the taxon in red, *Tarentola americana*, was the most exterior outgroup in the other two analyses (RAxML and MrBayes). Each taxon has its country/region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

As previously stated, we also performed a divergence time analysis using *BEAST2 for the three reptile genera, which was similar to the BEAST analysis with a few exceptions, namely: the introduction of three new taxa that had previously been used in the phylogenetic analysis that utilised external outgroups; and the use of multiple calibration points (see Table 2.3 for more information).

Overall, the results displayed similar topologies to the BEAST, ML and BI analysis. This analysis also achieved less uncertainty in most node ages, likely due to the introduction of several calibrations points (Figures 3.7, 3.8 and 3.9).

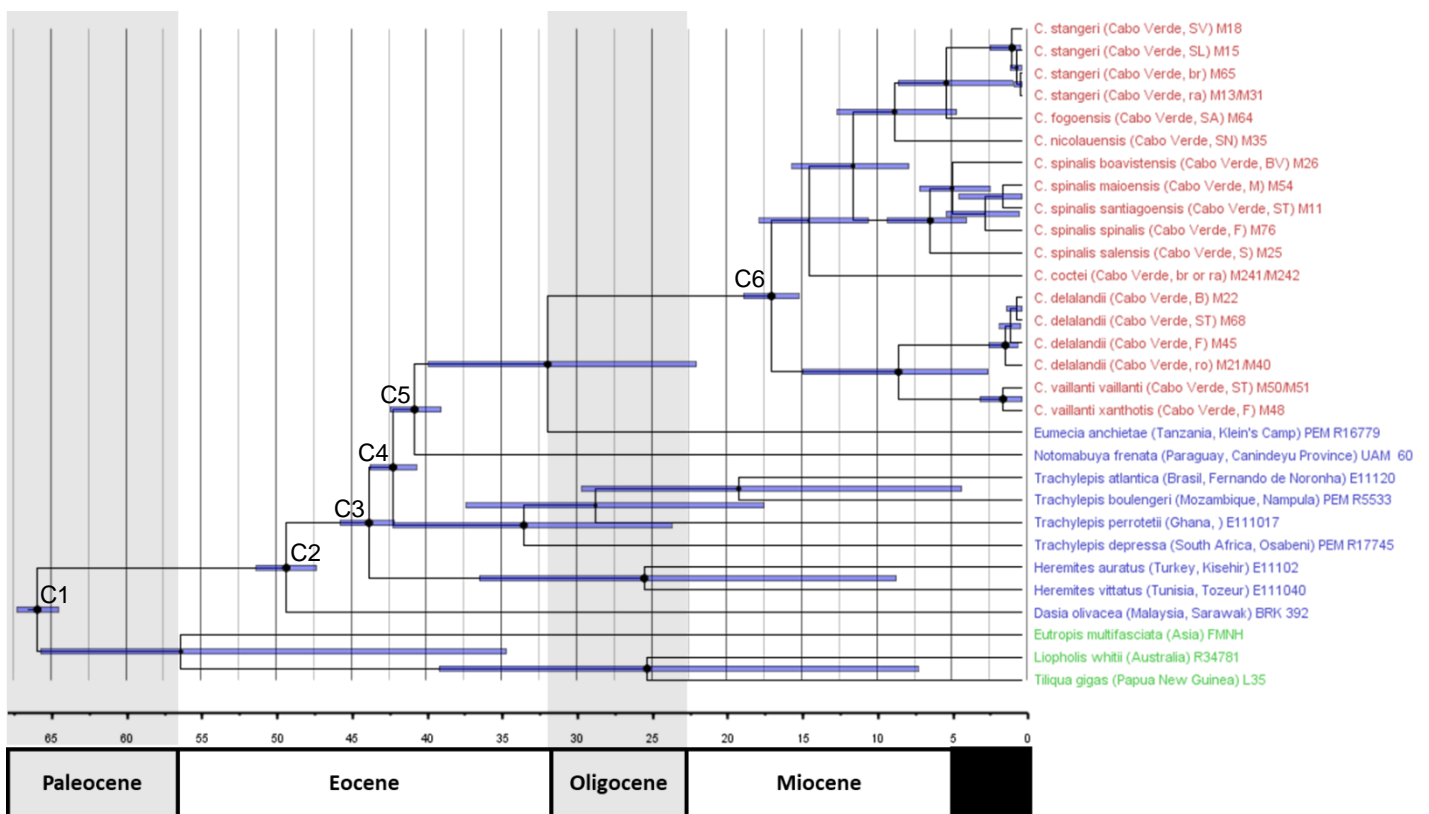


Figure 3.7. Results from *BEAST2 for the *Chioninia* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. The size of the circles on the nodes is representative of the posterior value, higher values being assigned larger circles. Cape Verdean species are marked in red, species from outside Cape Verde are in green, and the taxa in red were the most exterior outgroups in one of the sets of ML and BI analysis. Each taxon has its country/region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

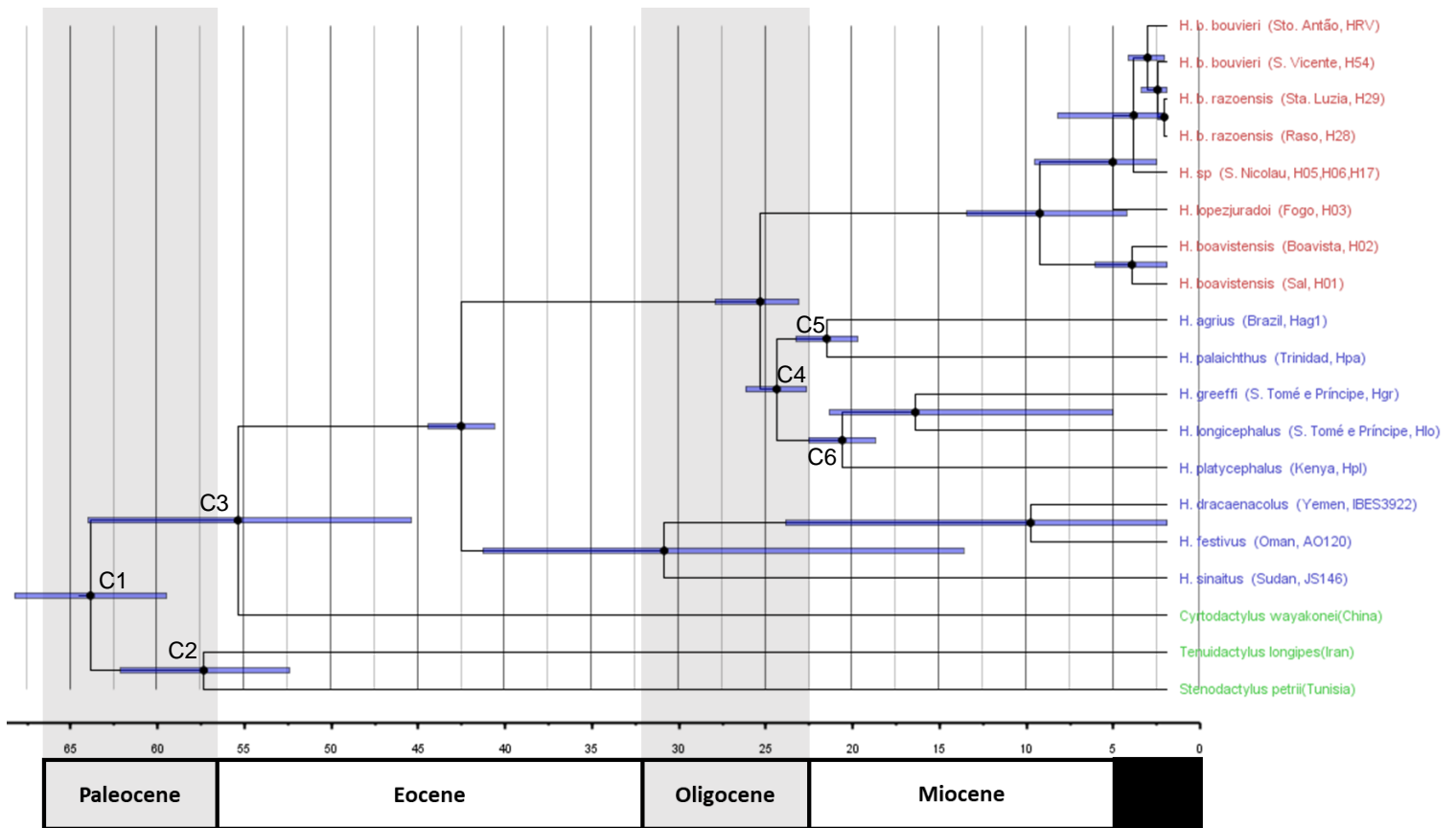


Figure 3.8. Results from *BEAST2 for the *Hemidactylus* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. The size of the circles on the nodes is representative of the posterior value, higher values being assigned larger circles. Cape Verdean species are marked in red, species from outside Cape Verde are in green, and the taxa in red were the most exterior outgroups in one of the sets of ML and BI analysis. Each taxon has its country/region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

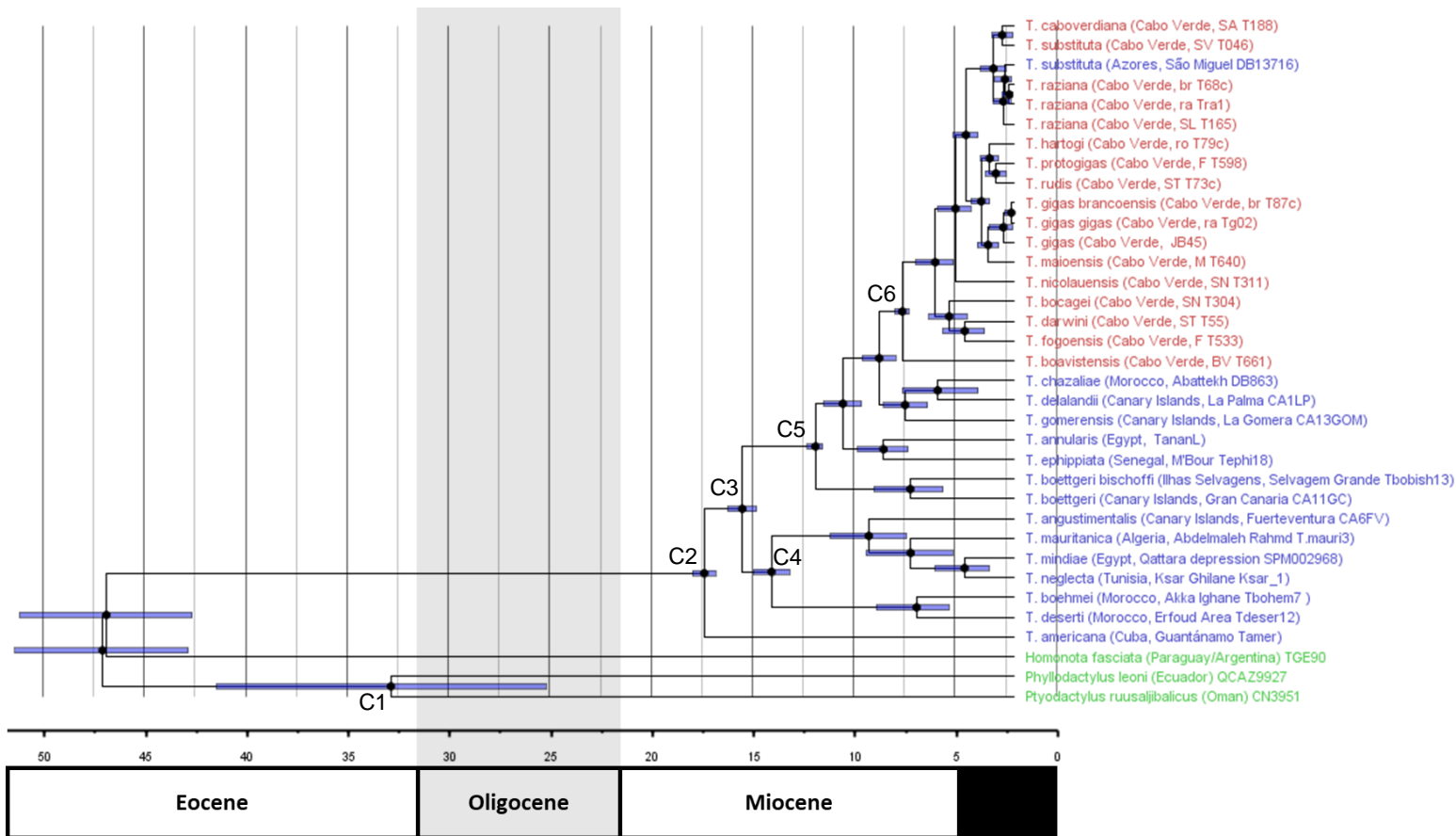


Figure 3.9. Results from *BEAST2 for the *Tarentola* genus with divergence times estimated using all the genes, as described in the methods section. The bars represent 95% confidence intervals on the node heights. The black circles on nodes show BI posterior probability $P \geq 0.95$. All the clades composed of the same species achieved BI posterior probability $P \geq 0.95$. The size of the circles on the nodes is representative of the posterior value, higher values being assigned larger circles. Cape Verdean species are marked in red, species from outside Cape Verde are in green, and the taxa in red were the most exterior outgroups in one of the sets of ML and BI analysis. Each taxon has its country/region, followed by the code given by the study that collected the sample in parenthesis (see Appendix VIII). For Cape Verde (Cabo Verde) the region acronyms are the following: BV, Boavista; SN, São Nicolau; F, Fogo; ST, Santiago; SV, São Vicente; SL, Santa Luzia; br, Branco; ra, Raso; SA, Santo Antão; B, Brava; ro, Rombos; M, Maio. Geological scaled based on (Karin et al. 2016).

3.3. Phylogenetic Networks

To test for possible events of hybridization and recombination, and also to take a closer look at the relations within the Cape Verde clade, we performed a total of four networks analysis with SplitsTree4, using the default settings of the program, as described in the methods section. For the *Chioninia* and *Hemidactylus* groups, we created a single network (Figure 3.10), which included the endemic Cape Verdean reptiles of each of these genus. For *Tarentola* we performed two separated analysis (Figure 3.11), one which included the introduced specimen in Azores, *T. substituta*, and other that excluded this individual.

Overall, the obtained networks showed the same patterns observed in the phylogenetic trees, namely the number and composition of the groups. The webbed pattern seen across the networks is also an evidence that several hybridization or recombination events may have taken place in these lineages across the span of evolution, which seems to indicate that interactions between these species was common, even across different islands. This pattern can also be a result of retention of ancestral polymorphism or perhaps missing haplotypes, and as such additional analyses with each individual gens could prove further insights.

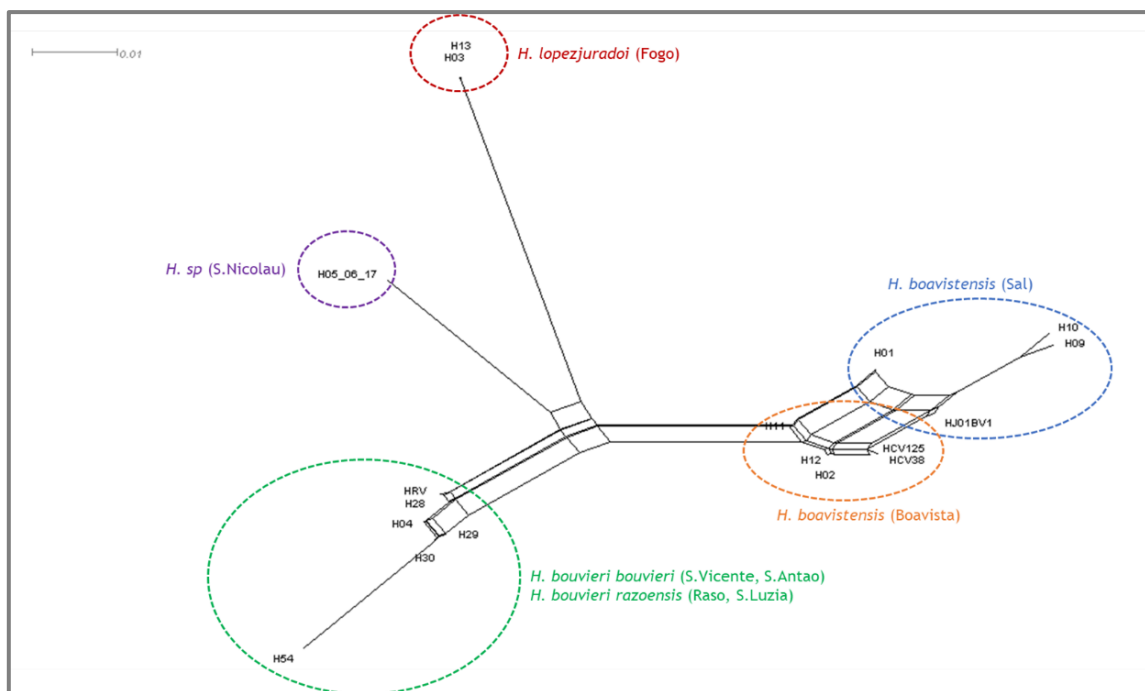


Figure 3.10. Phylogenetic networks created for *Hemidactylus*, using SplitsTree4 neighbour-net algorithm (see methods for more details). The shown codes are the ones given by the study that collected the sample, as seen in Appendix VIII.

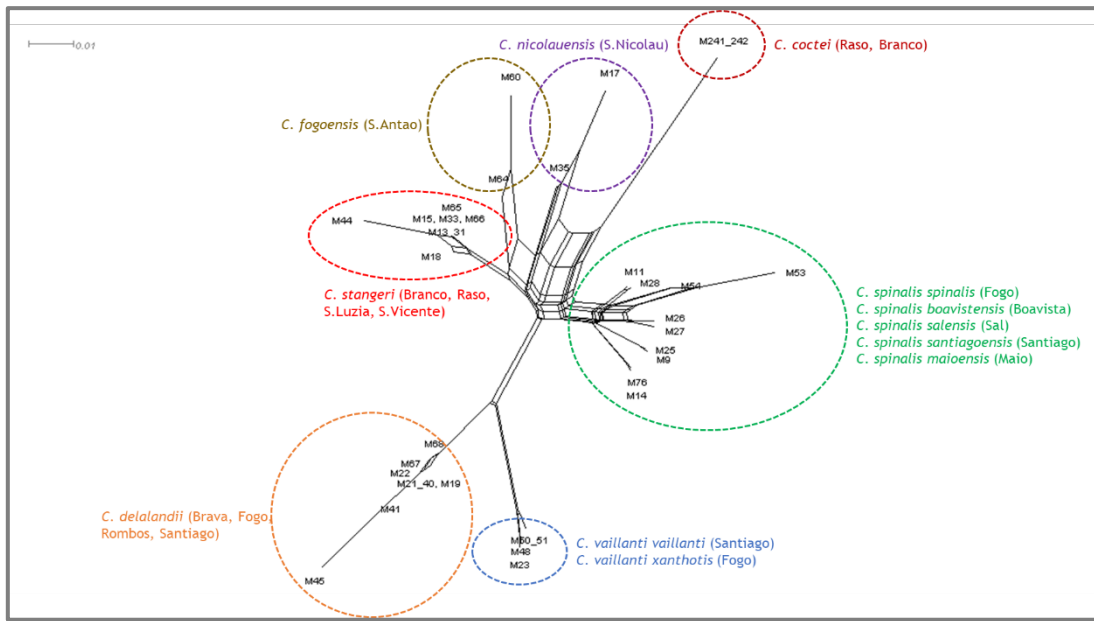


Figure 3.11. Phylogenetic networks created for *Chioninia*, using SplitsTree4 neighbour-net algorithm (see methods for more details). The shown codes are the ones given by the study that collected the sample, as seen in Appendix VIII.

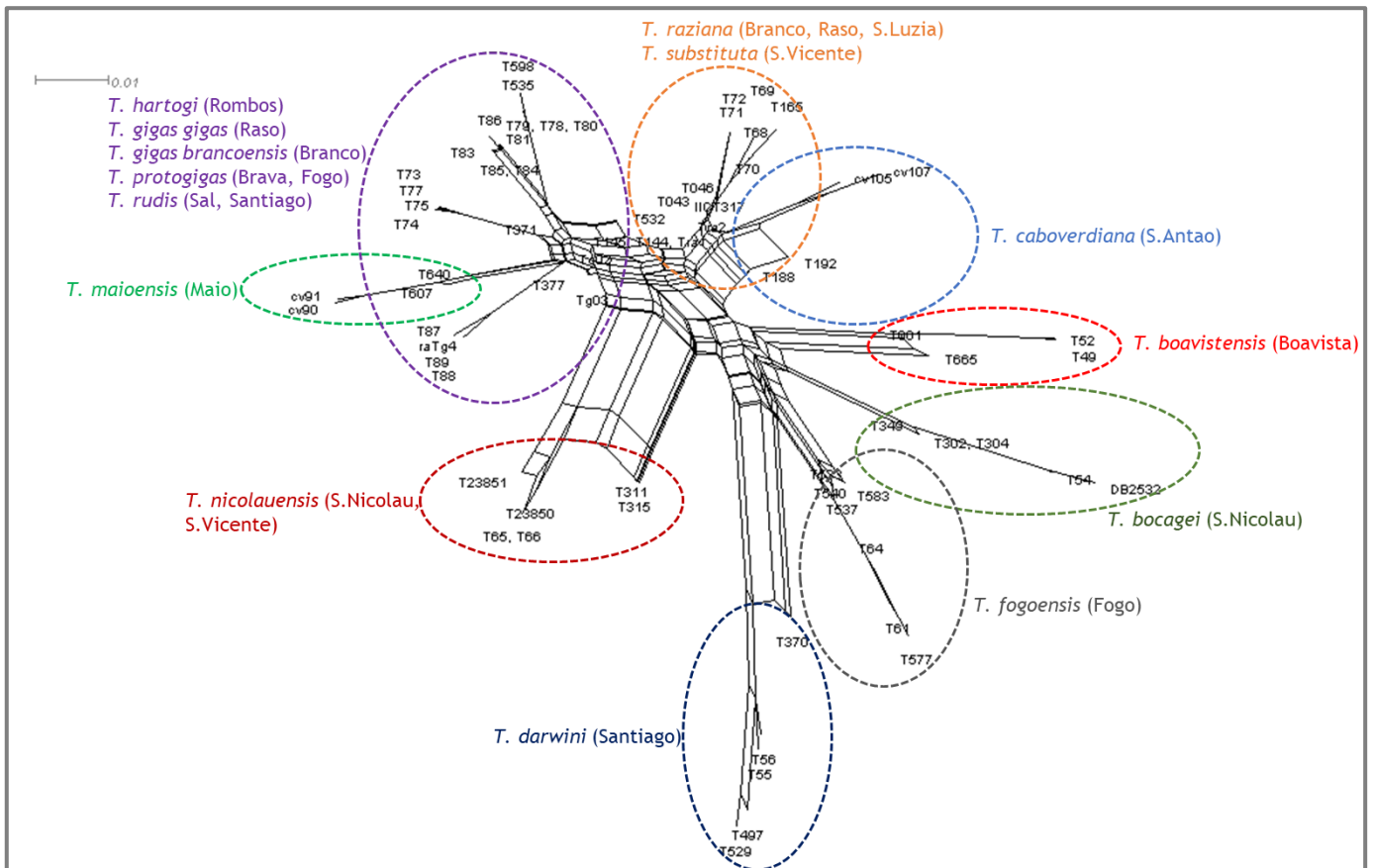


Figure 3.12. Phylogenetic networks created for *Tarentola* without the Azorean specimen, using SplitsTree4 neighbour-net algorithm (see methods for more details). The shown codes are the ones given by the study that collected the sample, as seen in Appendix VIII.

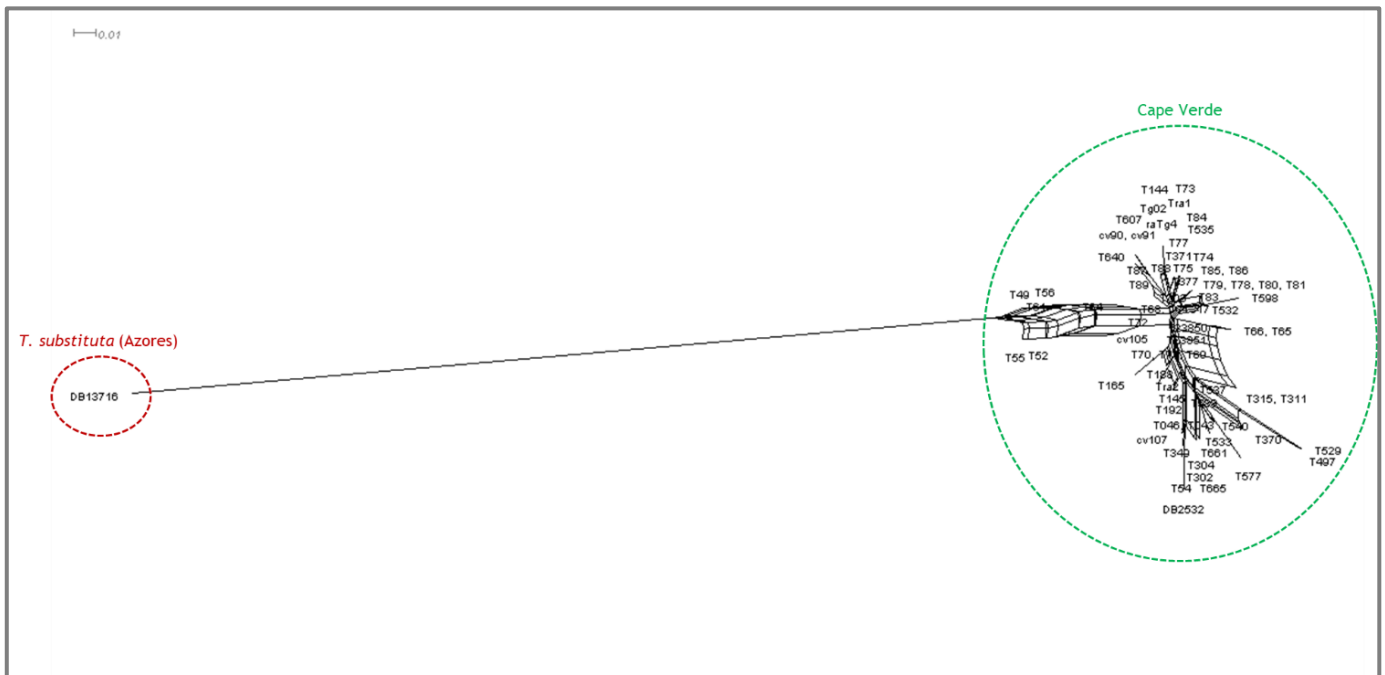


Figure 3.13. Phylogenetic networks created for *Tarentola* with the Azorean specimen, using SplitsTree4 neighbour-net algorithm (see methods for more details). The shown codes are the ones given by the study that collected the sample, as seen in Appendix VIII.

The resulting network for the *Hemidactylus* genus shows four distinct clusters: one containing *Hemidactylus boavistensis*, native to both Sal and Boavista; one with *Hemidactylus lopezjuradoi* from Fogo island; the single-island endemic of São Nicolau (*Hemidactylus sp.*); and finally one harboring both subspecies of *Hemidactylus bouvieri*. The interconnectivity between the lineages of *H. boavistensis* shows that the present populations share a lot of similarities across islands, though some differentiation is starting to develop. Accordingly, the group of *H. bouvieri bouvieri* and *H. bouvieri razoensis*, points to a strong connectivity between individuals from São Vicente, Santo Antão, Santa Luzia and Raso, an indicator that crossings between islands was possible during low sea-level periods, a pattern that is also noticeable in *Chioninia* and *Tarentola*, as well as other Cape Verde species. Lastly, the relationship between the *H. boavistensis* and the remaining groups, shows that the islands of Sal and Boavista were the possible colonization entry points in the archipelago, which goes in accordance to the fact that these are the oldest islands of Cape Verde.

The two clades seen in the phylogenetic inference analysis are also clearly present in the generated network for the *Chioninia* genus – one composing itself of *C. delalandii* and *C. vaillantii* (both subspecies); and the other of the remaining species. This seems to indicate two separate radiation events that followed different directions, with one group ranging to the southern islands of Brava, Fogo and Santiago, and the other going north to the Barlavento islands. The network shows different possibilities of points of origin in Cape Verde, but the distribution of the several lineages of *C. spinalis* on the islands that are both oldest and closest to the mainland, and their relative position on the network, may suggest that these were the first to arrive to the archipelago.

Last but not least, the obtained network for the *Tarentola* reveals similar patterns to those seen in the phylogenetic trees for this group. Even though there's clear distinction between the four species, a big cluster containing *T. darwini*, *T. fogoensis*, *T. bocagei* and *T. boavistensis*, is perceptible within the network. The connection between the groups of *T. caboverdiana*, *T. substituta* and *T. raziana*, and the connection between the *T. maioensis*, *T. hartogi*, *T. gigas* (both subspecies), *T. protogigas* and *T. rudis*, is also evident as seen in the obtained phylogenetic topologies. The species of *T. nicolauensis* is found by itself around the middle of the network, suggesting a different evolutionary pattern from the other endemic reptiles. As this genus' web pattern is very complex, a clear point of colonization is not noticeable. However, taken in account the location and age of the islands, it's possible that the first specimens arrived at Boavista, with some colonizing the northern islands from São Nicolau. Once again, the high level of interconnectivity across species and islands, strongly suggests that periods of coexistence occurred in the past, as the sea level dropped and increased the area of these land masses.

Regarding the introduced species of *Tarentola substituta* found in the island of São Miguel in the Azores, the created network using this lineage shows a remarkable distinction between this exemplar and the remaining Cape Verdean specimens. As such, it's likely that the introduction of this species is not recent or that adaptation occurred very rapidly, as this lineage seems to have differentiate itself quite well from its ancestral relatives. To further investigate this introduction, a more complete sampling of this specimens would be necessary, as at present only one gene is available.

3.4. Reptiles richness and endemism

As previously referred, they're six reptile genera (i.e. *Chalcides*, *Chioninia*, *Gallotia*, *Hemidactylus*, *Tarentola* and *Teira*) with several endemic species that have diversified within the Macaronesian Region (Figure 3.12; and see details on ecology and conservation in Appendix I). However the Cape Verde Islands shows the huge level of reptiles diversification within the Macaronesian Islands. With three genera, and around half of the described endemic species in the region, it's representative of a big slice of biodiversity in Macaronesia.

The *Tarentola* genus, which can be found in both Cape Verde and in the Canary Islands, has the higher number of species already assessed in IUCN (14 in total), which have diverse conservation status as well. Most of the species are in the lowest risk of extinction (Least Concern), but some were ranking in threatened categories, namely *T. gigas* spp. and *T. boavistensis*. On the hand, *T. rudis* has insufficient of data to estimate a reliable status, but its possible its also under risk due to low population density and small distribution areas.

The *Hemidactylus* genus in the other hand, has only three species already assessed in IUCN, all considered under threat or data deficient. As this genus is quite widespread around the world, is possible that several extinctions may have occurred in the island and may still occur if conservation efforts are not made.

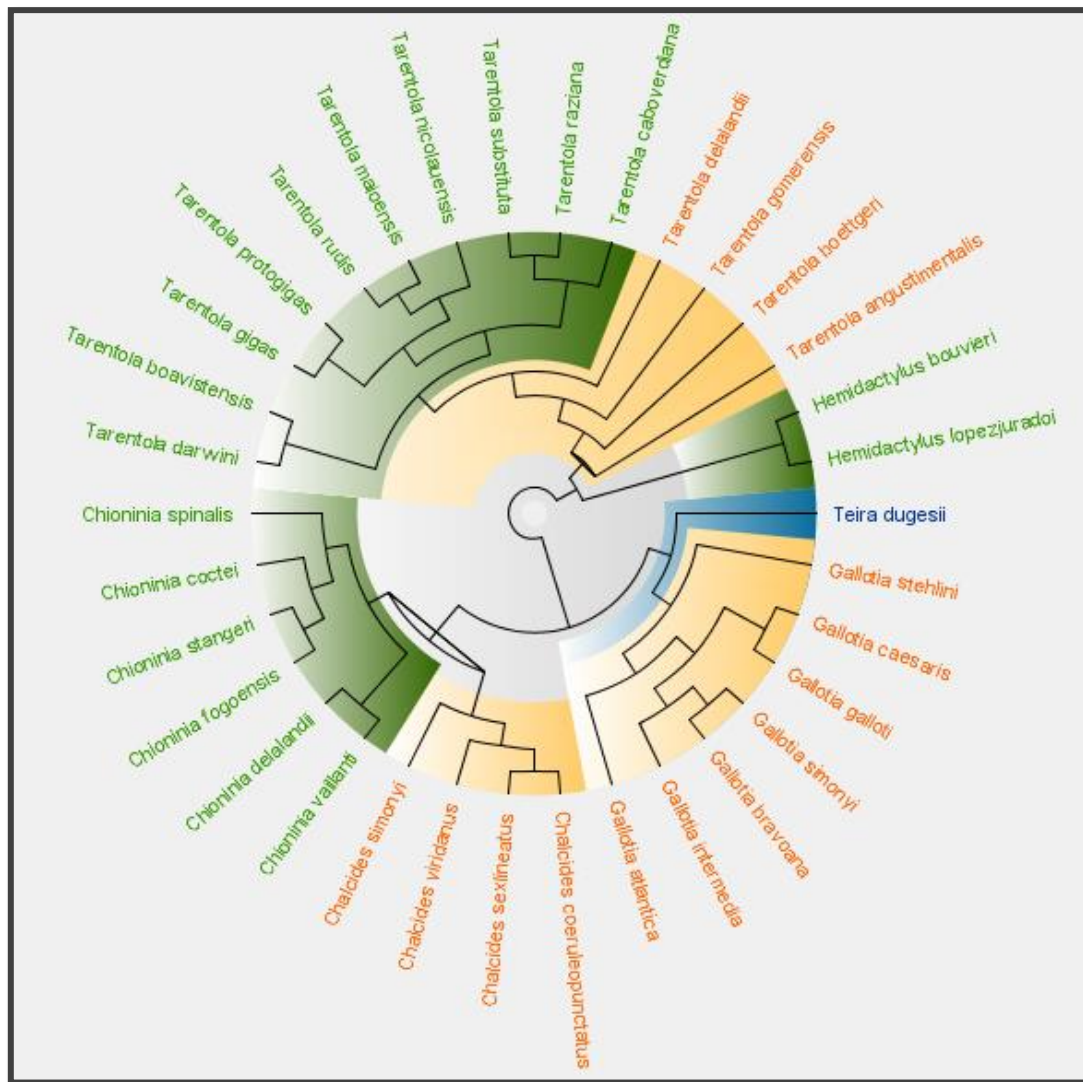


Figure 3.14 The reptile diversity in the Macaronesia region (excluding subspecies). Cape Verde specimens are highlighted in green; Canary Islands species are marked in orange; and Madeira originated reptiles are blue. Raw tree obtained from TimeTree (Kumar et al. 2017) and manipulated with FigTree

The endemic Cape Verdean genus *Chioninia*, has also a high number of described species and varied conserved status. *C. coctei* is considered extinct, and two others, *C. stangeri* and *C. vaillantii*, are threatened and endangered, respectively. This is a clear sign that these reptiles are also under an extreme pressure, particularly due to predation from invasive species and human influence, and should also be protected.

Lastly the *Gallotia* genus, endemic to the Canary Islands, has eight IUCN species, most of them critically endangered. Domestic animals such as cats, and the loss of habitat due to agriculture, may have caused the extinction *G. avariatae*, and as such preventive measures should be put in place before other species are lost.

It is clear that the Macaronesia region is rich in biodiversity, which is under a severe pressure due to human presence.

4. Discussion

4.1. Using Bioinformatic tools to study evolution on island species

Bioinformatics is changing the way we analyse biological data, and as we've seen throughout this study, it's extremely proficient and capable of doing in doing so. Using a great volume of data created across several studies, we review and assess all the endemic reptiles in Cape Verde, which are a big portion of the biodiversity found in Macaronesia.

An important highlight of this study is the importance of data sharing that is universally accepted nowadays, without it many software packages, algorithms and ideas wouldn't be possible, as each research had to start almost from scratch. Another invaluable resource for any bioinformatician is without a doubt the World Wide Web. Doing a simple search, we can find numerous research articles and their respective data; tutorials and courses in bioinformatics; and a vast amount of software packages for all kind of necessities one can find. Cloud services are also extremely helpful when we're working with substantial quantity of data, which can take a long time to analyse. Services like the ones provided in Cipres Gateway, allowed us to cut the running time of analysis substantially – analysis that would take a day or two in our personal computer, finished in a matter of hours in their servers thanks to their powerful hardware.

To better choose the several software packages in this study (see Table 1.1) we examined the several references that analysed the focused reptiles in our study, and other similar recent phylogenetic studies. Starting in the alignment algorithm, we opted for the MAFFT algorithm as it was available in Geneious, and provided accurate and rapid alignments. As it was built-in in Geneious, it was also easy to verify and correct possible errors that could occur during the aligning process.

Geneious, DNASp, BioEdit and MEGA, provided reliable and easy tools to verify the quality of the sequences along the study. DNASp also allowed for haplotype examinations, and the use of fastPHASE to remove nucleotide ambiguities when they were found.

To perform phylogenetic inference, we selected the most commonly used software's in today research studies: RAxML, MrBayes and BEAST. When calculating ML trees we used RAxMLGUI, for two main reasons: first it presented an interface that allowed for better control and input of parameters; and second because it had the option to upload directly to the Cipres Gateway, where the analysis were run. Both MrBayes and BEAST used BI to assess phylogenies, however BEAST is more adapted to use that methodology when calculating divergence time as well. As such, we preferred to use MrBayes for our "simple" phylogenetic calculations using BI; and BEAST to evaluate the dating of our phylogenies.

BEAST2 (Bouckaert et al. 2013) has been developed recently and it includes the option to install several helpful plugins that focus in a range of analysis. One of such plugins, *BEAST2 (Ogilvie et al. 2017), seems to be performing very well and seems to achieve better results comparing the previous versions of the software, as well as working faster than the base software program.

For the phylogenetic networks, we used SplitsTree4, as it provided an easy and yet resourceful way to obtain these results. Finally, when creating a visual representation of all the attained trees, we selected the FigTree software program, as it offered the most intuitive interface and methodology to manipulate and produce the final results.

4.2. Phylogenetic relationships among Cape Verde reptiles

The taxonomic and molecular data of the chosen genera was diverse, and as a result the obtained phylogenies varied in scope, consistency and possible colonization points.

Chioninia (Scincidae)

The *Chioninia* analysis differentiated itself from the other examined Cape Verde reptilians in gene and taxonomic diversity. The outgroups used for this group included other genus, due to the endemic nature of this genus. As such we included five different genera that were shown to be closer related in recent studies (Pyron et al. 2013; Metallinou et al. 2016; Karin et al. 2016; Zheng & Wiens 2016) – *Eumecia*, *Trachylepis*, *Heremites*, *Notomabuya* and *Dasia*. This group had also the biggest gene pool of the three, a total of 13 genes, composed of both nuclear and mitochondrial DNA. One of the challenges in this analysis was finding the most adequate genes and specimens to combine in our study, since the majority of available *Chioninia* sequences were of mitochondrial origin (Carranza et al. 2001; Carranza & Arnold 2003) and most outgroups had few mtDNA samples (Karin et al. 2016; Metallinou et al. 2016). Even after a careful selection and examination of the sequences, we ended up with a high percentage of missing data (73.2%), which was still lower than the one found in one of the other two analysis. As described in (Zheng & Wiens 2016), combining several gene matrixes usually results in a high amount of missing data, which by itself doesn't seem to negatively affect the overall results.

Even if this approach gave us the most reliable sequence set to start it, this genus proved to be challenging to analyse. During preliminar examinations, it was common to find low bootstrap values and inaccurate clustering, some outgroups being found inside the Cape Verde cluster. A species in particular, *Eumecia anchietae*, was commonly found alongside *Chioninia* species in single gene analysis, and as a result we analysed how the trees performed without this outgroup. In the end, there was no significant improvement in tree support without this species, and the outgroup was used in the combined analysis sets. There was also a high number of ambiguities in the DNA sequences, which added yet another stage to the examination process, as we analysed the set with and without those nucleotides to look at the several results. The results with and without discriminated haplotypes yielded similar results, and as such we used the set with differentiated haplotypes.

Overall the trees obtained for this group were reasonably supported within the *Chioninia* and outgroup clades, the smallest values being found closer to the root. The results were generally better using the BI approach, except when we used multiple outgroups or a midpoint root in the ML trees. This seems to corroborate with (Karin et al. 2016) claims that this is a molecularly complex group, and that a bigger sample of both mitochondrial and nuclear genes across the different species is necessary. Nevertheless, there's a clear indication of two separate groups within Cape Verde: one with *C. delalandii* and both subspecies of *C. vaillanti*, found in most southern islands; and another with the extinct *C. coctei* and the remaining species. Of all the outgroups, *Eumecia anchietae* was the closest taxa in normally rooted trees, which resembles the results of (Pyron et al. 2013) and (Zheng & Wiens 2016), which contrasts with the trees seen in (Karin et al. 2016) and (Metallinou et al. 2016). The two latter studies indicated a closer proximity to the genus *Trachylepis*, which is indeed what we see on the trees with multiple outgroups or midpoint groups.

***Hemidactylus* (Gekkonidae)**

The selected species used for the analysis of *Hemidactylus* were all found within this genus. Due to the high diversification and wide distribution of this group (Bansal & Karanth 2013), a first version of the analysis had a vast outgroup sample to understand which would be better suited for the tests. The final version contained all the found sequences of Cape Verde specimens and a total of eight outgroup species. Similarly to what we saw in the *Chioninia* analysis set, there was a bigger pool of mtDNA samples for the Cape Verde species, with only one study having used a nuclear gene (Arnold et al. 2008). Fortunately, the two mitochondrial genes sequenced for these specimens, 12S and cyt b, and the single nDNA strand, RAG2, were also available for most of the outgroup species, which allowed a more comprehensive examination.

The results for this genus were generally well supported, with most nodes reaching or nearly reaching the maximum bootstrap value for each of the phylogenetic methodologies. Across the created trees, the *Hemidactylus* clade found in Cape Verde was closer related to specimens found in São Tomé e Príncipe, with also a resemblance to species originated from Brazil, Trinidad Island and Kenya. These results match the ones found in the previous study by (Arnold et al. 2008), which also took in account morphological data. The resemblance between specimens divided by massive distances of both land and sea, is yet another proof of this genus remarkable dispersion capacity (Bansal & Karanth 2010). Regarding the Cape Verde clade, they're are notoriously two groups that spawn proximate islands – two strands of *H. boavistensis* are found across the neighbour islands of Boavista e Sal; *H. lopezjuradoi* occurs in the island of Fogo; and the remaining species spawn across the northern islands.

***Tarentola* (Phyllodactylidae)**

In the *Tarentola* analysis, we found a substantial number of DNA sequences for Cape Verdean specimens, namely for the cyt b gene, that were produced from several extensive sampling in previous works in the region (Carranza et al. 2000; Jesus et al. 2002; Vasconcelos et al. 2010; Vasconcelos et al. 2012). As such, we selected a smaller set of data, composed of 88 sequences in total, which maximized the number of available sequences of each specimen and represented most of the several lineages previously described in the genus.

The trees created for the single gene analysis were better supported when using the BI method and mtDNA sequences, with cyt b showing the best overall results. This goes in accordance to what was seen in the previous analysis, and it's a likely outcome of the wider sampling of mitochondrial material. Regarding the three concatenated analysis, the results were generally well supported, with only poorer results in the nDNA set. The mitochondrial and complete sequence sets showed a similar topology, the latter obtaining slightly better bootstrap values, possibility due to the introduction of nuclear information. As seen in previously works in this genus, the Cape Verdean clade shows a closer resemblance to specimens found in the Canary Islands and Morocco, and other African regions. Within the Cape Verde clade, two main groups are observable: one across the southeast part of the archipelago (Boavista, São Nicolau, Santiago e Fogo), with four species in each island, *T. boavistensis* in Boavista, *T. bocagei* in São Nicolau, *T. darwini* in Santiago, and *T. fogoensis* in Fogo; and the other more widely dispersed.

4.3. Patterns of islands diversity and diversification hypotheses

Taking in account all the obtained results in both classic phylogenetic analysis (ML and BI), divergence time dating and phylogenetic networks, we can create some hypotheses for diversification patterns in the Cape Verdean species. Overall, our results are in accordance previous studies done for each of the genus, and the recent work of (Grehan 2016) that proposes diversification patterns for several Macaronesian species. In all genera was also common to find similarities between species from neighbouring islands, which clearly showed a northern/southern colonization pattern along the islands. This fits the descriptions of (Caujapé-Castells et al. 2017) relating to the geographical history of the Cape Verde archipelago, namely that during the Last Glacier Maximum (LGM) the lower sea-level allowed for communication between several of islands, namely: Santo Antão, São Vicente and Santa Luzia, and by a lesser margin the island São Nicolau; the islands of Sal, Boavista and Maio, which are also the oldest, and connected to Santiago as well; and finally Brava and Fogo.

Chioninia (Scincidae)

The *Chioninia* genus seems to have originated 21.7 – 39.6 Mya and diverged 18.5-14.9 Mya, from specimens that came from African regions possibly in a single colonization event. After arriving to the islands two radiations spread out along the archipelago, some going to the northern islands of Santo Antão, São Vicente and Santa Luzia (and the nearby islets of Raso and Branco), and another branch to the islands of Brava, Santiago and Fogo. The proximity to the *Trachylepis* and *Heremites* specimens, seems to support the idea proposed by (Karin et al. 2016) that this genus colonized Africa, and then dispersed to the Cape Verde Islands and the South American subcontinent. Relatively to the single specimen of *Trachylepis* found outside Africa (*Trachylepis atlantica*), in the Brazilian island of Fernando de Noronha, its radiation seems to have been originated from continental Africa and not from the Cape Verde archipelago, as it was proposed by (Mausfeld et al. 2002), although further sampling of South American lineages is necessary to confirm this possibility.

Hemidactylus (Gekkonidae)

As we discussed before, the *Hemidactylus* genus is one of the most diverse and widespread reptiles found in the world, which seems to have emerged around 40 Mya in the Middle-East regions, with Cape Verdean lineages dating 21.2 – 26.0 Mya and dispersing along the archipelago 2.4 – 11.5 Mya. The Cape Verdean clade shows a close resemblance to species found in São Tomé e Príncipe, and perhaps more remarkably to those found in Brazil and Trinidad. As the South American species show a higher proximity to the Cape Verdean clade instead of the other African species, is possible that those specimens reached the American Continent from the Cape Verde Islands somewhere around 7.5 Mya. Similarly to what is seen in the other analysed genera, the results suggest a northern/southern pattern along the islands, something also commonly found in the endemic plants of the region as well (Romeiras et al. 2015), as well as some bird (e.g. *Falco tinnunculus alexandri*) and fish species. The southern clade found in the islands of Boavista and Sal, were perhaps the first to arrive at the archipelago, as these are the oldest islands, and likely made their way to the islands after.

***Tarentola* (Phyllodactylidae)**

Lastly, in *Tarentola*, we found the same colonization pattern described in both of the previous works with the species found in Macaronesia, that suggest a possible origin of the genus in West Indies around 45 Mya, and an origin and dispersion on the Cape Verde islands of 5.8 – 7.5 Mya and 5.1 – 5.8 Mya, respectively. The close relation between the Cape Verdean species and the ones found in the Canary Islands and Morocco indicate a common ancestral of these lineages. Within Cape Verde the connection between the obtained results and island geography is noticeable, with species coming closer to those found in nearby islands. The species of *T. boavistensis* is likely the most ancestral species on Cape Verde, which then made their way to the southern and northern islands and originated two separate clades.

The introduced specimen of *T. substituta* seems to be indeed related to the Cape Verde specimen with the same name, even though in some cases it was nested closer to *T. raziana* specimens. The study of (Rato et al. 2015) that first discovered this introduction suggested that it probably occurred during the ship voyages made between the two archipelagos, but the diversity shown in the phylogenetic trees, and particularly in the network analysis, indicate that perhaps their arrival to Azores was earlier and not due to humans, or maybe that their evolution occurred very rapidly since then, though more sampling is necessary to thoroughly evaluate this introduction.

5. Final remarks and perspectives

With the high number of methods, programs and data used across this study, it was possible to perform a comprehensive and interesting bioinformatic analysis. The variety of phylogenetic software's available today, as well as the current ideology of sharing data, software packages and tutorials online and for free, allow for a much sturdier base to start any analysis, and were detrimental to our work. The endemic reptiles of Cape Verde, were also an intriguing and fascinating case-study, as they provided a good amount of data and knowledge already available, but at the same time several gaps and obstacles that could be improved upon.

By working with three main phylogenetic methods – phylogenetic inference with ML and BI; divergence time estimation with BEAST and *BEAST2; and phylogenetic networks with SplitsTree4 – we extended the range of our conclusions by combining their results. As such, this workflow could possibly benefit many other cases, specifically those where there's a good source of dispersed data that was never analysed as a whole.

Future studies on this case-study could build upon our research by introducing the other three genera found in the Macaronesian region, *Chalcides*, *Gallotia* and *Teira*, as doing so would allow for a reassess on all the reptiles in the region. The *Chioninia* genus could also benefit from a wider sampling of genes (both mitochondrial and nuclear alike), as to allow for a more complete comparison with their relatives, as their relations are clearly very complex. Finally, an integration of biogeographical data could also prove useful, especially to better clarify some of the radiation patters along the Cape Verde archipelago and the colonization of these islands.

6. References

- Aldrich, J., 1997. R.A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3), pp.162–176.
- Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403–410.
- Arechavaleta, M., Zurita, N., Marrero, M. C. & Martín, 2005. *Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres)* M. A. Hernández, ed.,
- Arnold, E.N. et al., 2008. Systematics, biogeography and evolution of the endemic *Hemidactylus* geckos (Reptilia, Squamata, Gekkonidae) of the Cape Verde Islands: Based on morphology and mitochondrial and nuclear DNA sequences. *Zoologica Scripta*, 37(6), pp.619–636.
- Arnold, N. & Ovenden, D., 2002. *Reptiles and Amphibians of Britain and Europe (Collins Field Guide)*, Collins.
- Askira Gelman, I., 2011. GIGO or not GIGO. *Journal of Data and Information Quality*, 2(2), pp.1–27.
- Austin, J.J. & Arnold, E.N., 2006. Using ancient and recent DNA to explore relationships of extinct and endangered *Leiolopisma* skinks (Reptilia: Scincidae) in the Mascarene islands. *Molecular Phylogenetics and Evolution*, 39(2), pp.503–511.
- Bansal, R. & Karanth, K.P., 2010. Molecular phylogeny of *Hemidactylus* geckos (Squamata: Gekkonidae) of the Indian subcontinent reveals a unique Indian radiation and an Indian origin of Asian house geckos. *Molecular Phylogenetics and Evolution*, 57(1), pp.459–465.
- Bansal, R. & Karanth, K.P., 2013. Phylogenetic Analysis and Molecular Dating Suggest That *Hemidactylus anamallensis* Is Not a Member of the *Hemidactylus* Radiation and Has an Ancient Late Cretaceous Origin. *PLoS ONE*, 8(5).
- Bauer, A.M., Böhme, W. & Weitschat, W., 2005. An Early Eocene gecko from Baltic amber and its implications for the evolution of gecko adhesion. *Journal of Zoology*, 265(4), pp.327–332.
- Bellard, C. et al., 2014. Vulnerability of biodiversity hotspots to global change. *Global Ecology and Biogeography*, 23(12), pp.1376–1386.
- Benton, M., Donoghue, P.C.J. & Asher, R.J., 2009. Calibrating and constraining molecular clocks. *The Timetree of Life*, pp.35–86.
- Borsuk-Bialynicka, M., 1984. Anguimorphans and related lizards from the Late Cretaceous of the Gobi Desert, Mongolia. *Palaeontologia Polonica*, 46, pp.5–105.
- Bouckaert, R. et al., 2013. Beast2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.*, 10(4), p.1003537.
- Bryan C. Carstens, J.D.M., 2013. Phylogenetic Model Choice: Justifying a Species Tree or Concatenation Analysis. *Journal of Phylogenetics & Evolutionary Biology*, 1(3), pp.1–8.
- Bryant, D. & Moulton, V., 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, 21(2), pp.255–265.
- Bryant, L.J., 1989. Non-dinosaurian lower vertebrates across the Cretaceous-Tertiary boundary in northeastern Montana. *Univ. Calif. Pubs. Geol. Sci*, 134, pp.1–107.
- Carranza, S. et al., 2000. Long-distance colonization and radiation in gekkonid lizards, *Tarentola* (Reptilia: Gekkonidae), revealed by mitochondrial DNA sequences. *Proceedings. Biological sciences / The Royal Society*, 267(1444), pp.637–49.
- Carranza, S. et al., 2001. Parallel gigantism and complex colonization patterns in the Cape Verde scincid lizards *Mabuya* and *Macroscoincus* (Reptilia: Scincidae) revealed by mitochondrial DNA sequences. *Proceedings. Biological sciences / The Royal Society*, 268(1476), pp.1595–1603.

- Carranza, S. et al., 2002. Relationships and evolution of the North African geckos, *Geckonia* and *Tarentola* (Reptilia: Gekkonidae), based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, 23(2), pp.244–256.
- Carranza, S. & Arnold, E.N., 2003. lizards (Reptilia, Scincidae) crossed the Atlantic twice. *Systematics and Biodiversity*, 1(2), pp.275–282.
- Carranza, S. & Arnold, E.N., 2006. Systematics, biogeography, and evolution of *Hemidactylus* geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 38(2), pp.531–545.
- Caujapé-Castells, J. et al., 2010. Conservation of oceanic island floras: Present and future global challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, 12(2), pp.107–129.
- Caujapé-Castells, J. et al., 2017. Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the Canarian endemic flora. *Perspectives in Plant Ecology, Evolution and Systematics*, 27(March), pp.9–22.
- Clark, K. et al., 2016. GenBank. *Nucleic Acids Research*, 44(D1), pp.D67–D72.
- Clement, M., Posada, D. & Crandall, K.A., 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9(10), pp.1657–1659.
- Conrad, J.L., 2008. *Phylogeny and systematics of squamata (reptilia) based on morphology*,
- Cox, S.C., Carranza, S. & Brown, R.P., 2010. Divergence times and colonization of the Canary Islands by *Gallotia* lizards. *Molecular Phylogenetics and Evolution*, 56(2), pp.747–757.
- Crottini, A. et al., 2012. Vertebrate time-tree elucidates the biogeographic pattern of a major biotic change around the K-T boundary in Madagascar. *Proceedings of the National Academy of Sciences*, 109(14), pp.5358–5363.
- Cunha, R.L. et al., 2014. Evolution at a different pace: Distinctive phylogenetic patterns of cone snails from two ancient oceanic archipelagos. *Systematic Biology*, 63(6), pp.971–987.
- Darwin, C., 2009. Coleção PLANETA DARWIN: © Planeta Vivo. *Livro*, 1(5), p.42.
- Denton, R.K. & O'Neill, R.C., 1995. *Prototeius stageri*, gen. et sp. nov., a new teiid lizard from the upper cretaceous marshalltown formation of new jersey, with a preliminary phylogenetic revision of the teiidae. *Journal of Vertebrate Paleontology*, 15(2), pp.235–253.
- Donoghue, P.C.J. & Benton, M.J., 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution*, 22(8), pp.424–431.
- Doucelance, R. et al., 2003. Pb-Sr-He isotope and trace element geochemistry of the Cape Verde Archipelago. *Geochimica et Cosmochimica Acta*, 67(19), pp.3717–3733.
- Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1), p.214.
- Duarte, M.C. & Romeiras, M.M., 2009. Cape Verde Islands. In *Encyclopedia of Islands*. University of California Press, pp. 143–148.
- Duda, T.F. & Rolán, E., 2005. Explosive radiation of Cape Verde Conus, a marine species flock. *Molecular Ecology*, 14(1), pp.267–272.
- Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), pp.1792–1797.
- Edwards, S. V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution*, 63(1), pp.1–19.
- EMBL, SIB Swiss Institute of Bioinformatics & Protein Information Resource (PIR), 2013. UniProt. In *Nucleic acids research*. p. 41: D43-D47.

- Estes, R., 1969. A scincoid lizard from the Cretaceous and Paleocene of Montana. *Breviora. Univ. Calif. Pubs. Geol. Sci.*, 331, pp.1–9.
- Estes, R., 1964. Fossil vertebrates from the late Cretaceous Lance Formation, eastern Wyoming. *Univ. Calif. Pubs. Geol. Sci.*, 49, pp.1–180.
- Evans, S.E., 2003. At the feet of the dinosaurs: the early history and radiation of lizards. *Biological reviews of the Cambridge Philosophical Society*, 78(4), pp.513–551.
- Farris, J.S., 1970. Methods for Computing Wagner Trees. *Systematic Biology*, 19(1), pp.83–92.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), pp.368–76.
- Fienberg, S.E., 2006. When did Bayesian inference become “Bayesian”? *Bayesian Analysis*, 1(1 A), pp.1–40.
- Fitch, W.M., 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. *Systematic Biology*, 20(4), pp.406–416.
- Francisco-Ortega, J. et al., 2000. Plant genetic diversity in the Canary Islands: A conservation perspective. *American Journal of Botany*, 87(7), pp.909–919.
- Gamble, T. et al., 2011. Coming to America: multiple origins of New World geckos: Origins of New World geckos. *Journal of Evolutionary Biology*, 24(2), pp.231–244.
- Gamble, T. et al., 2008a. Evidence for Gondwanan vicariance in an ancient clade of gecko lizards. *Journal of Biogeography*, 35(1), pp.88–104.
- Gamble, T. et al., 2008b. Out of the blue: a novel, trans-Atlantic clade of geckos (Gekkota, Squamata). *Zoologica Scripta*, 37(4), pp.355–366.
- Gillespie, R.G. & Clague, D.A., 2009. *Encyclopedia of islands*, Berkeley: Univ. of California Press.
- Grehan, J.R., 2016. Biogeographic relationships between Macaronesia and the Americas. *Australian Systematic Botany*, 29(6), pp.447–472.
- Grismer, L.L. et al., 2015. Repeated evolution of sympatric, palaeoendemic species in closely related, co-distributed lineages of Hemiphyllodactylus Bleeker, 1860 (Squamata: Gekkonidae) across a sky-island archipelago in Peninsular Malaysia. *Zoological Journal of the Linnean Society*, 174(4), pp.859–876.
- Guindon, S. et al., 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, 537, pp.113–137.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98. Available at: <http://jwbrown.mbio.ncsu.edu/JWB/papers/1999Hall1.pdf>.
- Ho, S.Y.M., 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, 38(4), pp.409–414.
- Ho, S.Y.W. & Duchêne, S., 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24), pp.5947–5965.
- Ho, S.Y.W. & Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58(3), pp.367–380.
- Huelsenbeck, J.P. & Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), pp.754–755.
- Hugall, A.F., Foster, R. & Lee, M.S.Y., 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology*, 56(4), pp.543–563.

- Huson, D.H. & Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), pp.254–267.
- Huson, D.H. & Bryant, D., 2001. Application of Phylogenetic Networks in Evolutionary Studies.
- Huson, D.H. & Scornavacca, C., 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, 61(6), pp.1061–1067.
- IUCN, 2017. The IUCN Red List of Threatened Species Version 2017-2. <<http://www.iucnredlist.org>>. Downloaded on 14 September 2017.
- Jesus, J. et al., 2002. Relationships of Tarentola (Reptilia: Gekkonidae) from the Cape Verde Islands estimated from DNA sequence data. *Amphibia-Reptilia*, 23, pp.47–54.
- Karin, B.R. et al., 2016. Resolving the higher-order phylogenetic relationships of the circumtropical Mabuya group (Squamata: Scincidae): An out-of-Asia diversification. *Molecular Phylogenetics and Evolution*, 102, pp.220–232.
- Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.
- Kearse, M. et al., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.
- Keqin, G. & Norell, M. a, 2000. Taxonomic Composition And Systematics Of Late Cretaceous Lizard Assemblages From Ukhaa Tolgod And Adjacent Localities, Mongolian Gobi Desert. *Bulletin of the American Museum of Natural History New York Bull Am Mus Nat Hist N Y*, 249(1), pp.1–118.
- Kluge, A.G., 1969. The evolution and geographical origin of the New World Hemidactylus mabouia-brookii complex (Gekkonidae, Sauria). *Miscellaneous Publications of the Museum of Zoology, University of Michigan*, (138), pp.1–78.
- Kodama, Y. et al., 2012. The DNA Data Bank of Japan launches a new resource, the DDBJ Omics Archive of functional genomics experiments. *Nucleic Acids Research*, 40(D1).
- Krause, D.W., Evans, S.E. & Gao, K., 2003. First definitive record of Mesozoic lizards from Madagascar. *Journal of Vertebrate Paleontology*, 23(4), pp.842–856.
- Kulikova, T. et al., 2007. EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Research*, 35(SUPPL. 1).
- Kumar, S. et al., 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, 34(7), pp.1812–1819.
- Kumar, S., Stecher, G. & Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, 33(7), pp.1870–1874.
- Lanfear, R. et al., 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular biology and evolution*, p.msw260.
- Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947–2948.
- Letunic, I. & Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, 44(W1), pp.W242–W245.
- Librado, P. & Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), pp.1451–1452.

- Machín-Sánchez, M. et al., 2014. A combined barcode and morphological approach to the systematics and biogeography of *Laurencia pyramidalis* and *Laurenciella marilzae* (Rhodophyta). *European Journal of Phycology*, 49(1), pp.115–127.
- Maddison, W.P. & Maddison, D.R., 2008. Mesquite: A modular system for evolutionary analysis. *Evolution*, 62, pp.1103–1118.
- Marsh, O.C., 1892. Notice of new reptiles from the Laramie Formation. *Am. J. Sci*, 257, pp.449–453.
- Mausfeld, P. et al., 2002. Phylogenetic affinities of *Mabuya atlantica* Schmidt, 1945, endemic to the Atlantic ocean archipelago of Fernando de Noronha (Brazil): necessity of partitioning the genus *Mabuya* Fitzinger, 1826 (Scincidae: Lygosominae). *Zoologischer Anzeiger*, 241, pp.281–293.
- Metallinou, M. et al., 2016. A single origin of extreme matrotrophy in African mabuyine skinks. *Biology Letters*, 12(8), p.20160430.
- Metallinou, M. et al., 2012. Conquering the Sahara and Arabian deserts: systematics and biogeography of *Stenodactylus* geckos (Reptilia: Gekkonidae). *BMC Evolutionary Biology*, 12(1), p.258.
- Miller, M.A., Pfeiffer, W. & Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop, GCE 2010*.
- Miralles, A. et al., 2011. An integrative taxonomic revision of the Cape Verdean skinks (Squamata, Scincidae). *Zoologica Scripta*, 40(1), pp.16–44.
- Moore, J.H., 2007. Bioinformatics. *Journal of Cellular Physiology*, 213(2), pp.365–369.
- Mulcahy, D.G. et al., 2012. Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution*, 65(3), pp.974–991.
- Nydam, R.L., 2000. A new taxon of helodermatid-like lizard from the Albian–Cenomanian of Utah. *Journal of Vertebrate Paleontology*, 20(2), pp.285–294.
- Ogilvie, H.A., Bouckaert, R.R. & Drummond, A.J., 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, 34(8), pp.2101–2114.
- Orellana, J.A.V. et al., 2016. A phytosociological classification of bramble communities in Macaronesia (Azores, Madeira and Canary Islands). *Phytocoenologia*, 46(2), pp.185–200.
- Pina-Martins, F. & Paulo, O.S., 2008. Concatenator: Sequence data matrices handling made easy. *Molecular Ecology Resources*, 8(6), pp.1254–1255.
- Planet, P.J., 2006. Tree disagreement: Measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics*, 39(1 SPEC. ISS.), pp.86–102.
- Posada, D., 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, 25(7), pp.1253–1256.
- Pyron, R., Burbrink, F.T. & Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, 13(1), p.93.
- Rambaut, A., 2009. FigTree, a graphical viewer of phylogenetic trees. *Institute of Evolutionary Biology University of Edinburgh*.
- Rambaut, A. & Drummond, A.J., 2013. Tracer v1.6.
- Rato, C. et al., 2015. First record of *Tarentola substituta* JOGER, 1984, and genetic identification of *Tarentola mauritanica* (LINNAEUS, 1758). *Herpetozoa*, 27, pp.182–187.
- Rato, C., Carranza, S. & Harris, D.J., 2012. Evolutionary history of the genus *Tarentola* (Gekkota: Phyllodactylidae) from the Mediterranean Basin, estimated using multilocus sequence data. *BMC Evolutionary Biology*, 12(1), p.14.

- Rocha, S., Carretero, M.A. & Harris, D.J., 2005. Diversity and phylogenetic relationships of Hemidactylus geckos from the Comoro islands. *Molecular Phylogenetics and Evolution*, 35(1), pp.292–299.
- Romeiras, M.M. et al., 2016. IUCN Red List assessment of the Cape Verde endemic flora: Towards a global strategy for plant conservation in Macaronesia. *Botanical Journal of the Linnean Society*, 180(3), pp.413–425.
- Romeiras, M.M., Monteiro, F., Cristina Duarte, M., et al., 2015. Patterns of genetic diversity in three plant lineages endemic to the Cape Verde Islands. *AoB Plants*, 7(1), pp.1–11.
- Romeiras, M.M., Monteiro, F., Duarte, M.C., et al., 2015. Patterns of genetic diversity in three plant lineages endemic to the Cape Verde Islands. *AoB Plants*, 7, pp.1–11.
- Romeiras, M.M. et al., 2016. Species Conservation Assessments in Oceanic Islands: the Consequences of Precautionary Versus Evidentiary Attitudes. *Conservation Letters*, 9(4), pp.275–280.
- Roux, J.P., 2004. (Pteropsida: Dryopteridaceae) in Macaronesia. *Systematics and Biodiversity*, 2(2), pp.147–161.
- Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), pp.406–425.
- Sanderson, M.J., 2003. Evolution and Divergence Times in the Absence of a Molecular Clock. *Evolution*, 19(2), pp.301–302.
- Santos, C. et al., 2010. Mitochondrial DNA patterns in the macaronesia Islands: Variation within and among archipelagos. *American Journal of Physical Anthropology*, 141(4), pp.610–619.
- Shenbagarathai, R., 2007. Bioinformatics tools and applications. In *Current Science*. p. 898.
- Silvestro, D. & Michalak, I., 2012. RaxmlGUI: A graphical front-end for RAxML. *Organisms Diversity and Evolution*, 12(4), pp.335–337.
- Skinner, A., Hugall, A.F. & Hutchinson, M.N., 2011. Lygosomine phylogeny and the origins of Australian scincid lizards. *Journal of Biogeography*, 38(6), pp.1044–1058.
- Šmíd, J. et al., 2013. Out of Arabia: A Complex Biogeographic History of Multiple Vicariance and Dispersal Events in the Gecko Genus Hemidactylus (Reptilia: Gekkonidae). *PLoS ONE*, 8(5).
- Sokal, R.R. & Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, (38), pp.1409–1438.
- Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), pp.1312–1313.
- Stratton, M.R., Campbell, P.J. & Andrew F, P., 2009. The cancer genome. *Nature*, 458(7239), pp.719–724.
- Sues, H.D. & Olsen, P.E., 1990. Triassic vertebrates of Gondwanan aspect from the Richmond Basin of Virginia. *Science*, 249(4972), pp.1020–1023.
- Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. *Sinauer Associates, Sunderland, Massachusetts*.
- Takayama, K. et al., 2015. Genetic consequences of cladogenetic vs. anagenetic speciation in endemic plants of oceanic islands. *AoB Plants*, 7, p.plv102.
- Uetz, P., 2010. The original descriptions of reptiles. *Zootaxa*, 68(2334), pp.59–68.
- Vasconcelos, R. et al., 2012. An integrative taxonomic revision of the Tarentola geckos (Squamata, Phyllodactylidae) of the Cape Verde Islands. *Zoological Journal of the Linnean Society*, 164(2), pp.328–360.

- Vasconcelos, R. et al., 2013. Review of the distribution and conservation status of the terrestrial reptiles of the Cape Verde Islands. *Oryx*, 47(1), pp.77–87.
- Vasconcelos, R., Carranza, S. & James Harris, D., 2010. Insight into an island radiation: The Tarentola geckos of the Cape Verde archipelago. *Journal of Biogeography*, 37(6), pp.1047–1060.
- Vasconcelos, R.C.S., 2010. Integrative Approaches to the Systematics and Conservation of the Reptiles of the Cape Verde Islands Integrative Approaches to the Systematics and Conservation of the Reptiles of the Cape Verde Islands. , p.365.
- Vidal, N. & Hedges, S.B., 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *Comptes Rendus - Biologies*, 328(10–11), pp.1000–1008.
- Wiens, J.J., Brandley, M.C. & Reeder, T.W., 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60(1), pp.123–141.
- Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), pp.1586–1591.
- Yang, Z. & Rannala, B., 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13(5), pp.303–314.
- Zheng, Y. & Wiens, J.J., 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, 94, pp.537–547.
- Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. *Philosophy*, p.115.

SUPPORTING INFORMATION

Appendix I – List of the endemic reptiles from Macaronesia assessed by the IUCN Red List, with detailed information on their habitat and ecology, main threats and conservation actions.

Appendix II – List of the selected set of DNA sequences used in this study, organized by their reference, taxa, code (as seen in the referenced study) and accession numbers for each gene, separated by brackets.

Appendix III – Microsoft Excel Spreadsheet containing the complete list of sequences used in the study and their corresponding information.

Appendix IV – Microsoft Excel Spreadsheet containing comprehensive data for each of the genes in this study.

Appendix V – Microsoft Excel Spreadsheet detailing the results attained in PartitionFinder.

Appendix VI – All phylogenetic trees created for *Chioninia* without using external outgroups.

Appendix VII – All phylogenetic trees created for *Hemidactylus* without using external outgroups.

Appendix VIII – All phylogenetic trees created for *Tarentola* without using external outgroups.

Appendix IX – All phylogenetic trees created for *Chioninia* using external outgroups.

Appendix X – All phylogenetic trees created for *Hemidactylus* using external outgroups.

Appendix XI – All phylogenetic trees created for *Tarentola* using external outgroups.

Appendix I - The endemic reptiles from Macaronesia assessed by the IUCN Red List, with detailed information on their habitat and ecology, main threats and conservation actions.

Location / Family / Species	Red List Category	Habitat & Ecology	Threats	Conservation Actions
Canary Islands				
Lacertidae				
<i>Gallotia atlantica</i>	LC	Wide range of dry, open habitats	No major threats	It is protected by national legislation, and it is present in several protected areas
<i>Gallotia auaritae</i>	CR (Possibly Extinct)	Through the littoral zone of La Palma, living in xerophytic vegetation	Introduced cats; Consumption by people; Conversion of land to agricultural use	No conservation measures
<i>Gallotia bravoana</i>	CR	This species is once widespread in many habitat types on La Gomera. It is now restricted to dry cliffs with sparse vegetation	The main threats to this species are predation by feral cats, and rock falls within its restricted range	The species is protected by international legislation. A captive breeding program has been established on La Gomera, and a species recovery plan is in place
<i>Gallotia caesaris</i>	LC	Open areas; Shrubland; Cultivated and urban areas	No major threats	Protected by international legislation; Occurs in several protected areas
<i>Gallotia galloti</i>	LC	This species is common in various open, rocky and shrubland habitats. It is commonly associated with stone walls in cultivated areas	No major threats	It is protected by international legislation and is present in several protected areas. The subspecies <i>insulanagae</i> is considered Near Threatened on the Spanish Red List
<i>Gallotia intermedia</i>	CR	This species inhabits rugged terrain, with rocks and boulders, often found on small rock ledges with sparse vegetation	The main threat to this species is predation by feral cats and, to a lesser degree, by rats	Measures to control access by cats to some of the remaining populations. A recovery action plan has been developed for this species
<i>Gallotia simonyi</i>	CR	It is now confined to a small area of cliffs with sparse vegetation	The major threat to this species is predation by feral cats, and possibly also by dogs and rats	A recovery plan for this species has been developed, and captive-breeding and reintroduction programs are in place
<i>Gallotia stehlini</i>	LC	This species is present in a wide variety of habitats. It occurs in open areas, shrubland, rocky humid gorges	Although this species is generally still abundant, it is impacted by predation by feral cats and rats.	This species is protected by international legislation. It occurs in several protected areas
Scincidae				
<i>Chalcides sexlineatus</i>	LC	It is found in a wide variety of habitats including, humid meadows and valleys, arid gullies and sandy areas, woodland and plantations	There is some collection of this species for the pet trade, but not at a level to constitute a threat to the species	This species is protected by national and international (Bern Convention) legislation. It occurs in several protected areas
<i>Chalcides viridanus</i>	LC	It is found in most coastal, arid and moist habitats, being only largely absent from densely wooded areas	It is locally threatened by predation by cats.	This species is listed on the Bern Convention. It occurs in several protected areas
<i>Chalcides simonyi</i>	EN	It lives in fields, orchards, gardens and in rocky areas. It is usually found in areas with well-formed soil	The main threat is fragmentation of populations due to climatic changes (increasing aridity), loss of soil due to erosion	It occurs in some protected areas
Phyllodactylidae				
<i>Tarentola angustimentalis</i>	LC	This species is found in most habitats within its range	No major threats	This species is protected by international legislation and occurs in several protected areas
<i>Tarentola delalandii</i>	LC	This species is found in a wide variety of habitats	No major threats	This species is protected by international legislation. It occurs in several protected areas
<i>Tarentola boettgeri</i>	LC	This species is found in rocky areas, often near to the coast.	It is preyed upon by introduced mammal species such as cats and hedgehogs	It is protected by international legislation and occurs in several protected areas
<i>Tarentola gomerensis</i>	LC	This species is found in rocky areas, stone walls, disturbed habitats, plantations and urban areas	No major threats	It is protected by international legislation and in some of protected areas

Location / Family / Species	Red List Category	Habitat & Ecology	Threats	Conservation Actions
Cape Verde				
Scincidae				
<i>Chioninia coctei</i>	Extinct	Inhabited rocky clefts within seabird colonies	Rats, cats and dogs represented the major predators; Used as a food source by humans	No species-specific conservation measures
<i>Chioninia delalandii</i>	LC	Rocky walls including the remains of agricultural structures and beneath rocks; Also found in waste dumps	No species-specific threats	No species-specific conservation measures
<i>Chioninia fogoensis</i>	LC	Agrarian environment; Some remnants of dry forests and scrub; Rocky walls, under rocks in dry, barren areas and in anthropogenic habitats including croplands, plantations and villages	Habitat that has been extensively altered from dry savannah and grasslands into desert scrub; Introduced rats, sheeps and cattle	No species-specific conservation measures; Occurs in planned protected areas
<i>Chioninia nicolauensis</i>	LC	Dry and humid lowland habitats and in highlands; Occurs in vegetated areas, on rocky walls and beneath rocks in more barren habitats; Anthropogenic situations as croplands, plantations and villages	No species-specific threats	No species-specific conservation measures
<i>Chioninia spinalis</i>	LC	From vegetated dunes on salt flats to sandy deserts; Saltmarsh vegetation; Barren rocky areas; Agricultural stone walls; Under rocks and, in sandy areas, fallen palm trees and limestone plates	Susceptible to stochastic events as a consequence of the highly restricted range of some subpopulations and their physical and genetic isolation	No species-specific conservation measures
<i>Chioninia stangeri</i>	Near Threatened	Scrubland with a preference for sandy soils and dune vegetation, as well as on rocky walls and beneath rocks	Introduced rodents; Predation by domestic cats; Introduction of exotic lizard	First Cape Verde Red List identified it as being Rare in Branco and Raso and Data Deficient in São Vicente
<i>Chioninia vaillantii</i>	EN	Dry scrub; Conifer and eucalypt plantations; Remains of stone walls covered by vegetation in abandoned agricultural land in more humid areas	Road developments; Agricultural developments; Human pressure	First Cape Verde Red List determined this species to be in urgent need of protection; Protected under Cape Verde national legislation
Gekkonidae				
<i>Hemidactylus boavistensis</i>	Near Threatened	Very arid, open areas with little vegetation; Rocky and sandy areas, including ruined stone walls and sand dunes	Ongoing drought conditions	No species-specific conservation measures
<i>Hemidactylus bouvieri</i>	CR	Restricted to humid, high montane areas between 600-700 m in elevation; Humid valleys from 250 m all on São Nicolau; Along dry streams in inland areas with dense vegetation and uses cavities in rock or tree roots as shelter sites on Raso	Extensive vegetation loss (overgrazing of the remaining vegetation by goats and woodcutting by local people); Predation by cats and rats	Listed as threatened and in need of urgent conservation attention in the first national Red List of Cape Verde
<i>Hemidactylus lopezjuradoi</i>	DD	Nocturnal gecko restricted to specialized, humid habitats, known only from well-vegetated areas in deep valleys around 300 m in elevation	Extensive vegetation loss (overgrazing of the remaining vegetation by goats and woodcutting by local people); Predation by cats and rats; Stochastic events	No species-specific conservation measures
Phyllodactylidae				
<i>Tarentola boavistensis</i>	VU	Nocturnal gecko found beneath rocks on barren rocky or gravelly plains in arid areas or those with sparse vegetation	Ongoing drought conditions	No species-specific conservation measures
<i>Tarentola bocagei</i>	LC	Nocturnal gecko found in rocky, barren plains mainly in arid areas at low elevations	No major threats	No species-specific conservation measures

Location / Family / Species	Red List Category	Habitat & Ecology	Threats	Conservation Actions
Cape Verde				
Phyllodactylidae				
<i>Tarentola caboverdiana</i>	LC	Rocky, arid plains where this nocturnal gecko shelters under rocks during the day	Unknown	No species-specific conservation measures
<i>Tarentola darwini</i>	LC	Nocturnal gecko associated with barren, rocky plains in lowlands and with arid areas where it shelters beneath stones during the day	No major threats	No species-specific conservation measures
<i>Tarentola fogoensis</i>	LC	Rocky, arid plains and arid areas while it can occur in somewhat higher elevation, more humid and sandier areas	No major threats	No species-specific conservation measures
<i>Tarentola gigas</i>	EN	Cliff-holes and burrows used by seabirds, typically beneath sandstone blocks	Infrastructure development, improved access to refrigeration technology and the depletion of fish stocks following an increase in international fishing prompted local fisherman to extend their fishing expeditions in the islets, increasing human pressure	Listed as Endangered in the First Red List of Cape Verde which considered this species in need or urgent protection
<i>Tarentola maioensis</i>	LC	Under rocks on rocky barren plains and in sandy areas under calcareous plates	Droughts	No species-specific conservation measures
<i>Tarentola nicolauensis</i>	LC	Under rocks on rocky, barren arid plains and other arid areas; It can occur in somewhat higher elevation, more humid and sandier areas	No major threats	No species-specific conservation measures
<i>Tarentola protogigas</i>	LC	Predominantly nocturnal species found beneath rocks in rocky, barren areas and also on rocky walls	Droughts; Volcanoes; Earthquakes	No species-specific conservation measures
<i>Tarentola raziana</i>	Near Threatened	Under rocks on rocky barren plains and other arid areas	Introduced rodents; Infrastructure development; Increasing of human pressure; Climate change	Listed as Rare in the First Red List of Cape Verde
<i>Tarentola rudis</i>	DD	Predominantly nocturnal species found under rocks in rocky barren and arid areas	High rate of hybridization	It is included in conservation legislation
<i>Tarentola substituta</i>	LC	Under rocks in rocky barren plains and other arid areas where this strictly nocturnal gecko shelters under large stones during the day	No major threats	No species-specific conservation measures
Madeira				
Lacertidae				
<i>Teira dugesii</i>	LC	This species is generally found in arid areas with sparse vegetation	No major threats	This species occurs in some protected areas. It is listed on Appendix II of the Bern Convention

Appendix II The selected set of DNA sequences used in this study, organized by their reference, taxa, code (as seen in the referenced study) and accession numbers for each gene, separated by brackets. For *Hemidactylus* and *Tarentola* some of the cytochrome b sequences were divided in two parts and others were composed of a single one with the same region. As such, this genes' accession numbers are split across three columns: CytB1, CytB2, CytB, composed of first half's, second half's and complete sets, respectively.

Chioninia			
Reference	Taxa	Code	Accession No. (CytB/12S/Cmos/16S/COI/RAG2/BDNF/BRCA1/BRCA2/EXPH5/KIF24/MC1R/RAG1)
(Carranza et al. 2001)	<i>Chioninia coctei</i>	M241	AF280314/AF280182/AF280248
	<i>Chioninia coctei</i>	M242	AF280313/AF280181/AF280247
	<i>Chioninia delalandii</i>	M41	AF280326/AF280194/AF280260/HQ316444
	<i>Chioninia delalandii</i>	M22	AF280325/AF280193/AF280259/HQ316443
	<i>Chioninia delalandii</i>	M19	AF280315/AF280183/HQ316441
	<i>Chioninia delalandii</i>	M40	AF280324/AF280192/HQ316446
	<i>Chioninia delalandii</i>	M21	AF280322/AF280190/HQ316447
	<i>Chioninia delalandii</i>	M68	AF280321/AF280189/HQ316438
	<i>Chioninia delalandii</i>	M67	AF280320/AF280188/HQ316439
	<i>Chioninia fogoensis</i>	M64	AF280312/AF280180/AF280246/HQ316459
	<i>Chioninia nicolauensis</i>	M35	AF280306/AF280174/AF280240/HQ316463
	<i>Chioninia spinalis boavistensis</i>	M27	AF280276/AF280144/AF280210/HQ316466
	<i>Chioninia spinalis boavistensis</i>	M26	AF280275/AF280143/AF280209/HQ316465
	<i>Chioninia spinalis maioensis</i>	M54	AF280292/AF280160/AF280226/HQ316481
	<i>Chioninia spinalis salensis</i>	M25	AF280285/AF280153/AF280219/HQ316468
	<i>Chioninia spinalis salensis</i>	M9	AF280284/AF280152/AF280218/HQ316469
	<i>Chioninia spinalis santiagoensis</i>	M28	AF280279/AF280147/AF280213/HQ316473
	<i>Chioninia spinalis santiagoensis</i>	M11	AF280278/AF280146/AF280212/HQ316472
	<i>Chioninia spinalis spinalis</i>	M76	AF280289/AF280157/AF280223/HQ316478
	<i>Chioninia spinalis spinalis</i>	M14	AF280288/AF280156/AF280222/HQ316479
	<i>Chioninia stangeri</i>	M66	AF280303/AF280171/AF280237/HQ316455
	<i>Chioninia stangeri</i>	M65	AF280302/AF280170/AF280236/HQ316454
	<i>Chioninia stangeri</i>	M31	AF280295/AF280163/AF280229/HQ316451
	<i>Chioninia stangeri</i>	M13	AF280294/AF280162/AF280228/HQ316452
	<i>Chioninia stangeri</i>	M33	AF280298/AF280166/AF280232/HQ316450
	<i>Chioninia stangeri</i>	M15	AF280297/AF280165/AF280231/HQ316449
	<i>Chioninia stangeri</i>	M18	AF280301/AF280169/AF280235/HQ316453
	<i>Chioninia vaillanti vaillanti</i>	M51	AF280332/AF280200/AF280266/HQ316433
	<i>Chioninia vaillanti vaillanti</i>	M50	AF280331/AF280199/AF280265/HQ316432
	<i>Chioninia vaillanti xanthotis</i>	M48	AF280329/AF280197/AF280263/HQ316434
<i>Chioninia vaillanti xanthotis</i>	M23	AF280328/AF280196/AF280262/HQ316435	
(Carranza & Arnold 2003)	<i>Chioninia delalandii</i>	M45	AF280317/AF280185/AY151482/AF280251
	<i>Chioninia fogoensis</i>	M60	AF280309/AF280177/AF280243
	<i>Chioninia nicolauensis</i>	M17	AF280304/AF280172/AF280238
	<i>Chioninia spinalis maioensis</i>	M53	AF280291/AF280159/AF280225

	<i>Chioninia stangeri</i>	M44	AF280299/AF280167/AF280233
	<i>Heremites auratus</i>	E11102	AY151507/AY151435/AY151469
	<i>Heremites vittatus</i>	E111040	AY151514/AY151442/AY151488
	<i>Trachylepis atlantica</i>	E11120	AY151501/AY151429/AY151463
	<i>Trachylepis perrotetii</i>	E111017	AY151489/AY151417/AY151451
	<i>Trachylepis socotrana</i>	M73	AF280272/AF280140/AY151476
(Karin et al. 2016)	<i>Dasia olivacea</i>	BRK 392	KX231476/KX231445/KX231429/KX231492/KX231533/KX231411/KX231393/KX231375
	<i>Notomabuya frenata</i>	UAM 60	KX231453/KX231437/KX231499/KX231419/KX231526/KX231401/KX231383
	<i>Trachylepis depressa</i>	PEM R17745	KX231484/KX231455/KX231501/KX231541/KX231421/KX231528/KX231403/KX231385
Metallinou et al. 2016	<i>Eumecia anchietae</i>	PEM R16779	KX231478/KX231448/KX231432/KX231494/KX231535/KX231414/KX231522/KX231396/KX231378
Metallinou et al. 2016	<i>Trachylepis boulengeri</i>	PEM R5533	KX231483/KX231454/KX231438/KX231500/KX231540/KX231420/KX231527/KX231402/KX231384
Whiting et al. 2006	<i>Eutropis multifasciata</i>	FMNH 255530	DQ239138/DQ239219/AY444029/DQ238897/AY444055
Brandley et al. 2005	<i>Liopholis whitii</i>	R34781	AY649109/AY649150/KP843150/KP843157
Whiting et al. 2003	<i>Tiliqua gigas</i>	L35	AY217811/AY218015/AY217862/AY217965

Hemidactylus

Reference	Taxa	Code	Accession No. (RAG2/CytB1/CytB2/CytB/12S)
(Arnold et al. 2008)	<i>Hemidactylus agrius</i>	Hag1	EF540746/DQ120262/DQ120433
	<i>Hemidactylus b. bowvieri</i>	H54	DQ120253/EU730668/EU730648
	<i>Hemidactylus b. bowvieri</i>	HRV	EF540744/EU730659/EU730669/EU730649
	<i>Hemidactylus b. razoensis</i>	H04	EF540738/EU730655/EU730664/EU730644
	<i>Hemidactylus b. razoensis</i>	H28	EF540740/EU730656/EU730665/EU730645
	<i>Hemidactylus b. razoensis</i>	H29	EU730683/EU730657/EU730666/EU730646
	<i>Hemidactylus b. razoensis</i>	H30	EU730684/EU730658/EU730667/EU730647
	<i>Hemidactylus boavistensis</i>	H01	EU730677/DQ120247/EU730670/DQ120418
	<i>Hemidactylus boavistensis</i>	H09	DQ120248/EU730671/DQ120419
	<i>Hemidactylus boavistensis</i>	H10	DQ120249/EU730672/DQ120420
	<i>Hemidactylus boavistensis</i>	H02	EU730678/DQ120251/EU730673/DQ120422
	<i>Hemidactylus boavistensis</i>	H11	EU730679/DQ120251/EU730674/DQ120422
	<i>Hemidactylus boavistensis</i>	H12	EU730680/DQ120250/EU730675/DQ120421
	<i>Hemidactylus boavistensis</i>	HJ01BV1	AF324811/AF324812
	<i>Hemidactylus boavistensis</i>	HCV38	AF324809/AF324810
	<i>Hemidactylus boavistensis</i>	HCV125	AF324807/AF324808
	<i>Hemidactylus greeffi</i>	Hgr	DQ120244/DQ120415
	<i>Hemidactylus lopezjuradoi</i>	H03	EU730681/EU730650/EU730660/EU730639
	<i>Hemidactylus lopezjuradoi</i>	H13	EU730682/EU730651/EU730660/EU730640
	<i>Hemidactylus palaichthus</i>	Hpa	DQ120263/DQ120434
	<i>Hemidactylus platycephalus</i>	Hpl	EF540745/DQ120266/DQ120437
	<i>Hemidactylus sp</i>	H05	EF540737/EU730652/EU730661/EU730641
	<i>Hemidactylus sp</i>	H16	EF540742/EU730653/EU730662/EU730642
	<i>Hemidactylus sp</i>	H17	EF540743/EU730654/EU730663/EU730643
(Šmíd et al. 2013)	<i>Hemidactylus sinaitus</i>	JS146	JQ957446/KC818866/KC818712
Yuan et al. 2013	<i>Cyrtodactylus wayakonei</i>	R0010	KF258130/KF258110
Cervenka et al. 2010	<i>Tenuidactylus longipes</i>	R\IRA\1030 (R1030)	JQ945515/EU589170/EU589193
Metallinou et al. 2012	<i>Stenodactylus petrii</i>	E1505332	KC191101/KC190637

Tarentola

Reference	Taxa	Code	Accession No. (CytB/CytB1/CytB2/12S/PDC/ACM4/MC1R/Cmos)
(Carranza et al. 2000)	<i>Tarentola boavistensis</i>	T49	AF185009/AF186137
	<i>Tarentola boavistensis</i>	T52	AF185007/AF186135
	<i>Tarentola darwini</i>	T54	AF185036/AF186164
	<i>Tarentola darwini</i>	T55	AF185038/AF186166
	<i>Tarentola darwini</i>	T56	AF185039/AF186167
	<i>Tarentola darwini</i>	T61	AF185044/AF186172
	<i>Tarentola darwini</i>	T64	AF185047/AF186175/AF363565
	<i>Tarentola gigas brancoensis</i>	T89	AF185018/AF186146
	<i>Tarentola gigas gigas</i>	T88	AF185019/AF186147
	<i>Tarentola hartogi</i>	T78	AF185020/AF186148
	<i>Tarentola hartogi</i>	T79	AF185021/AF186149
	<i>Tarentola hartogi</i>	T80	AF185022/AF186150
	<i>Tarentola hartogi</i>	T81	AF185023/AF186151
	<i>Tarentola nicolauensis</i>	T65	AF185034/AF186162
	<i>Tarentola nicolauensis</i>	T66	AF185035/AF186163
	<i>Tarentola protogigas</i>	T83	AF185025/AF186153
	<i>Tarentola protogigas</i>	T84	AF185027/AF186155
	<i>Tarentola protogigas</i>	T85	AF185026/AF186154
	<i>Tarentola protogigas</i>	T86	AF185028/AF186156
	<i>Tarentola raziana</i>	T68	AF185029/AF186157
	<i>Tarentola raziana</i>	T69	AF185033/AF186161
	<i>Tarentola raziana</i>	T70	AF185032/AF186160
	<i>Tarentola rudis</i>	T74	AF185013/AF186141
	<i>Tarentola rudis</i>	T75	AF185014/AF186142
	<i>Tarentola rudis</i>	T77	AF185016/AF186144
	<i>Tarentola substituta</i>	T71	AF185030/AF186158
	(Carranza et al. 2002)	<i>Tarentola americana</i>	Tamer
<i>Tarentola angustimentalis</i>		CA6FV	AF184992/AF186120/AF363545
<i>Tarentola annularis</i>		TananL	AF364322/AF363571/AF363552
<i>Tarentola boehmei</i>		Tbohem7	AF364320/AF363569/AF363543
<i>Tarentola boettgeri</i>		CA11GC	AF184997/AF186125/AF363548
<i>Tarentola boettgeri bischoffi</i>		Tbobish13	AF185000/AF186128/AF363551
<i>Tarentola delalandii</i>		CA1LP	AF185001/AF186129/AF363557
<i>Tarentola deserti</i>		Tdeser12	AF364321/AF363570/AF363544
<i>Tarentola ehippiata</i>		Tephi18	AF364323/AF363572/AF363553
<i>Tarentola gigas brancoensis</i>		T87	AF185017/AF186145/AF363563
<i>Tarentola gomerensis</i>		CA13GOM	AF185004/AF186132/AF363560
<i>Tarentola mauritania</i>		Tarentolamauri3	AF364327/AF363576/AF363566
<i>Tarentola rudis</i>		T73	AF185012/AF186140/AF363562
<i>Tarentola substituta</i>		T72	AF185031/AF186159/AF363564
(Gamble et al. 2008a)	<i>Tarentola gigas</i>	JB45	EU293707/EU293662/EU293685

(Jesus et al. 2002)	<i>Tarentola gigas</i>	T23893	AF468806/AF468807
	<i>Tarentola nicolauensis</i>	T23850	AF468796/AF468797
	<i>Tarentola nicolauensis</i>	T23851	AF468794/AF468795
	<i>Tarentola raziana</i>	IICT317	AF468812/AF468813
(Rato et al. 2012)	<i>Tarentola chazaliae</i>	DB863	JQ300722/JQ301036/JQ301166
	<i>Tarentola mindiae</i>	SPM002968	JQ300592/JQ301061/JQ301144
	<i>Tarentola neglecta</i>	Ksar_1	JQ300649/JQ301083/JQ301152
(Rato et al. 2015)	<i>Tarentola substituta</i>	DB13716	KJ814253
(Vasconcelos et al. 2010)	<i>Tarentola caboverdiana</i>	cv105	GQ381120/GQ380712/GQ380699
	<i>Tarentola caboverdiana</i>	cv107	GQ381124/GQ380716/GQ380703
	<i>Tarentola darwini</i>	T497	GQ380890
	<i>Tarentola darwini</i>	T577	GQ380791
	<i>Tarentola gigas gigas</i>	raTg4	GQ381129
	<i>Tarentola maioensis</i>	cv90	GQ380762/GQ380720/GQ380707
	<i>Tarentola maioensis</i>	cv91	GQ380763/GQ380721/GQ380708
	<i>Tarentola protogigas</i>	T535	GQ380782/GQ380719/GQ380706
	<i>Tarentola protogigas</i>	T598	GQ380783
	<i>Tarentola raziana</i>	T165	GQ381022
(Vasconcelos et al. 2012)	<i>Tarentola boavistensis</i>	T661	GQ381016/JN208926/JN208998/JN209070
	<i>Tarentola boavistensis</i>	T665	GQ381012/JN208930/JN209002/JN209074
	<i>Tarentola bocagei</i>	DB2532	JN185933
	<i>Tarentola bocagei</i>	T302	GQ380950/JN208936/JN209008/JN209079
	<i>Tarentola bocagei</i>	T304	GQ380952/JN208937/JN209009/JN209080
	<i>Tarentola bocagei</i>	T349	GQ380954/JN208938/JN209010/JN209081
	<i>Tarentola caboverdiana</i>	T188	GQ381117/JN208977/JN209048/JN209159
	<i>Tarentola caboverdiana</i>	T192	GQ381119/JN208978/JN209049/JN209161
	<i>Tarentola darwini</i>	T370	GQ380825/JN208953/JN209025/JN209095
	<i>Tarentola darwini</i>	T529	GQ380932
	<i>Tarentola fogoensis</i>	T533	GQ380784/JN208942/JN209014/JN209084
	<i>Tarentola fogoensis</i>	T537	GQ380787/JN208943/JN209015/JN209085
	<i>Tarentola fogoensis</i>	T540	GQ380790/JN208944/JN209016/JN209086
	<i>Tarentola fogoensis</i>	T583	GQ380795/JN208945/JN209017/JN209087
	<i>Tarentola gigas gigas</i>	Tg02	GQ381127/JN208983/JN209055/JN209181
	<i>Tarentola gigas gigas</i>	Tg03	GQ381128/JN208984/JN209056/JN209182
	<i>Tarentola maioensis</i>	T607	GQ380743/JN208996/JN209068/JN209194
	<i>Tarentola maioensis</i>	T640	GQ380754/JN208997/JN209069/JN209195
	<i>Tarentola nicolauensis</i>	T311	GQ380972/JN208981/JN209053/JN209179
	<i>Tarentola nicolauensis</i>	T315	GQ380994/JN208982/JN209054/JN209180
	<i>Tarentola protogigas</i>	T532	GQ380781/JN208990/JN209062/JN209188
	<i>Tarentola raziana</i>	T144	GQ381027/JN208974/JN209045
	<i>Tarentola raziana</i>	T145	GQ381017/JN208975/JN209046/JN209139
	<i>Tarentola raziana</i>	Tra1	GQ381032/JN208976/JN209047/JN209151
	<i>Tarentola raziana</i>	Tra2	GQ381033/JN209152
	<i>Tarentola rudis</i>	T371	GQ380725/JN208985/JN209057/JN209183
	<i>Tarentola rudis</i>	T377	GQ380726/JN208986/JN209058/JN209184
	<i>Tarentola substituta</i>	T043	GQ381049/JN208972/JN209043/JN209122

Torres-Carvajal et al. 2013	<i>Tarentola substituta</i> <i>Phyllodactylus leoni</i>	T046 QCAZ9927	GQ381051/JN208973/JN209044/JN209123 JQ821782/JQ821749/JQ821815/JQ821738
Gamble et al. 2008 Carranza et al. 2009	<i>Homonota fasciata</i>	TG00085 E90814	FJ985032/FJ985052/EU293697/EU293651/EU293674
Simo-Riudalbas et al. 2017	<i>Ptyodactylus ruusaljibalicus</i>	CN3951	MF084691/MF084456/MF084578/MF084751/MF084517

Appendix III – Summary of the list of sequences used in this study. The complete list of sequences and their corresponding information is available in the digital version of this appendix as a Microsoft Excel Spreadsheet.

Table 1 - Summary of the sequences in this study.

	No. Sequences	No. CV seq.	No. Non-CV seq.
<i>Chioninia</i>	588	519	69
<i>Hemidactylus</i>	221	69	152
<i>Tarentola</i>	916	859	57
Total	1725	1447	278

Appendix IV - Comprehensive data for each of the genes in this study, also available as a Microsoft Excel Spreadsheet in the digital version of this appendix.

Chioninia

Gene	12S	16S	COI	CytB	BDN F	BRC A1	BRC A2	C- mos	EXP H5	KIF2 4	MC1 R	RAG 1	RAG 2
Number of sequences	41	11	31	41	4	5	4	4	5	4	5	5	29
Alignment length (base pairs)	384	535	498	716	687	978	776	373	867	560	660	1147	799
Number of Haplotypes	30	11	23	36	4	5	4	4	5	4	5	5	9
Number of Sites excluding gaps/missing data	379	412	498	305	662	344	447	356	787	439	622	148	799
Number of Sites with gaps or missing data	5	123	5	411	25	634	329	17	80	121	38	999	0
% of gaps and missing data in the matrix	1.3	23.0	1.0	57.40	3.6	64.8	42.4	4.6	9.2	21.6	5.8	87.1	0
Conserved sites with Cape Verde species only	267	NA	371	437	668	770	649	347	718	440	582	995	NA
Conserved sites with all ingroups	329	388	371	491	NA	NA	NA	NA	NA	NA	NA	NA	789
Variable (polymorphic) sites with Cape Verde species only	116	NA	132	279	17	180	103	26	149	108	78	101	NA
Variable (polymorphic) sites with all ingroups	54	139	132	225	NA	NA	NA	NA	NA	NA	NA	NA	10
Parsimony informative sites with Cape Verde species only	80	NA	126	238	2	18	14	7	23	7	13	7	NA
Parsimony informative sites with all ingroups	48	83	126	215	NA	NA	NA	NA	NA	NA	NA	NA	5
Singleton variable sites with Cape Verde species only	36	NA	6	41	15	162	78	19	122	98	65	67	NA
Singleton variable sites with all ingroups	6	54	6	10	NA	NA	NA	NA	NA	NA	NA	NA	5
GC Content	0.449	0.444	0.420	0.447	0.482	0.413	0.320	0.408	0.405	0.587	0.537	0.443	0.433

Hemidactylus

Gene	12S	CytB	RAG2
Number of sequences	28	28	36
Alignment length (base pairs)	421	1107	869
Number of Haplotypes	22	25	10
Number of Sites excluding gaps/missing data	327	253	867
Number of Sites with gaps or missing data	94	854	2
% of gaps and missing data in the matrix	22.3	77.2	0.2
Conserved sites with Cape Verde species only	229	389	826
Conserved sites with all ingroups	335	491	799
Variable (polymorphic) sites with Cape Verde species only	164	290	41
Variable (polymorphic) sites with all ingroups	42	178	3
Parsimony informative sites with Cape Verde species only	98	218	19
Parsimony informative sites with all ingroups	33	167	3
Singleton variable sites with Cape Verde species only	60	71	22
Singleton variable sites with all ingroups	9	11	0
GC Content	0.499	0.465	0.416

Tarentola

Gene	12S	CytB	ACM4	C-mos	MC1R	PDC
Number of sequences	53	83	30	16	29	27
Alignment length (base pairs)	425	708	429	375	669	392
Number of Haplotypes	39	70	11	11	17	9
Number of Sites excluding gaps/missing data	285	303	361	374	590	385
Number of Sites with gaps or missing data	140	405	68	1	79	7
% of gaps and missing data in the matrix	32.9	57.2	15.9	0.3	11.8	1.8
Conserved sites with Cape Verde species only	252	329	418	351	637	NA
Conserved sites with all ingroups	386	483	425	373	651	384
Variable (polymorphic) sites with Cape Verde species only	163	361	10	24	29	NA
Variable (polymorphic) sites with all ingroups	44	207	3	2	13	7
Parsimony informative sites with Cape Verde species only	101	287	6	7	12	NA
Parsimony informative sites with all ingroups	40	172	1	2	9	2
Singleton variable sites with Cape Verde species only	61	74	4	17	17	NA
Singleton variable sites with all ingroups	4	35	2	0	4	5
GC Content	0.514	0.481	0.457	0.434	0.538	0.454

Appendix V - Results attained in PartitionFinder, also available as a Microsoft Excel Spreadsheet in the digital version of this appendix.

Chioninia

Mitochondrial RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	384	12S
2	GTR+I+G	535	16S
3	GTR+I+G	407	CytB_pos1, COI_pos1
4	GTR+G	168	COI_pos2
5	GTR+G	167	COI_pos3
6	GTR+G	239	CytB_pos2
7	GTR+I+G	238	CytB_pos3

Mitochondrial RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	385	12S
2	GTR+I+G	537	16S
3	GTR+I+G	408	CytB_pos1, COI_pos1
4	GTR+G	168	COI_pos2
5	GTR+G	167	COI_pos3
6	GTR+G	240	CytB_pos2
7	GTR+I+G	239	CytB_pos3

Mitochondrial MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	384	12S
2	GTR+I+G	535	16S
3	F81+I	168	COI_pos1
4	GTR+G	168	COI_pos2
5	GTR+I	167	COI_pos3
6	HKY+I	239	CytB_pos1
7	GTR+G	239	CytB_pos2
8	K80+I+G	238	CytB_pos3

Mitochondrial MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	385	12S
2	GTR+I+G	537	16S
3	HKY+I	408	COI_pos1, CytB_pos1
4	GTR+G	168	COI_pos2
5	GTR+I	167	COI_pos3
6	GTR+G	240	CytB_pos2
7	SYM+I+G	239	CytB_pos3

Nuclear RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR	458	BDNF_pos1, BDNF_pos3
2	GTR	1250	Cmos_pos2, Cmos_pos3, BDNF_pos2, RAG1_pos3, RAG2_pos2, Cmos_pos1
3	GTR	652	BRCA1_pos1, BRCA1_pos2
4	GTR+G	1148	BRCA1_pos3, RAG2_pos1, RAG2_pos3, EXPH5_pos3
5	GTR+G	1602	EXPH5_pos2, RAG1_pos1, RAG1_pos2, EXPH5_pos1, BRCA2_pos1
6	GTR	259	BRCA2_pos2
7	GTR	258	BRCA2_pos3

8	GTR	187	KIF24_pos1
9	GTR	373	KIF24_pos2, KIF24_pos3
10	GTR+G	660	MC1R_pos2, MC1R_pos1, MC1R_pos3

Nuclear RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	586	BDNF_pos3, BDNF_pos1, Cmos_pos2
2	GTR	355	BDNF_pos2, Cmos_pos1
3	GTR	652	BRCA1_pos1, BRCA1_pos2
4	GTR	2087	EXPH5_pos2, EXPH5_pos3, RAG1_pos2, RAG2_pos2, BRCA1_pos3, RAG2_pos1, RAG2_pos3
5	GTR+G	548	EXPH5_pos1, BRCA2_pos1
6	GTR	259	BRCA2_pos2
7	GTR	258	BRCA2_pos3
8	GTR+G	124	Cmos_pos3
9	GTR	187	KIF24_pos1
10	GTR	757	RAG1_pos3, KIF24_pos2, KIF24_pos3
11	GTR+G	660	MC1R_pos2, MC1R_pos1, MC1R_pos3
12	GTR	385	RAG1_pos1

Nuclear MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	HKY	229	BDNF_pos1
2	K80	229	BDNF_pos2
3	F81	229	BDNF_pos3
4	HKY	652	BRCA1_pos1, BRCA1_pos2
5	GTR+I	1380	EXPH5_pos2, BRCA1_pos3, RAG1_pos1, RAG1_pos2
6	HKY+G	548	BRCA2_pos1, EXPH5_pos1
7	GTR	259	BRCA2_pos2
8	HKY	258	BRCA2_pos3
9	HKY	1288	RAG2_pos2, Cmos_pos3, Cmos_pos2, Cmos_pos1, RAG2_pos1, RAG1_pos3
10	HKY+I	555	RAG2_pos3, EXPH5_pos3
11	HKY	187	KIF24_pos1
12	HKY	373	KIF24_pos3, KIF24_pos2
13	HKY+I	440	MC1R_pos1, MC1R_pos3
14	GTR+G	220	MC1R_pos2

Nuclear MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	HKY+I	231	BDNF_pos1
2	HKY	355	BDNF_pos2, Cmos_pos1
3	K80+I	230	BDNF_pos3
4	HKY	652	BRCA1_pos1, BRCA1_pos2
5	GTR	1509	RAG2_pos2, RAG1_pos2, BRCA1_pos3, RAG2_pos3, RAG2_pos1
6	HKY+G	548	BRCA2_pos1, EXPH5_pos1
7	GTR	259	BRCA2_pos2
8	HKY	258	BRCA2_pos3
9	SYM+I	125	Cmos_pos2
10	GTR+G	124	Cmos_pos3
11	HKY+I	578	EXPH5_pos2, EXPH5_pos3
12	HKY	187	KIF24_pos1
13	HKY	373	KIF24_pos3, KIF24_pos2
14	HKY+I	440	MC1R_pos1, MC1R_pos3
15	GTR+G	220	MC1R_pos2
16	HKY	385	RAG1_pos1

17 K80 384 RAG1_pos3

All RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	1028	BDNF_pos1, RAG2_pos2, RAG2_pos1, RAG2_pos3
2	GTR+I+G	467	BDNF_pos2, CytB_pos3
3	GTR+G	396	BDNF_pos3, COI_pos3
4	GTR	652	BRCA1_pos1, BRCA1_pos2
5	GTR+G	1928	EXPH5_pos2, BRCA1_pos3, RAG1_pos1, RAG1_pos2, BRCA2_pos1, EXPH5_pos1
6	GTR	259	BRCA2_pos2
7	GTR+G	547	BRCA2_pos3, EXPH5_pos3
8	GTR	755	Cmos_pos3, Cmos_pos2, RAG1_pos3, Cmos_pos1
9	GTR	560	KIF24_pos1, KIF24_pos2, KIF24_pos3
10	GTR+G	660	MC1R_pos2, MC1R_pos1, MC1R_pos3
11	GTR+G	384	12S
12	GTR+I+G	535	16S
13	GTR+I+G	407	CytB_pos1, COI_pos1
14	GTR+G	168	COI_pos2
15	GTR+G	239	CytB_pos2

All RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	385	12S
2	GTR+I+G	537	16S
3	GTR+I+G	408	COI_pos1, CytB_pos1
4	GTR+G	168	COI_pos2
5	GTR+I+G	398	BDNF_pos1, COI_pos3
6	GTR+G	240	CytB_pos2
7	GTR+I+G	469	CytB_pos3, BDNF_pos3
8	GTR+I+G	1154	Cmos_pos1, BDNF_pos2, RAG2_pos2, RAG2_pos1, RAG2_pos3
9	GTR	1421	RAG1_pos1, RAG1_pos2, BRCA1_pos2, BRCA1_pos1
10	GTR+G	904	BRCA1_pos3, EXPH5_pos2, EXPH5_pos3
11	GTR+G	548	EXPH5_pos1, BRCA2_pos1
12	GTR	259	BRCA2_pos2
13	GTR	258	BRCA2_pos3
14	GTR+I+G	125	Cmos_pos2
15	GTR+G	124	Cmos_pos3
16	GTR	944	KIF24_pos1, RAG1_pos3, KIF24_pos2, KIF24_pos3
17	GTR+G	660	MC1R_pos2, MC1R_pos3, MC1R_pos1

All MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	HKY+I	1028	BDNF_pos1, RAG2_pos2, RAG2_pos1, RAG2_pos3
2	K80+I+G	591	Cmos_pos2, BDNF_pos2, CytB_pos3
3	GTR+I	396	COI_pos3, BDNF_pos3
4	HKY	652	BRCA1_pos1, BRCA1_pos2
5	GTR+I	1504	EXPH5_pos2, Cmos_pos3, BRCA1_pos3, RAG1_pos1, RAG1_pos2
6	HKY+I	548	BRCA2_pos1, EXPH5_pos1
7	GTR	259	BRCA2_pos2
8	HKY+I	547	BRCA2_pos3, EXPH5_pos3
9	HKY	507	RAG1_pos3, Cmos_pos1
10	HKY	187	KIF24_pos1
11	HKY	373	KIF24_pos3, KIF24_pos2
12	GTR+G	660	MC1R_pos1, MC1R_pos2, MC1R_pos3

13	GTR+G	384	12S
14	GTR+I+G	535	16S
15	F81	168	COI_pos1
16	GTR+G	168	COI_pos2
17	HKY+I	239	CytB_pos1
18	GTR+G	239	CytB_pos2

All MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	385	12S
2	GTR+I+G	537	16S
3	F81	168	COI_pos1
4	GTR+G	168	COI_pos2
5	GTR+I	559	RAG2_pos1, Cmos_pos1, COI_pos3
6	HKY+I	240	CytB_pos1
7	GTR+G	240	CytB_pos2
8	SYM+I+G	239	CytB_pos3
9	HKY+I	231	BDNF_pos1
10	HKY+I	762	RAG2_pos3, RAG2_pos2, BDNF_pos2
11	K80+I	230	BDNF_pos3
12	HKY	652	BRCA1_pos1, BRCA1_pos2
13	HKY	999	BRCA1_pos3, EXPH5_pos3, RAG1_pos2
14	HKY+G	837	EXPH5_pos2, BRCA2_pos1, EXPH5_pos1
15	GTR	259	BRCA2_pos2
16	HKY	258	BRCA2_pos3
17	SYM+I	125	Cmos_pos2
18	GTR+G	124	Cmos_pos3
19	HKY	187	KIF24_pos1
20	HKY	373	KIF24_pos3, KIF24_pos2
21	GTR+G	660	MC1R_pos1, MC1R_pos2, MC1R_pos3
22	HKY	385	RAG1_pos1
23	K80	384	RAG1_pos3

All BEAST (Without External Outgroups)

Subset	Best Model	# sites	Partition names
1	HKY+I+X	1028	RAG2_pos2, BDNF_pos1, RAG2_pos3, RAG2_pos1
2	K80+I+G	467	BDNF_pos2, CytB_pos3
3	TRN+I+X	396	COI_pos3, BDNF_pos3
4	HKY+X	652	BRCA1_pos2, BRCA1_pos1
5	HKY+I+X	615	BRCA1_pos3, EXPH5_pos3
6	HKY+G+X	548	BRCA2_pos1, EXPH5_pos1
7	GTR+X	259	BRCA2_pos2
8	HKY+X	258	BRCA2_pos3
9	HKY+X	507	RAG1_pos3, Cmos_pos1
10	GTR+G+X	248	Cmos_pos3, Cmos_pos2
11	TRN+G+X	1054	EXPH5_pos2, RAG1_pos2, RAG1_pos1
12	HKY+X	187	KIF24_pos1
13	HKY+X	373	KIF24_pos3, KIF24_pos2
14	HKY+I+X	440	MC1R_pos1, MC1R_pos3
15	TRN+G+X	220	MC1R_pos2
16	GTR+G+X	384	12S
17	GTR+I+G+X	535	16S
18	HKY+X	168	COI_pos1

19	GTR+G+X	168	COI_pos2
20	TRN+I+X	239	CytB_pos1
21	TRN+G+X	239	CytB_pos2

All StarBEAST (With External Outgroups)

Subset	Best Model	# sites	Partition names
1	GTR+G+X	385	12S
2	HKY+I+G+X	537	16S
3	HKY+X	168	COI_pos1
4	GTR+G+X	168	COI_pos2
5	TRN+I+X	523	COI_pos3, Cmos_pos1, BDNF_pos1
6	HKY+I+X	240	CytB_pos1
7	TRN+G+X	240	CytB_pos2
8	GTR+I+G+X	239	CytB_pos3
9	HKY+I+X	1029	BDNF_pos2, RAG2_pos3, RAG2_pos2, RAG2_pos1
10	K80+I	230	BDNF_pos3
11	HKY+X	652	BRCA1_pos2, BRCA1_pos1
12	HKY+X	999	BRCA1_pos3, EXPH5_pos3, RAG1_pos2
13	HKY+G+X	548	EXPH5_pos1, BRCA2_pos1
14	GTR+X	259	BRCA2_pos2
15	HKY+X	258	BRCA2_pos3
16	TRNEF+I	125	Cmos_pos2
17	TRN+G+X	124	Cmos_pos3
18	HKY+X	289	EXPH5_pos2
19	HKY+X	187	KIF24_pos1
20	HKY+X	373	KIF24_pos3, KIF24_pos2
21	GTR+G+X	660	MC1R_pos2, MC1R_pos3, MC1R_pos1
22	HKY+X	385	RAG1_pos1
23	K80	384	RAG1_pos3

Hemidactylus

Mitochondrial RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	421	12S
2	GTR+I+G	369	CytB_pos1
3	GTR+G	369	CytB_pos2
4	GTR+G	369	CytB_pos3

Mitochondrial RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	410	12S
2	GTR+G	369	CytB_pos1
3	GTR+G	369	CytB_pos2
4	GTR+G	369	CytB_pos3

Mitochondrial MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	421	12S
2	K80+I+G	369	CytB_pos1
3	GTR+I	369	CytB_pos2
4	GTR+G	369	CytB_pos3

Mitochondrial MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	410	12S
2	K80+G	369	CytB_pos1
3	GTR+I	369	CytB_pos2
4	GTR+G	369	CytB_pos3

Nuclear RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR	290	RAG2_pos1
2	GTR	290	RAG2_pos2
3	GTR+G	289	RAG2_pos3

Nuclear RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR	290	RAG2_pos1
2	GTR	290	RAG2_pos2
3	GTR+G	289	RAG2_pos3

Nuclear MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	F81	290	RAG2_pos1
2	HKY	290	RAG2_pos2
3	HKY+G	289	RAG2_pos3

Nuclear MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	F81	290	RAG2_pos1
2	HKY	290	RAG2_pos2
3	HKY+G	289	RAG2_pos3

All RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	421	12S
2	GTR+I+G	369	CytB_pos1
3	GTR+G	369	CytB_pos2
4	GTR+G	369	CytB_pos3
5	GTR	290	RAG2_pos1
6	GTR	290	RAG2_pos2
7	GTR+G	289	RAG2_pos3

All RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	410	12S
2	GTR+G	369	CytB_pos1
3	GTR+I+G	369	CytB_pos2
4	GTR+G	369	CytB_pos3
5	GTR	290	RAG2_pos1
6	GTR	290	RAG2_pos2
7	GTR+G	289	RAG2_pos3

All MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	421	12S
2	K80+I+G	369	CytB_pos1
3	GTR+I	369	CytB_pos2
4	GTR+G	369	CytB_pos3

5	F81+I	290	RAG2_pos1
6	HKY	290	RAG2_pos2
7	HKY+G	289	RAG2_pos3

All MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	410	12S
2	K80+G	369	CytB_pos1
3	GTR+I+G	369	CytB_pos2
4	GTR+G	369	CytB_pos3
5	F81	290	RAG2_pos1
6	GTR	290	RAG2_pos2
7	HKY+G	289	RAG2_pos3

All BEAST (Without External Outgroups)

Subset	Best Model	# sites	Partition names
1	GTR+G+X	421	12S
2	HKY+I+G+X	369	CytB_pos1
3	GTR+I+X	369	CytB_pos2
4	GTR+G+X	369	CytB_pos3
5	HKY+I+X	290	RAG2_pos1
6	TRN+X	290	RAG2_pos2
7	HKY+G+X	289	RAG2_pos3

All StarBEAST (With External Outgroups)

Subset	Best Model	# sites	Partition names
1	GTR+I+G+X	410	12S
2	HKY+G+X	369	CytB_pos1
3	GTR+I+G+X	369	CytB_pos2
4	GTR+G+X	369	CytB_pos3
5	HKY+X	290	RAG2_pos1
6	TRN+X	290	RAG2_pos2
7	HKY+G+X	289	RAG2_pos3

Tarentola

Mitochondrial RAxML - Without External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR+I+G	425	12S
2	GTR+I+G	236	CytB_pos1
3	GTR+I+G	236	CytB_pos2
4	GTR+G	236	CytB_pos3
Mitochondrial RAxML - With External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR+I+G	425	12S
2	GTR+I+G	228	CytB_pos1
3	GTR+I+G	228	CytB_pos2
4	GTR+G	228	CytB_pos3
Mitochondrial MrBayes - Without External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR+I+G	425	12S
2	SYM+G	236	CytB_pos1
3	HKY+G	236	CytB_pos2
4	GTR+G	236	CytB_pos3
Mitochondrial MrBayes - With External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR+I+G	435	12S
2	SYM+I+G	228	CytB_pos1
3	HKY+I+G	228	CytB_pos2
4	GTR+G	228	CytB_pos3
Nuclear RAxML - Without External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR	268	Cmos_pos1, ACM4_pos1
2	GTR	143	ACM4_pos2
3	GTR	524	PDC_pos1, Cmos_pos2, ACM4_pos3, Cmos_pos3
4	GTR+I+G	223	MC1R_pos1
5	GTR	223	MC1R_pos2
6	GTR+I+G	484	PDC_pos3, MC1R_pos3, PDC_pos2
Nuclear RAxML - With External Outgroups			
Subset	Best Model	# sites	Partition names
1	GTR+G	268	Cmos_pos1, ACM4_pos1
2	GTR	143	ACM4_pos2
3	GTR	398	PDC_pos1, Cmos_pos2, ACM4_pos3, Cmos_pos3
4	GTR+G	476	MC1R_pos1
5	GTR+G	352	MC1R_pos2
6	GTR+G	221	PDC_pos3, MC1R_pos3, PDC_pos2
Nuclear MrBayes - Without External Outgroups			
Subset	Best Model	# sites	Partition names
1	F81	143	ACM4_pos1
2	F81	143	ACM4_pos2
3	K80	399	PDC_pos1, Cmos_pos2, ACM4_pos3
4	JC	125	Cmos_pos1
5	HKY	125	Cmos_pos3
6	GTR+I	223	MC1R_pos1

7	HKY	223	MC1R_pos2
8	GTR+I	484	PDC_pos3, MC1R_pos3, PDC_pos2

Nuclear MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	HKY	143	ACM4_pos1
2	HKY	143	ACM4_pos2
3	K80	398	Cmos_pos3, PDC_pos3, ACM4_pos3
4	JC+I	125	Cmos_pos1
5	GTR	255	Cmos_pos2, PDC_pos2
6	GTR+G	352	MC1R_pos1, PDC_pos1
7	GTR+I	221	MC1R_pos2
8	HKY+I	221	MC1R_pos3

All RAxML - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+G	425	12S
2	GTR+I+G	236	CytB_pos1
3	GTR+I+G	236	CytB_pos2
4	GTR+G	236	CytB_pos3
5	GTR	523	PDC_pos3, Cmos_pos2, Cmos_pos1, ACM4_pos1
6	GTR	366	MC1R_pos3, ACM4_pos2
7	GTR+G	399	PDC_pos1, ACM4_pos3, Cmos_pos3
8	GTR+I+G	223	MC1R_pos1
9	GTR+G	223	MC1R_pos2
10	GTR+G	131	PDC_pos2

All RAxML - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	435	12S
2	GTR+I+G	228	CytB_pos1
3	GTR+I+G	228	CytB_pos2
4	GTR+G	228	CytB_pos3
5	GTR+G	268	ACM4_pos1, Cmos_pos1
6	GTR	143	ACM4_pos2
7	GTR+G	398	Cmos_pos3, ACM4_pos3, PDC_pos3
8	GTR+G	476	MC1R_pos2, Cmos_pos2, PDC_pos2
9	GTR+G	352	PDC_pos1, MC1R_pos1
10	GTR+G	221	MC1R_pos3

All MrBayes - Without External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	425	12S
2	SYM+I+G	236	CytB_pos1
3	HKY+I+G	236	CytB_pos2
4	GTR+G	236	CytB_pos3
5	F81	143	ACM4_pos1
6	F81+I	143	ACM4_pos2
7	K80+I	399	PDC_pos1, ACM4_pos3, Cmos_pos3
8	JC+I	125	Cmos_pos1
9	HKY	255	Cmos_pos2, PDC_pos3
10	GTR+I	223	MC1R_pos1
11	HKY+G	223	MC1R_pos2
12	SYM	223	MC1R_pos3

13 JC+I 131 PDC_pos2

All MrBayes - With External Outgroups

Subset	Best Model	# sites	Partition names
1	GTR+I+G	435	12S
2	SYM+I+G	228	CytB_pos1
3	HKY+I+G	228	CytB_pos2
4	GTR+G	228	CytB_pos3
5	F81+I	268	ACM4_pos1, Cmos_pos1
6	HKY	143	ACM4_pos2
7	K80+I	273	PDC_pos3, ACM4_pos3
8	GTR	255	Cmos_pos2, PDC_pos2
9	K80	125	Cmos_pos3
10	GTR+G	352	PDC_pos1, MC1R_pos1
11	GTR+I	221	MC1R_pos2
12	HKY+G	221	MC1R_pos3

All BEAST (Without External Outgroups)

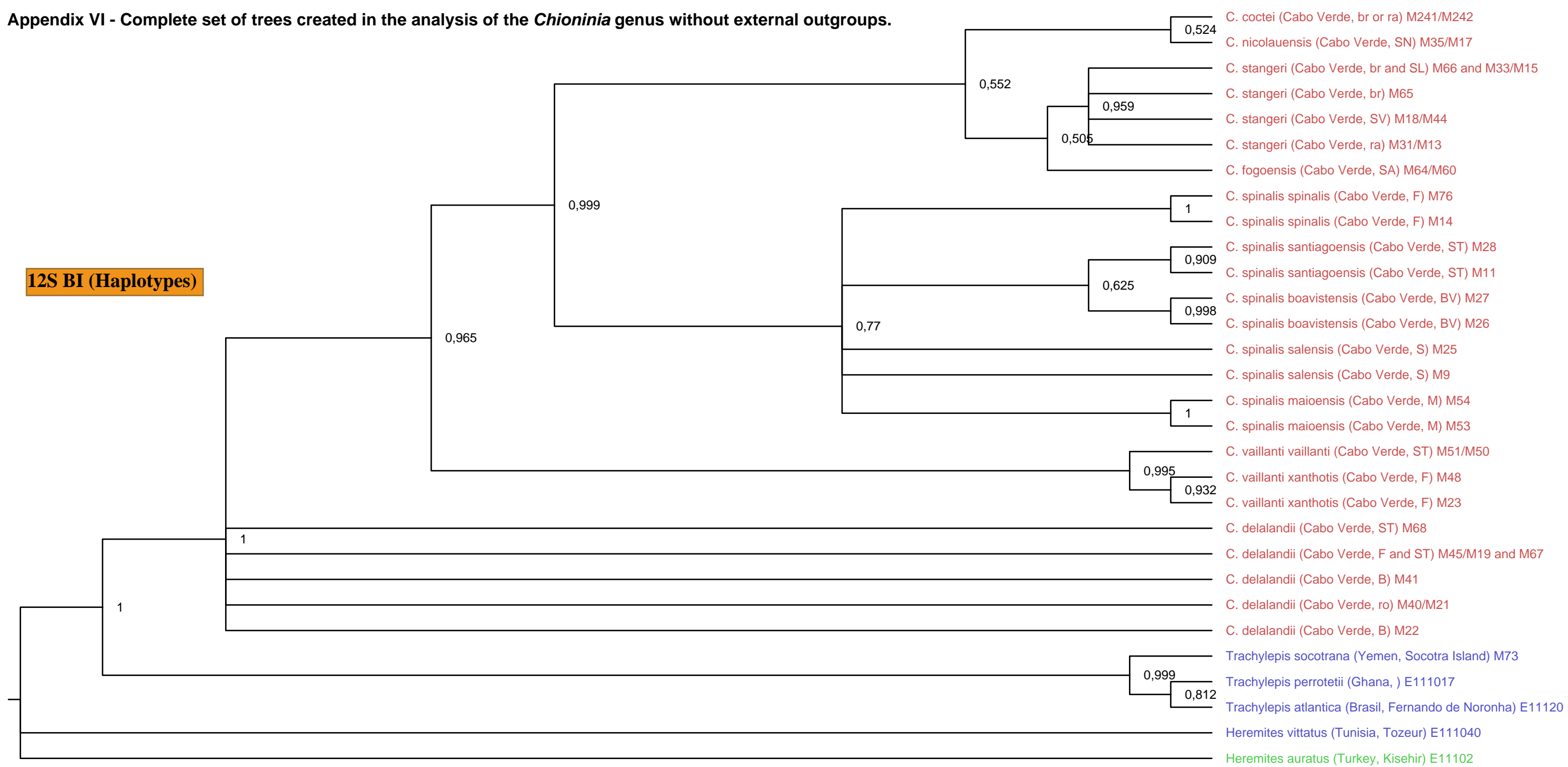
Subset	Best Model	# sites	Partition names
1	GTR+I+G+X	425	12S
2	SYM+I+G	236	CytB_pos1
3	HKY+I+G+X	236	CytB_pos2
4	GTR+G+X	236	CytB_pos3
5	HKY+X	143	ACM4_pos1
6	HKY+I+X	143	ACM4_pos2
7	K80+I	399	PDC_pos1, ACM4_pos3, Cmos_pos3
8	JC+I	125	Cmos_pos1
9	TRN+X	255	Cmos_pos2, PDC_pos3
10	GTR+I+X	223	MC1R_pos1
11	HKY+G+X	223	MC1R_pos2
12	SYM	223	MC1R_pos3
13	JC+I	131	PDC_pos2

All StarBEAST (With External Outgroups)

Subset	Best Model	# sites	Partition names
1	GTR+I+G+X	435	12S
2	SYM+I+G	228	CytB_pos1
3	HKY+I+G+X	228	CytB_pos2
4	GTR+G+X	228	CytB_pos3
5	HKY+X	143	ACM4_pos1
6	HKY+X	143	ACM4_pos2
7	K80+I	273	ACM4_pos3, PDC_pos3
8	JC+I	125	Cmos_pos1
9	GTR+X	255	Cmos_pos2, PDC_pos2
10	K80	125	Cmos_pos3
11	GTR+G+X	352	PDC_pos1, MC1R_pos1
12	GTR+I+X	221	MC1R_pos2
13	HKY+G+X	221	MC1R_pos3

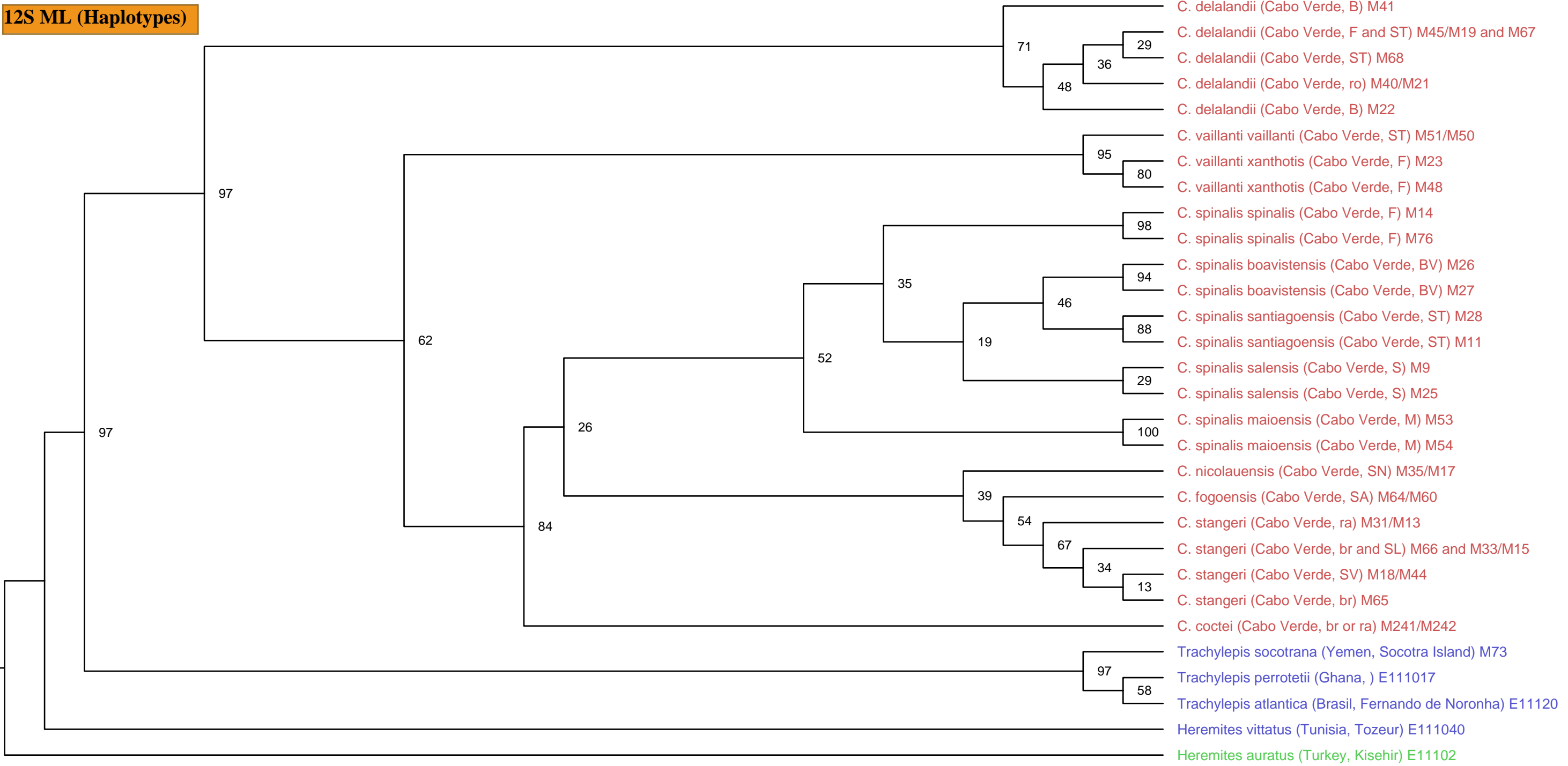
Appendix VI - Complete set of trees created in the analysis of the *Chioninia* genus without external outgroups.

12S BI (Haplotypes)



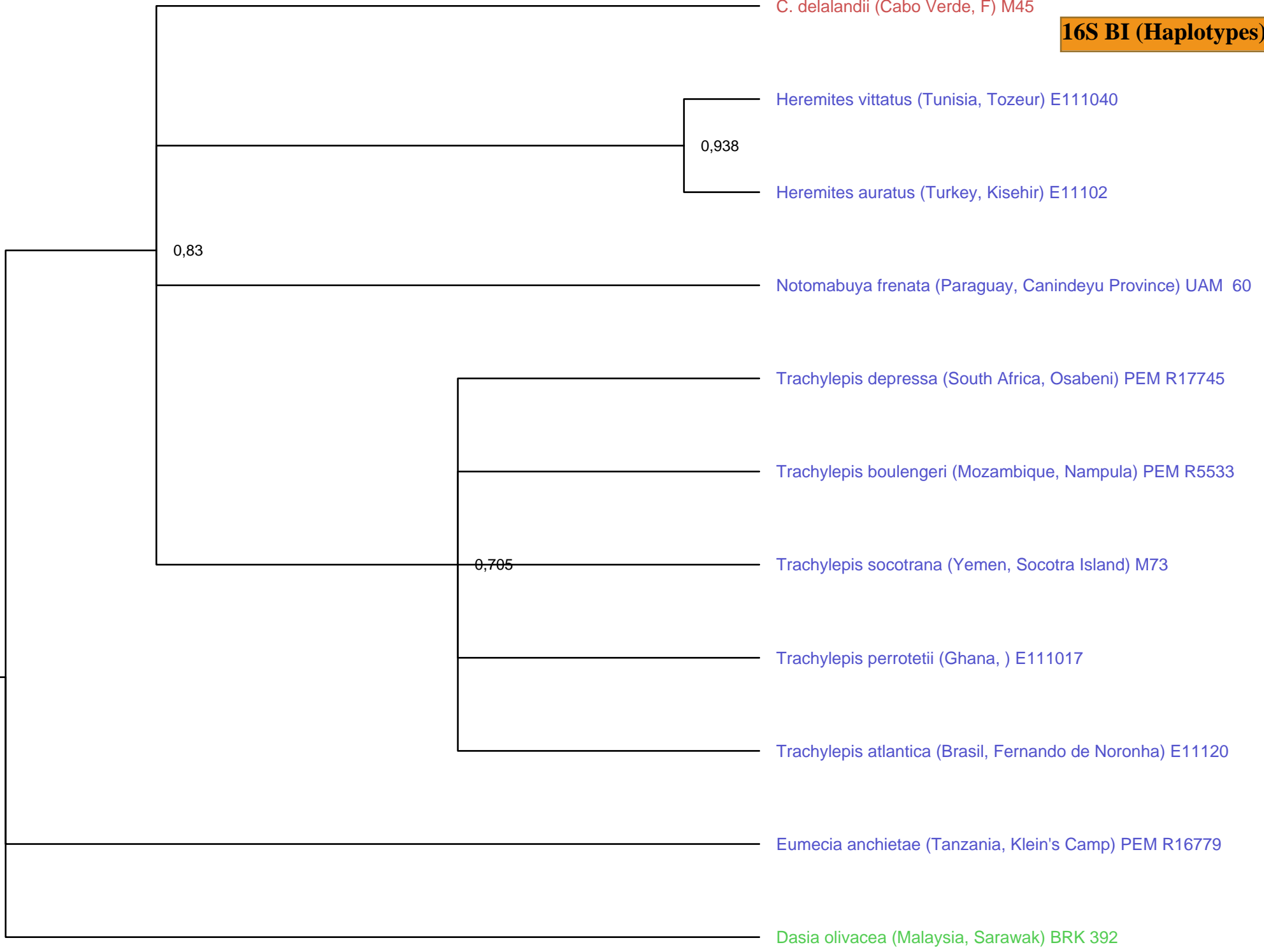
0.02

12S ML (Haplotypes)



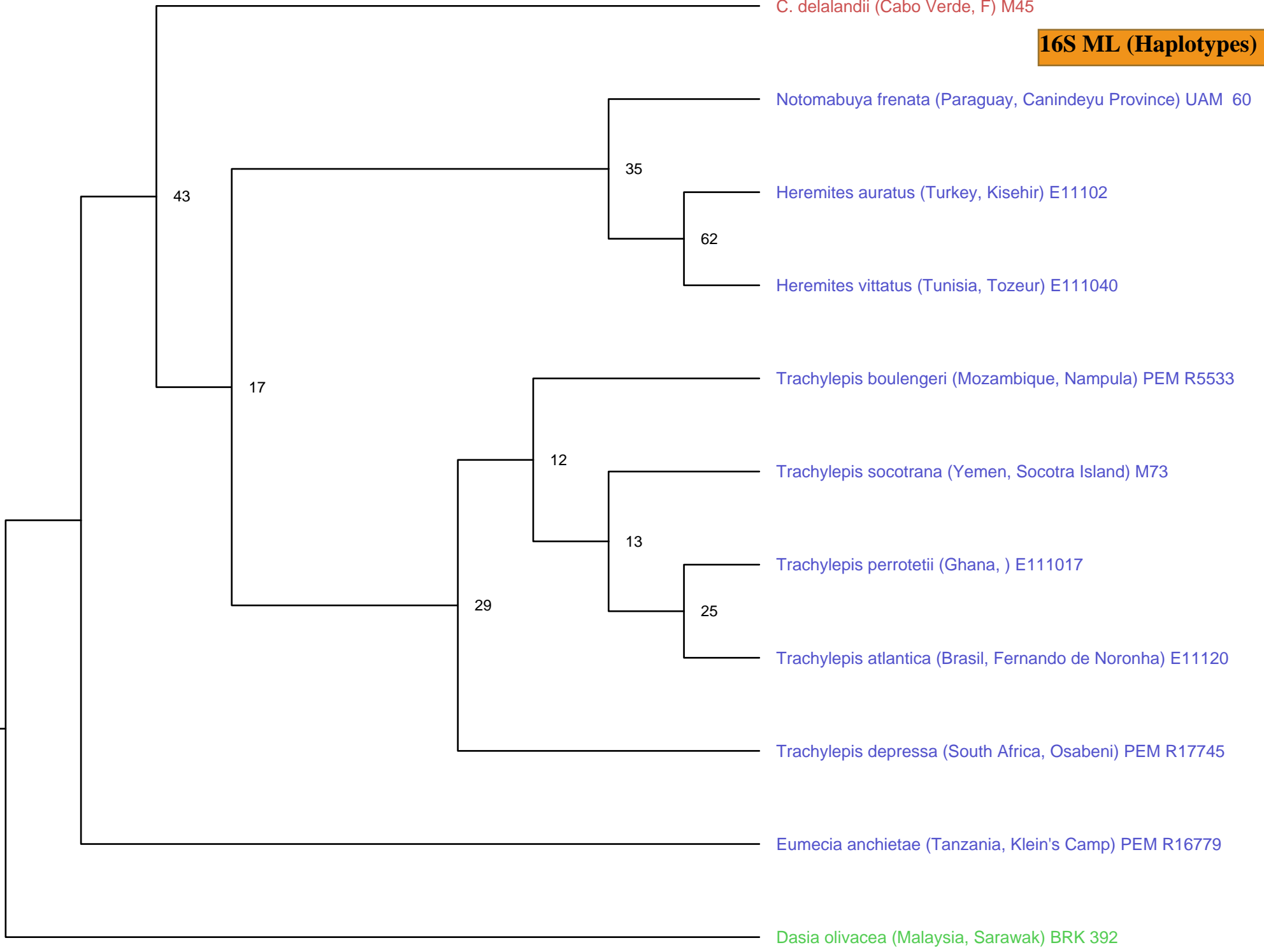
0.03

16S BI (Haplotypes)



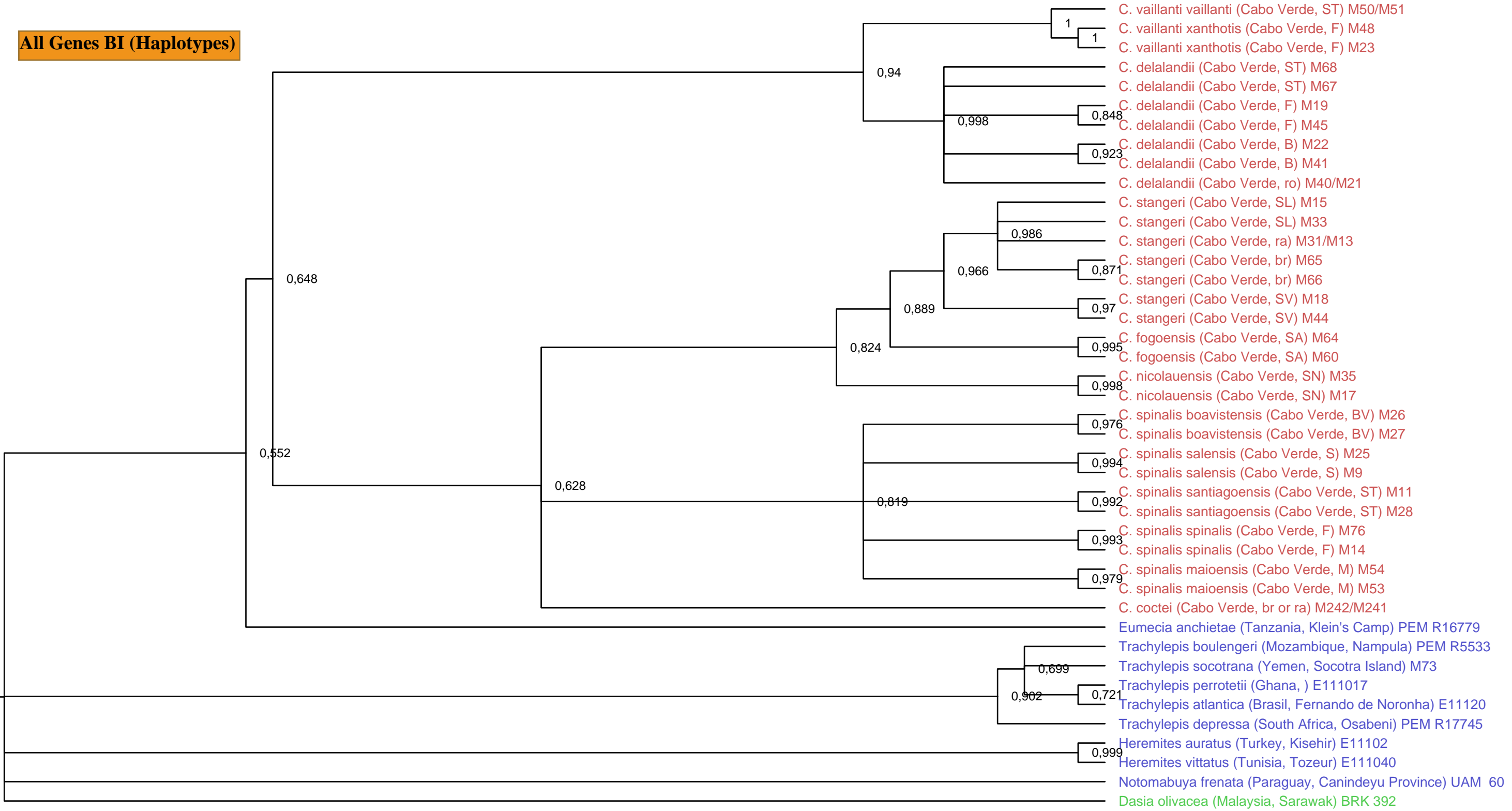
1.1

16S ML (Haplotypes)



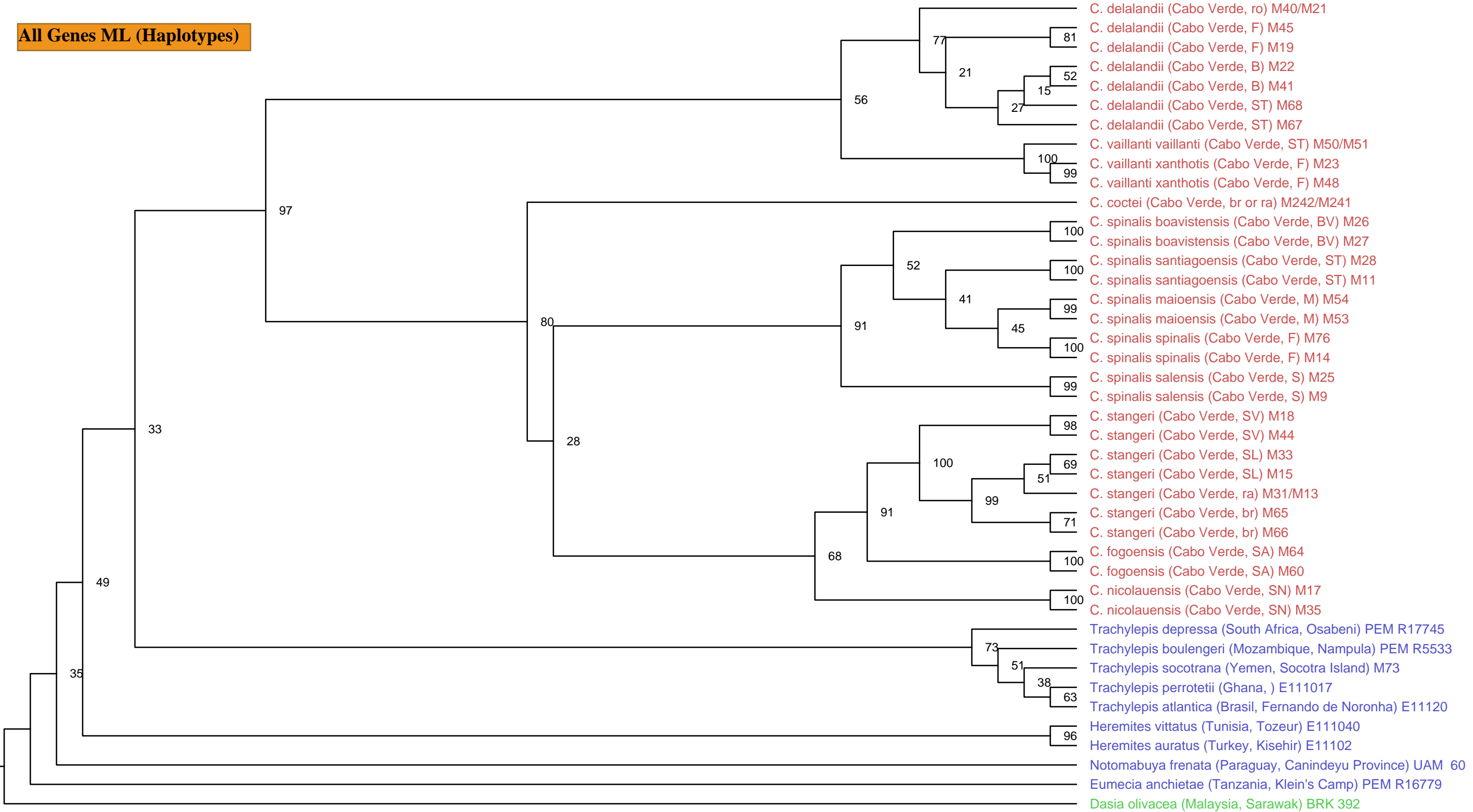
-
0.05

All Genes BI (Haplotypes)

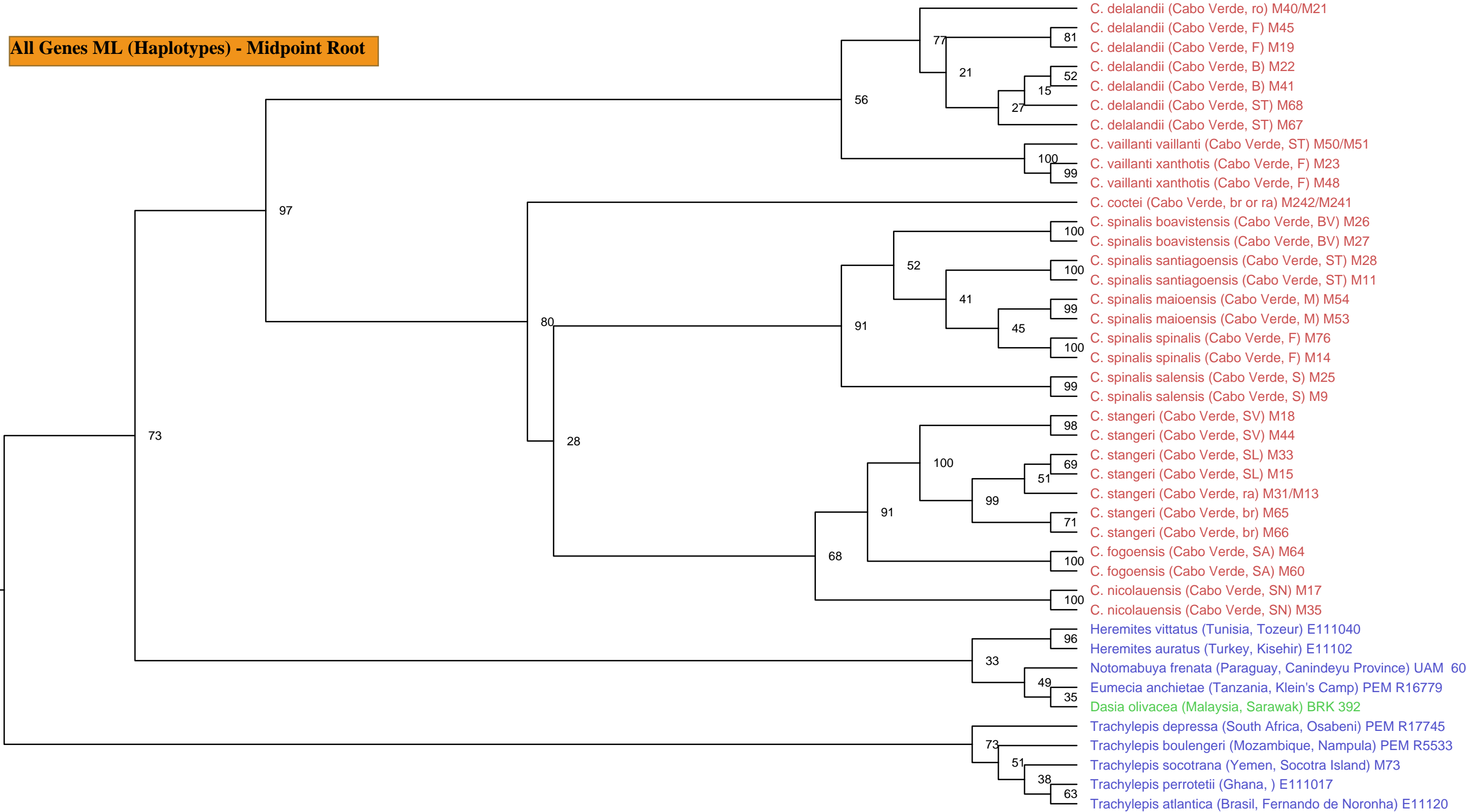


0.02

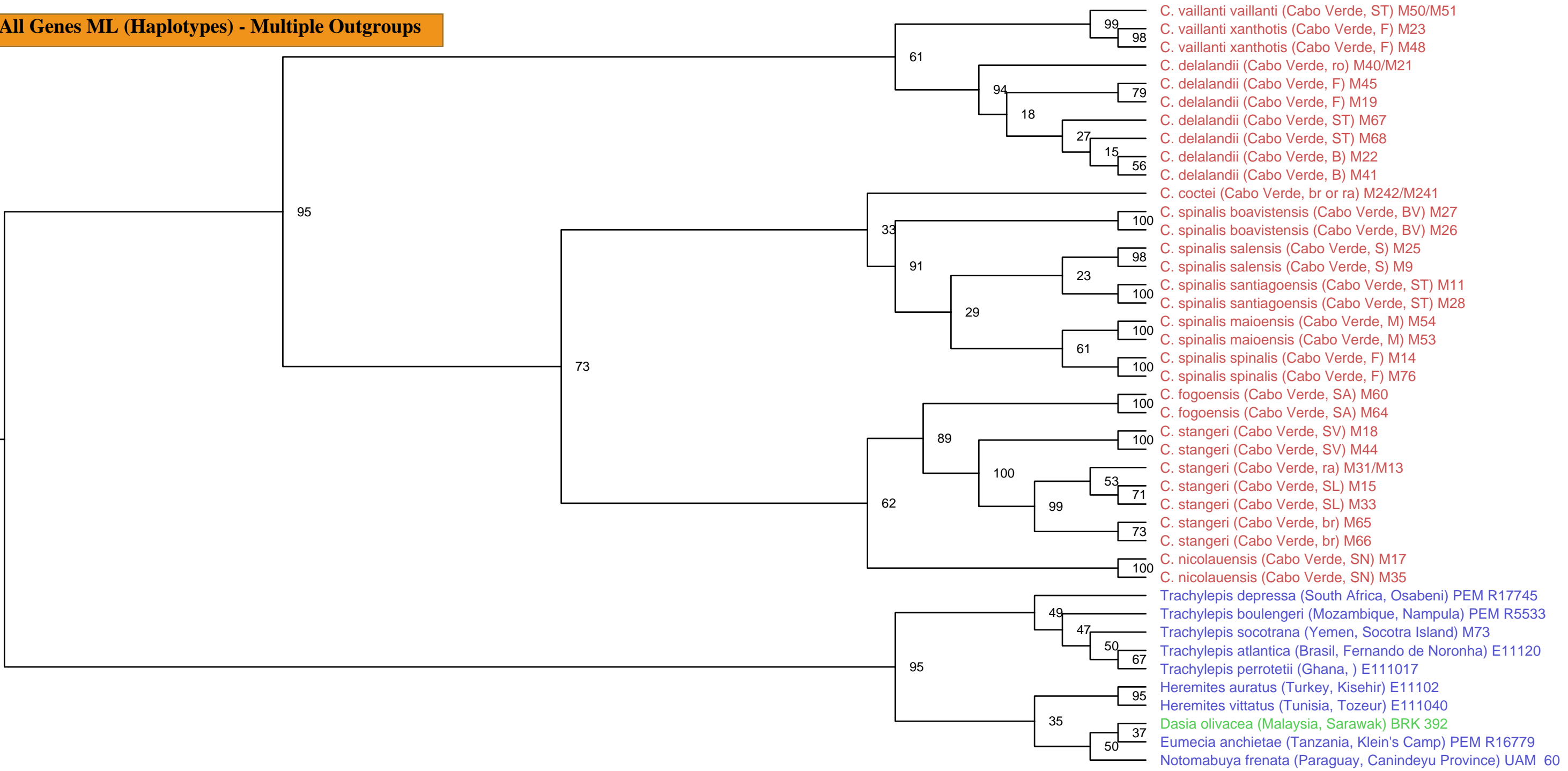
All Genes ML (Haplotypes)



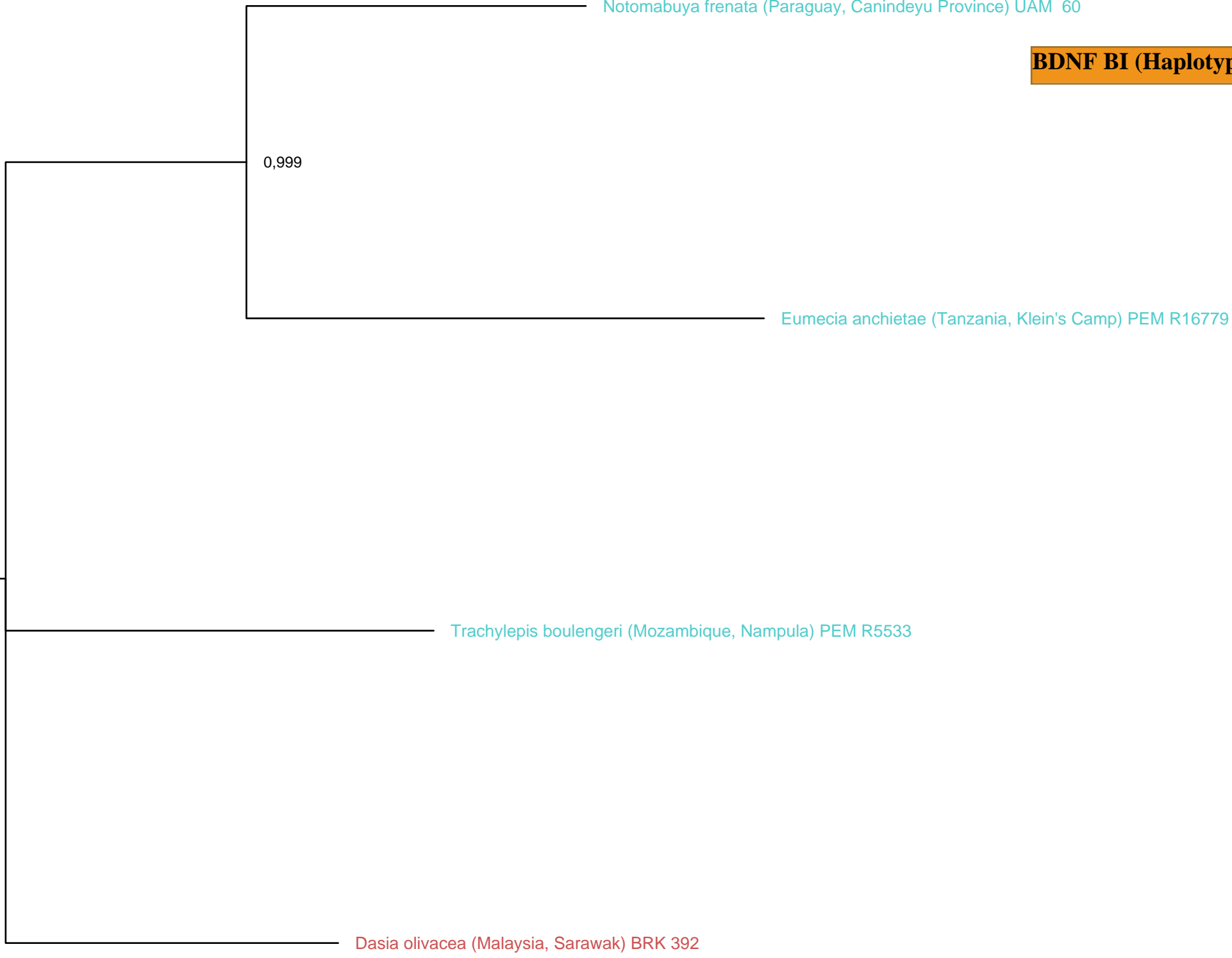
All Genes ML (Haplotypes) - Midpoint Root



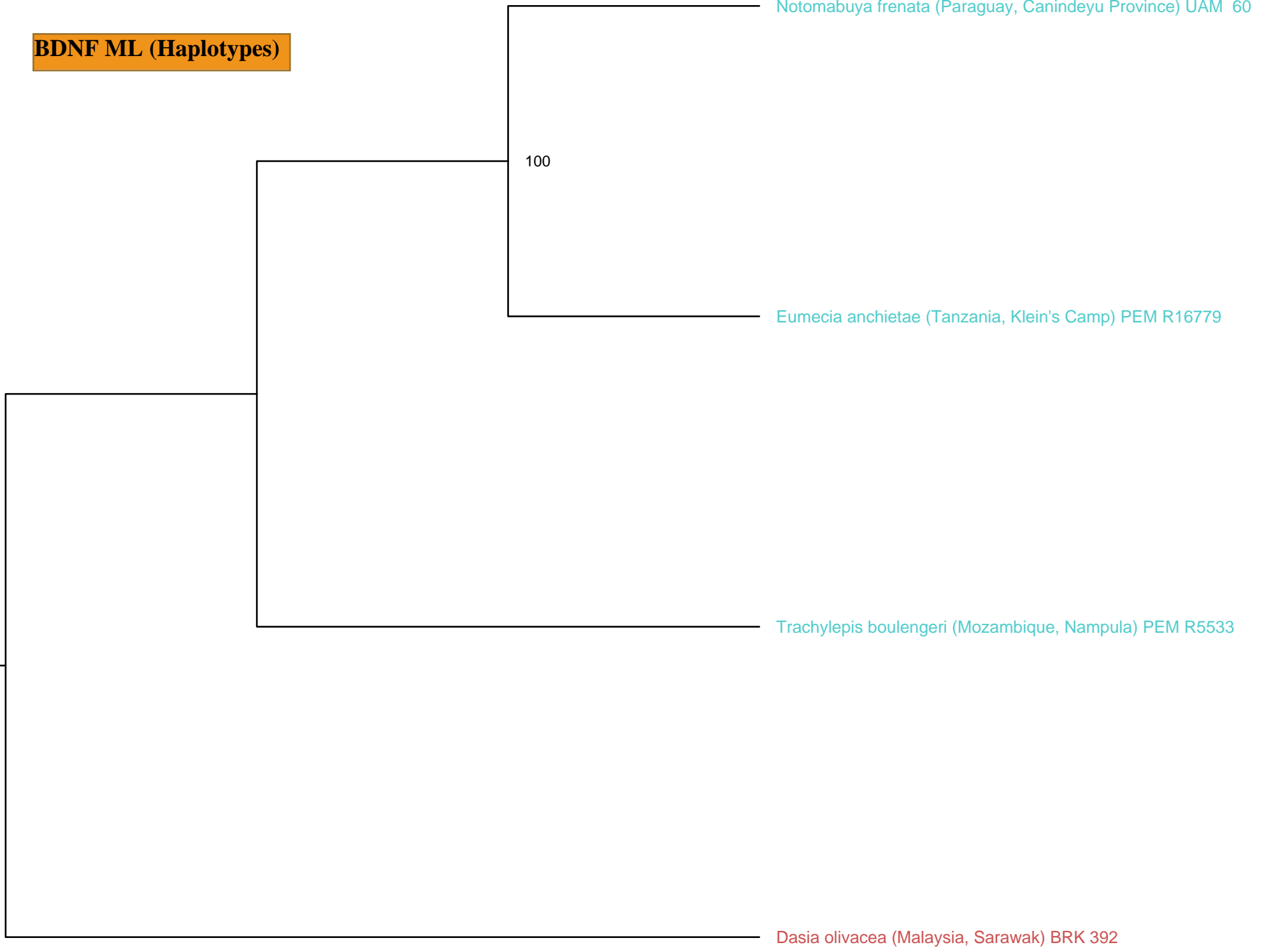
All Genes ML (Haplotypes) - Multiple Outgroups



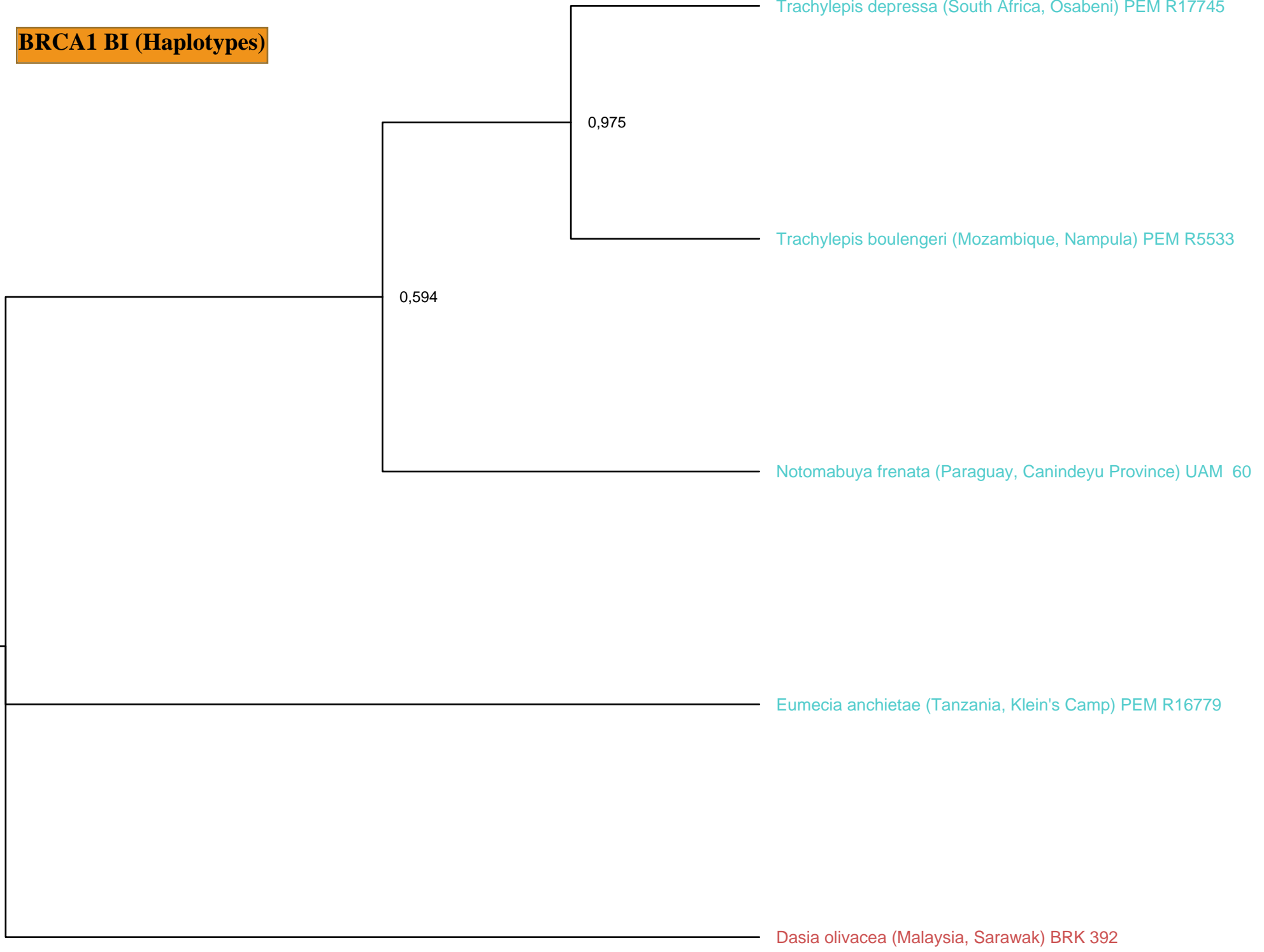
BDNF BI (Haplotypes)



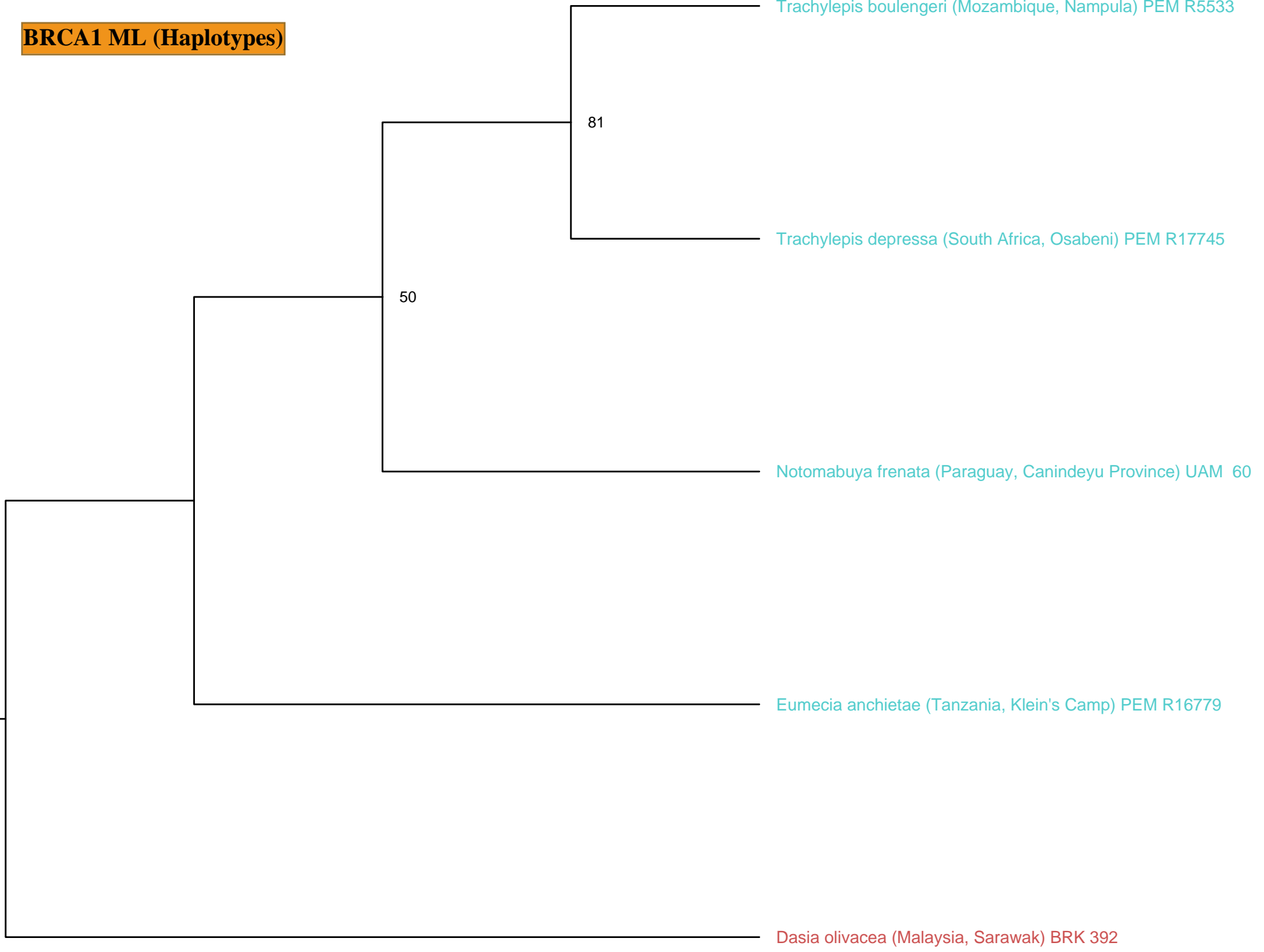
BDNF ML (Haplotypes)



BRCA1 BI (Haplotypes)

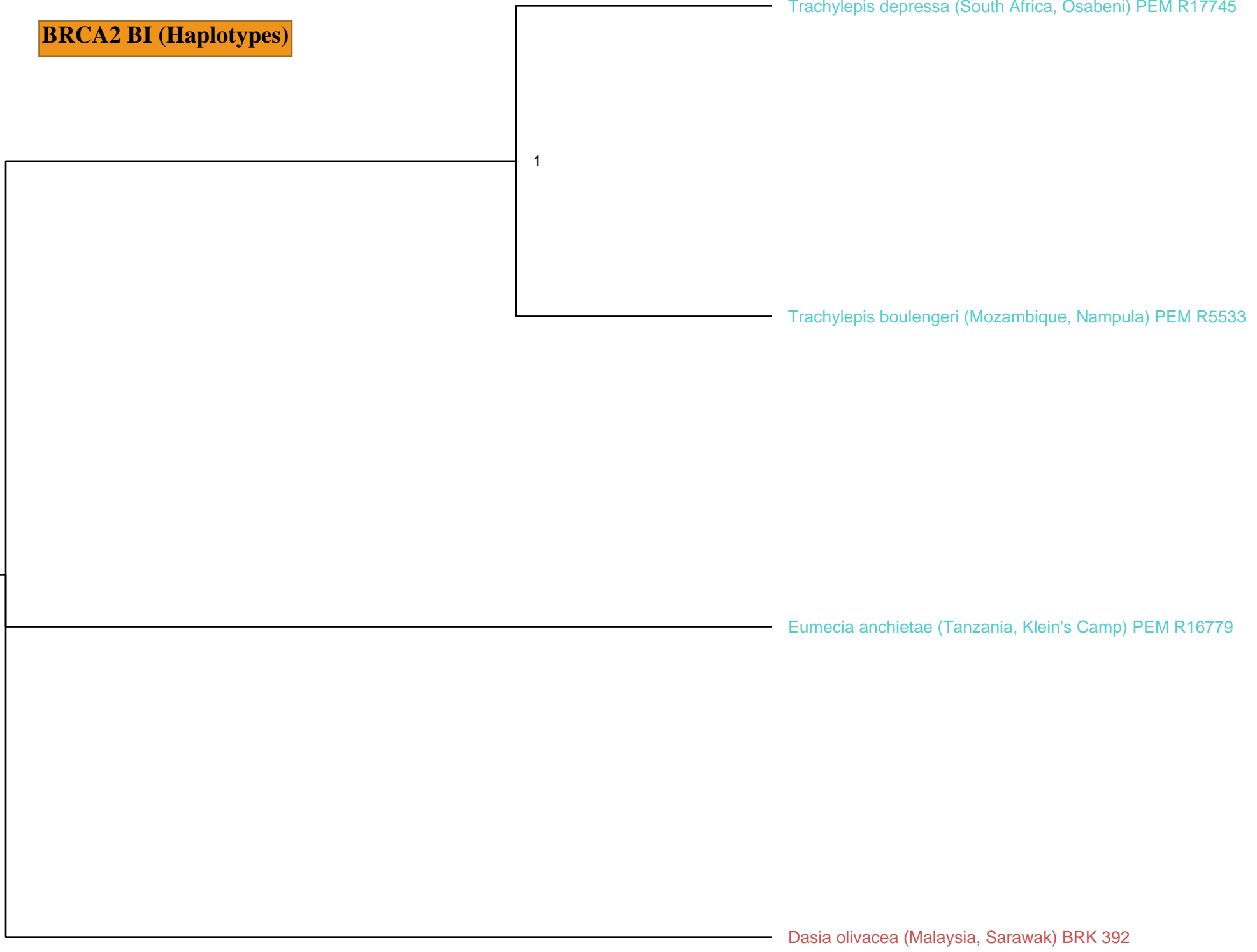


BRCA1 ML (Haplotypes)



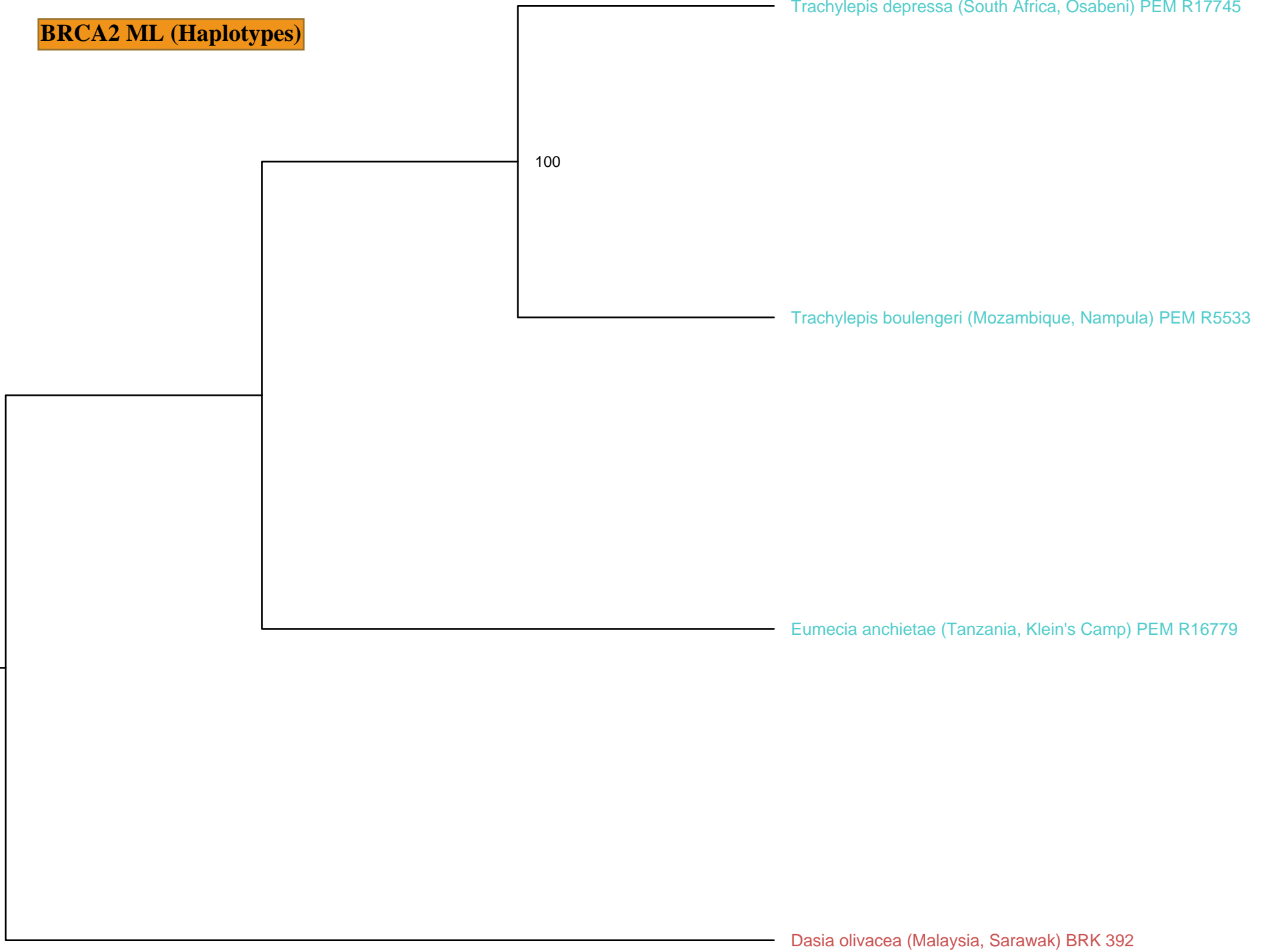
-
0.02

BRCA2 BI (Haplotypes)

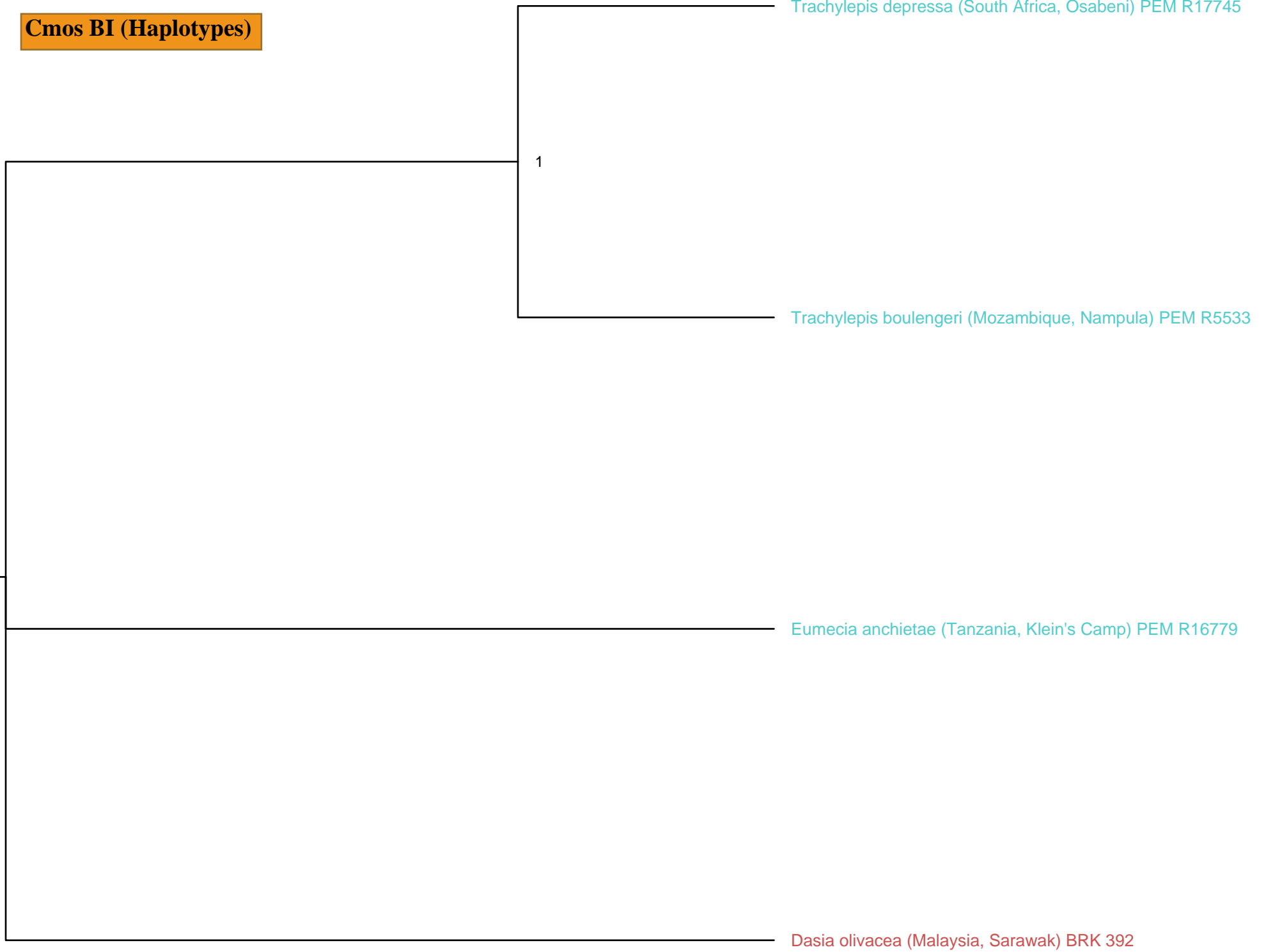


0.007

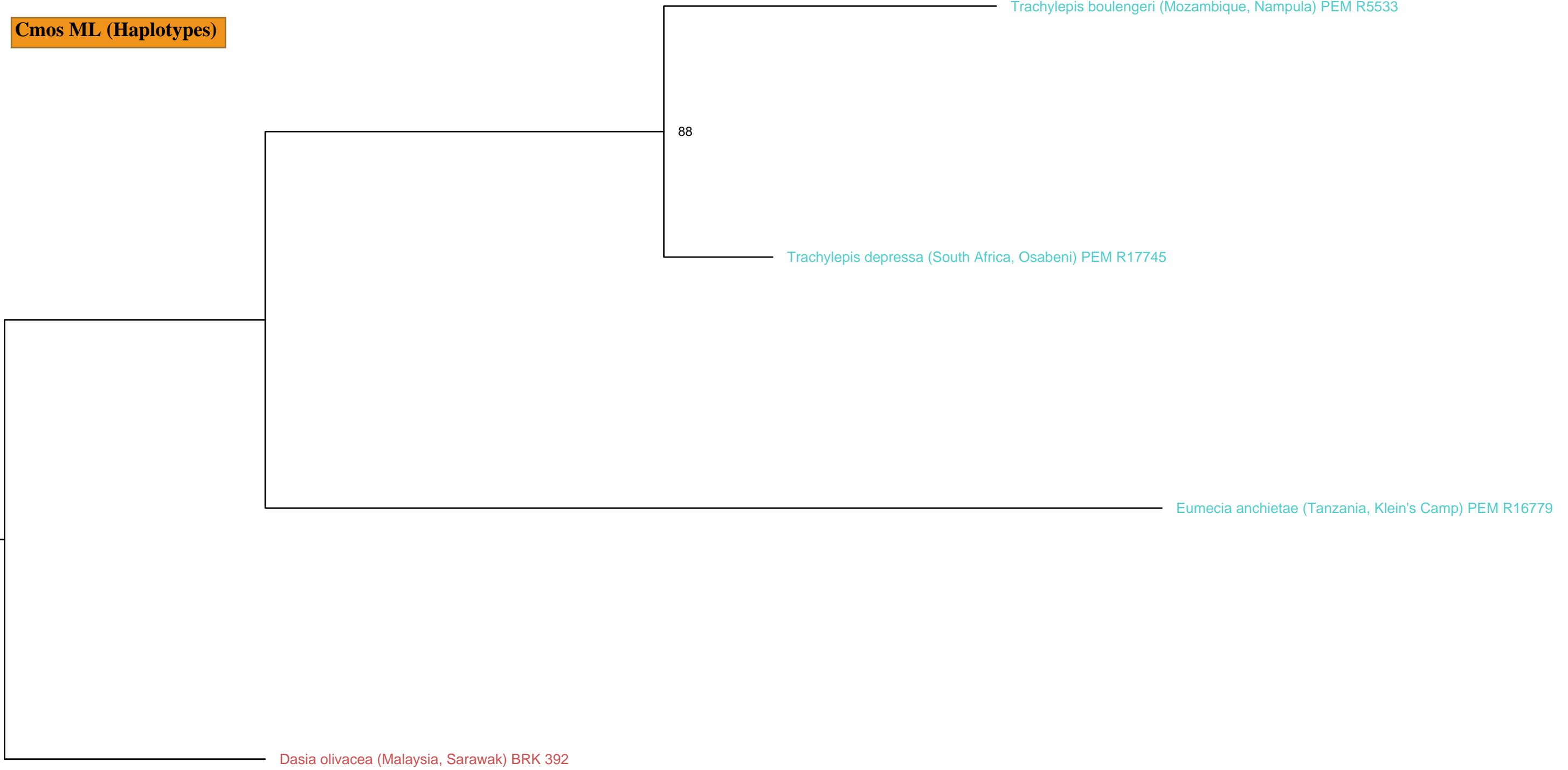
BRCA2 ML (Haplotypes)



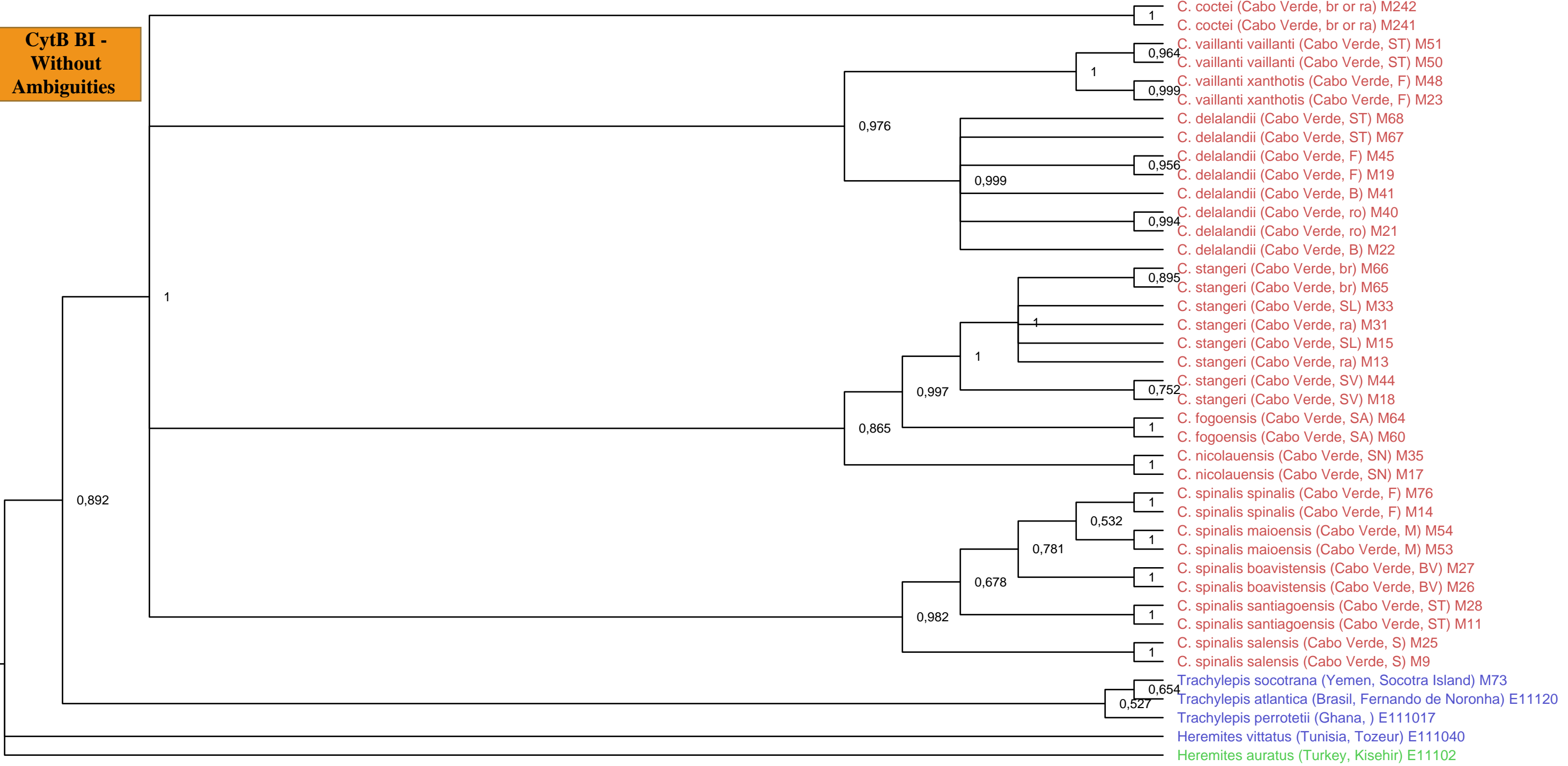
Cmos BI (Haplotypes)



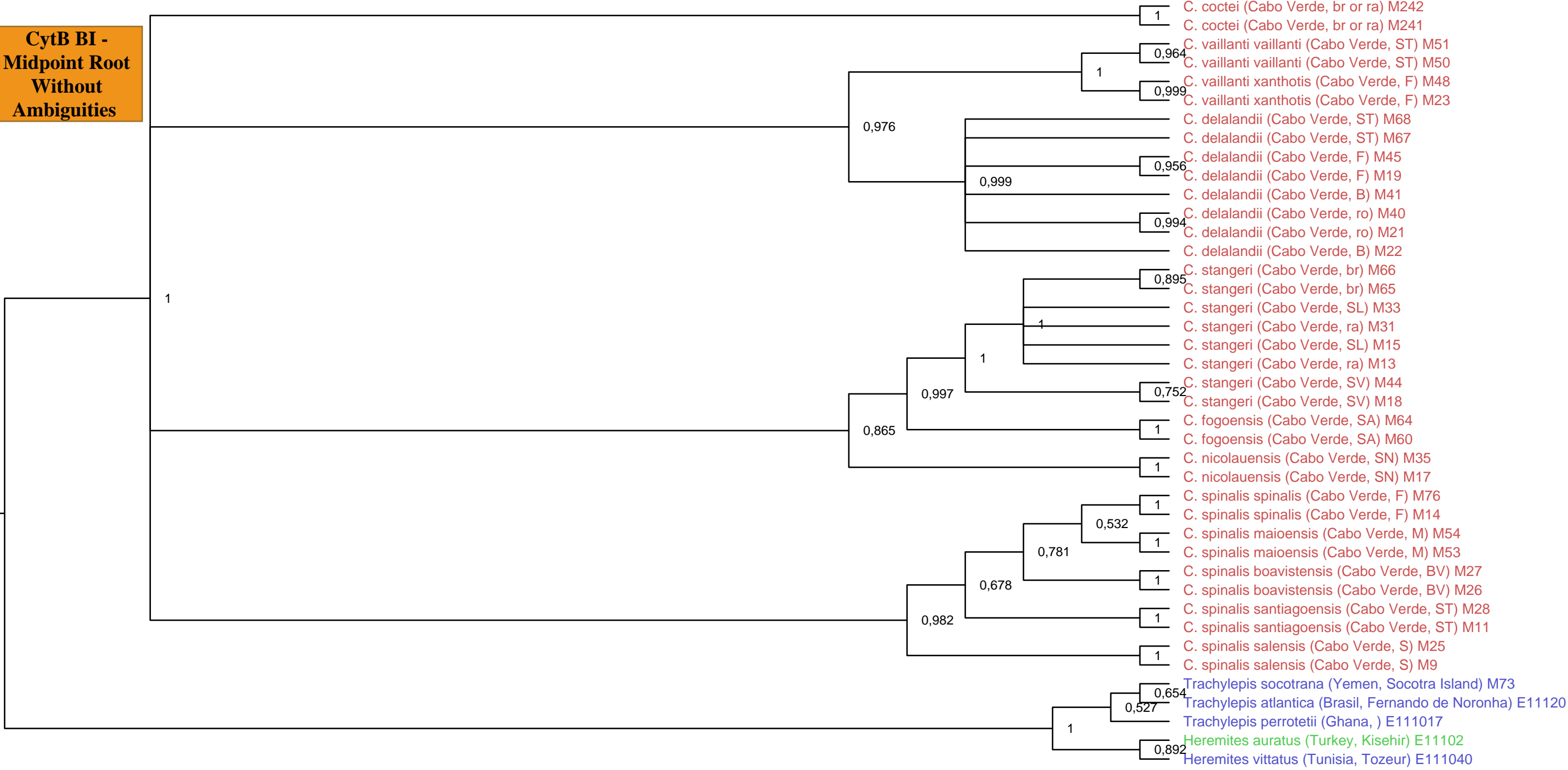
Cmos ML (Haplotypes)



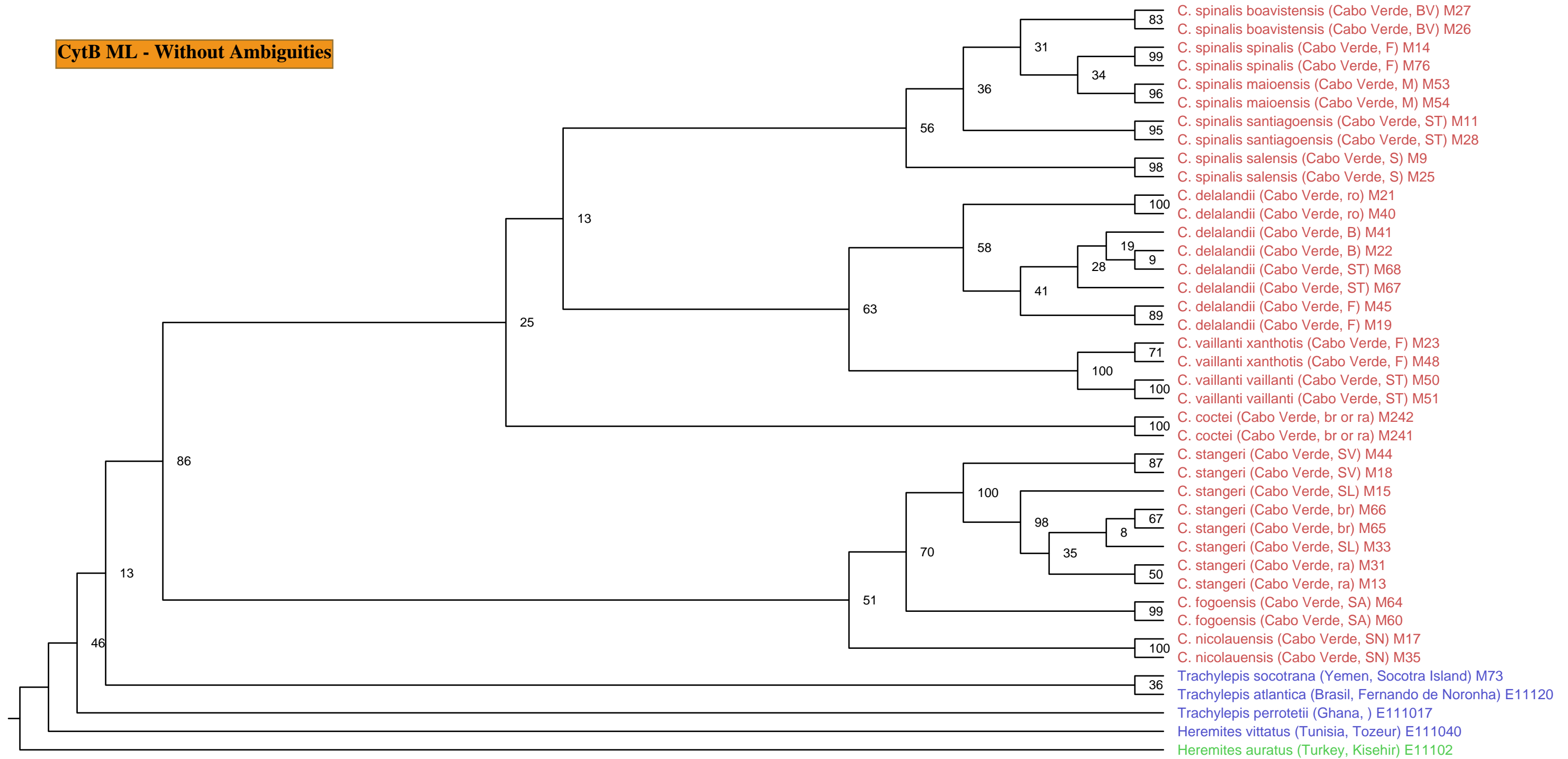
**CytB BI -
Without
Ambiguities**



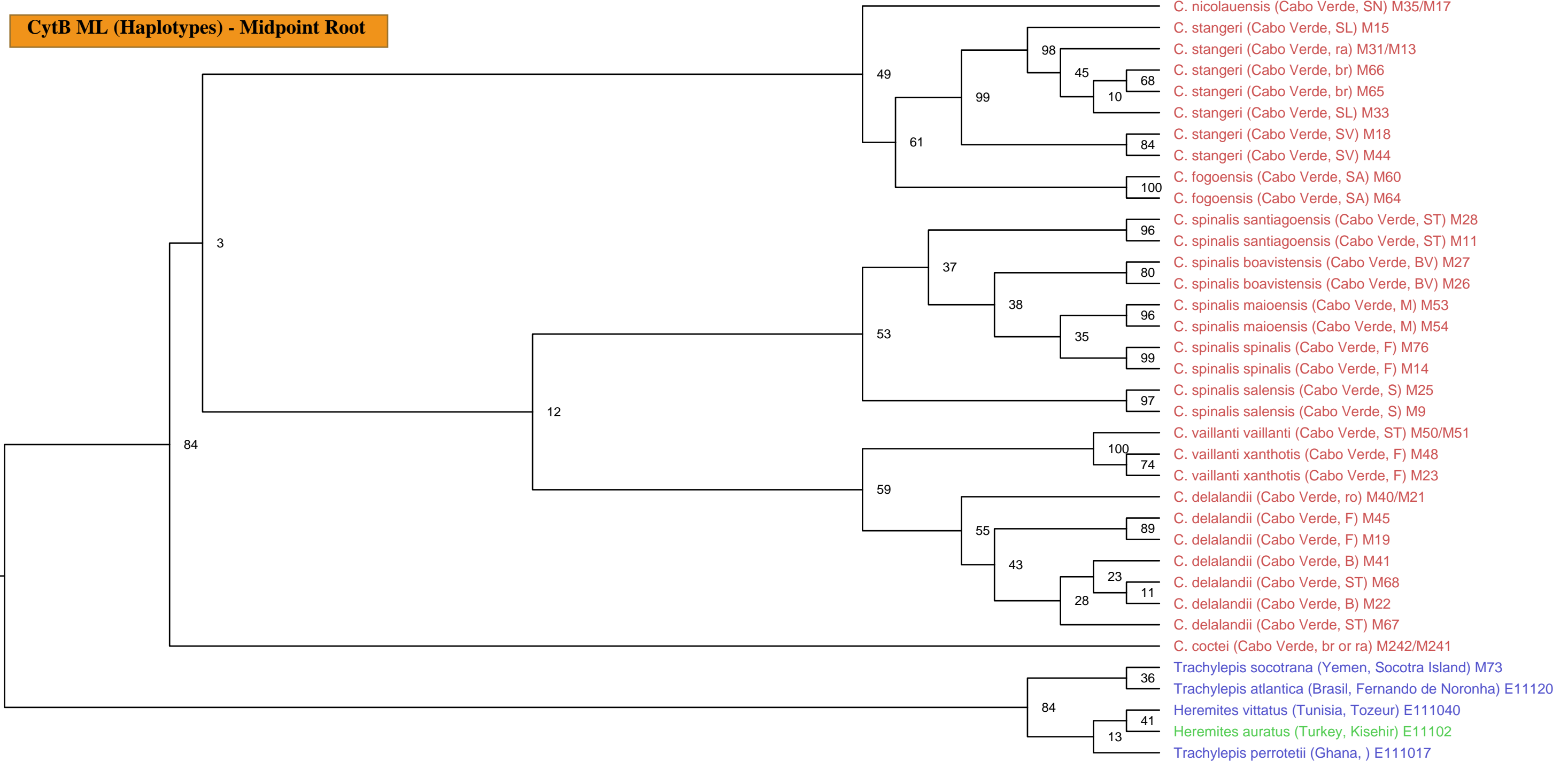
**CytB BI -
Midpoint Root
Without
Ambiguities**



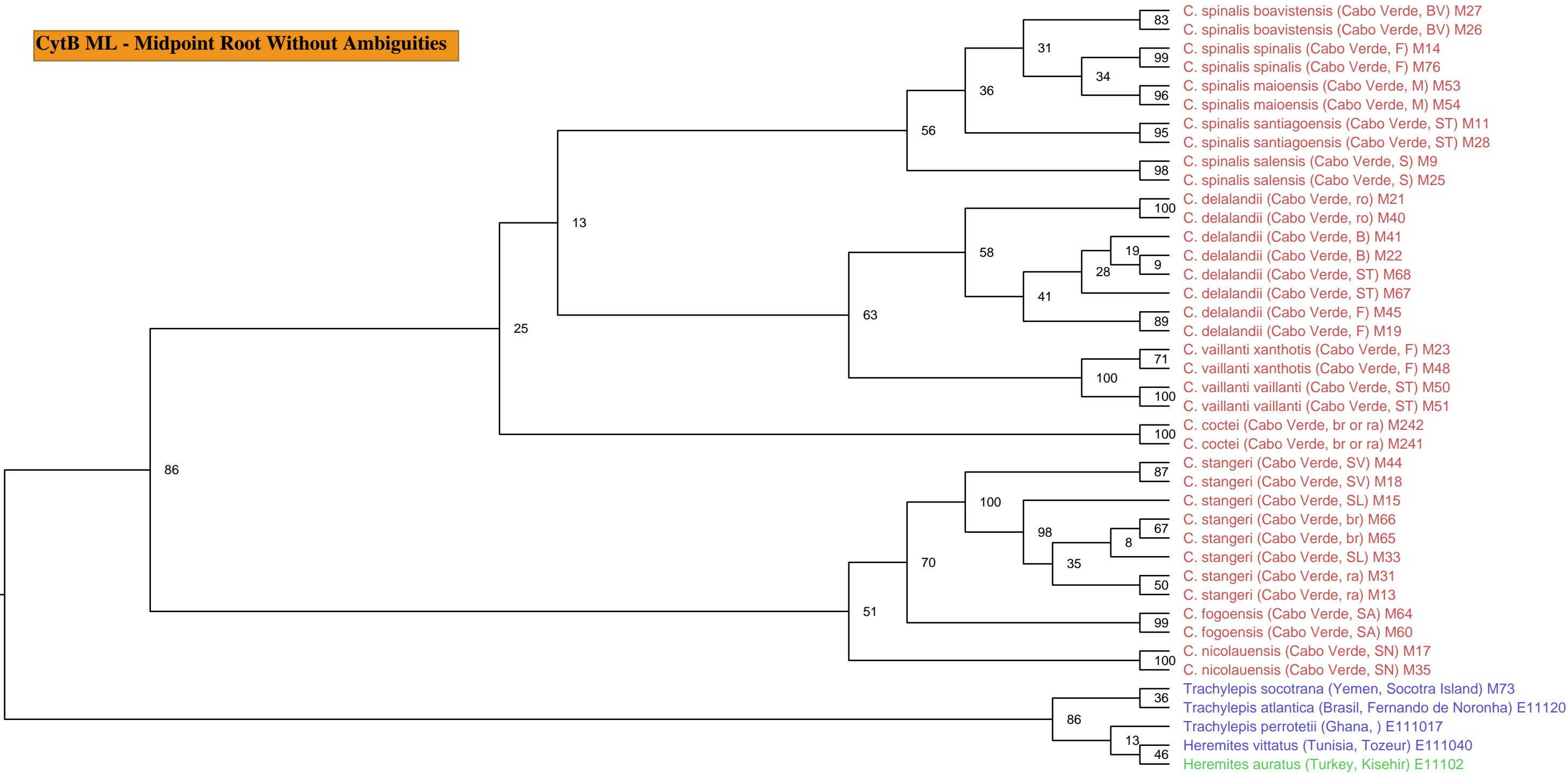
CytB ML - Without Ambiguities



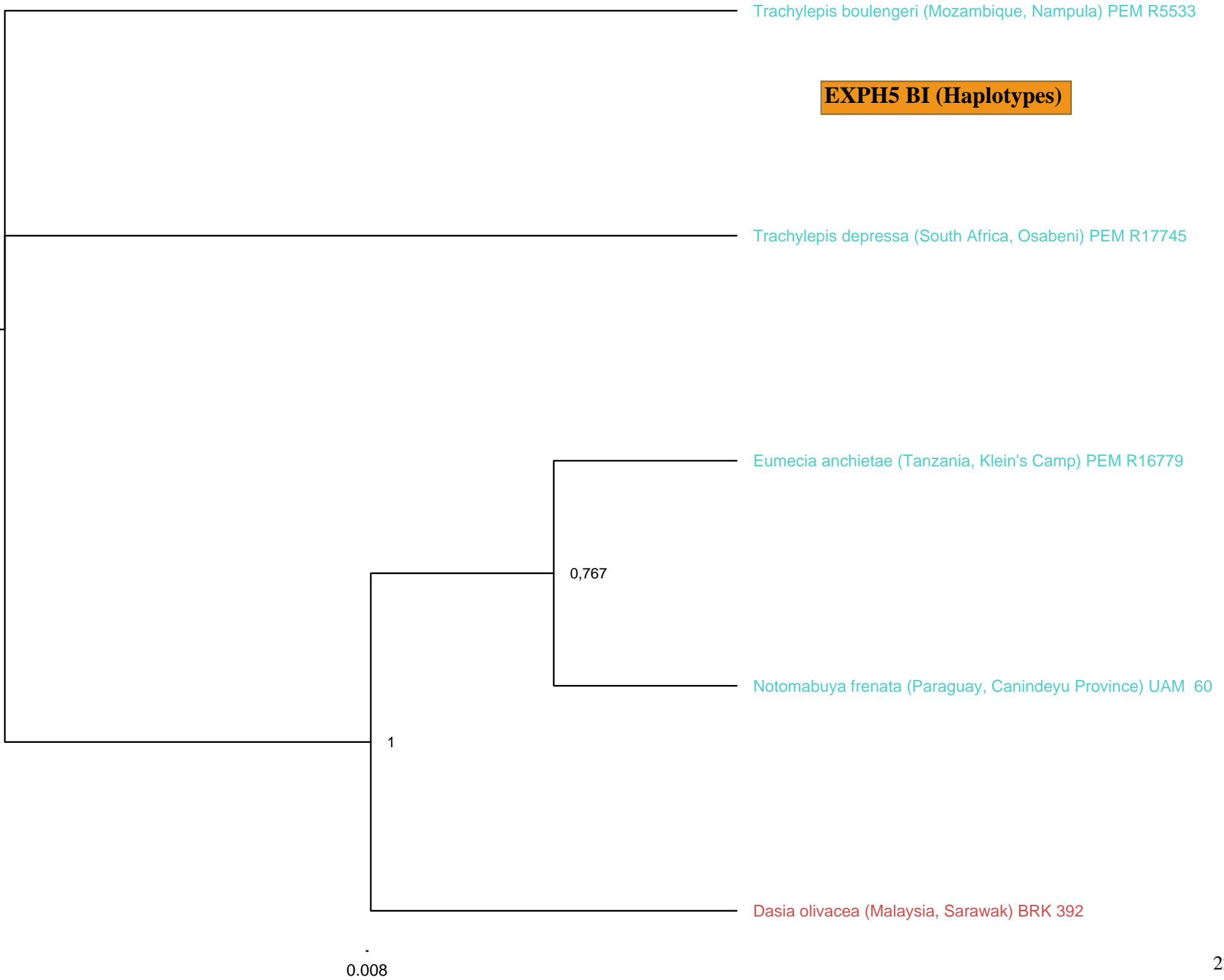
CytB ML (Haplotypes) - Midpoint Root



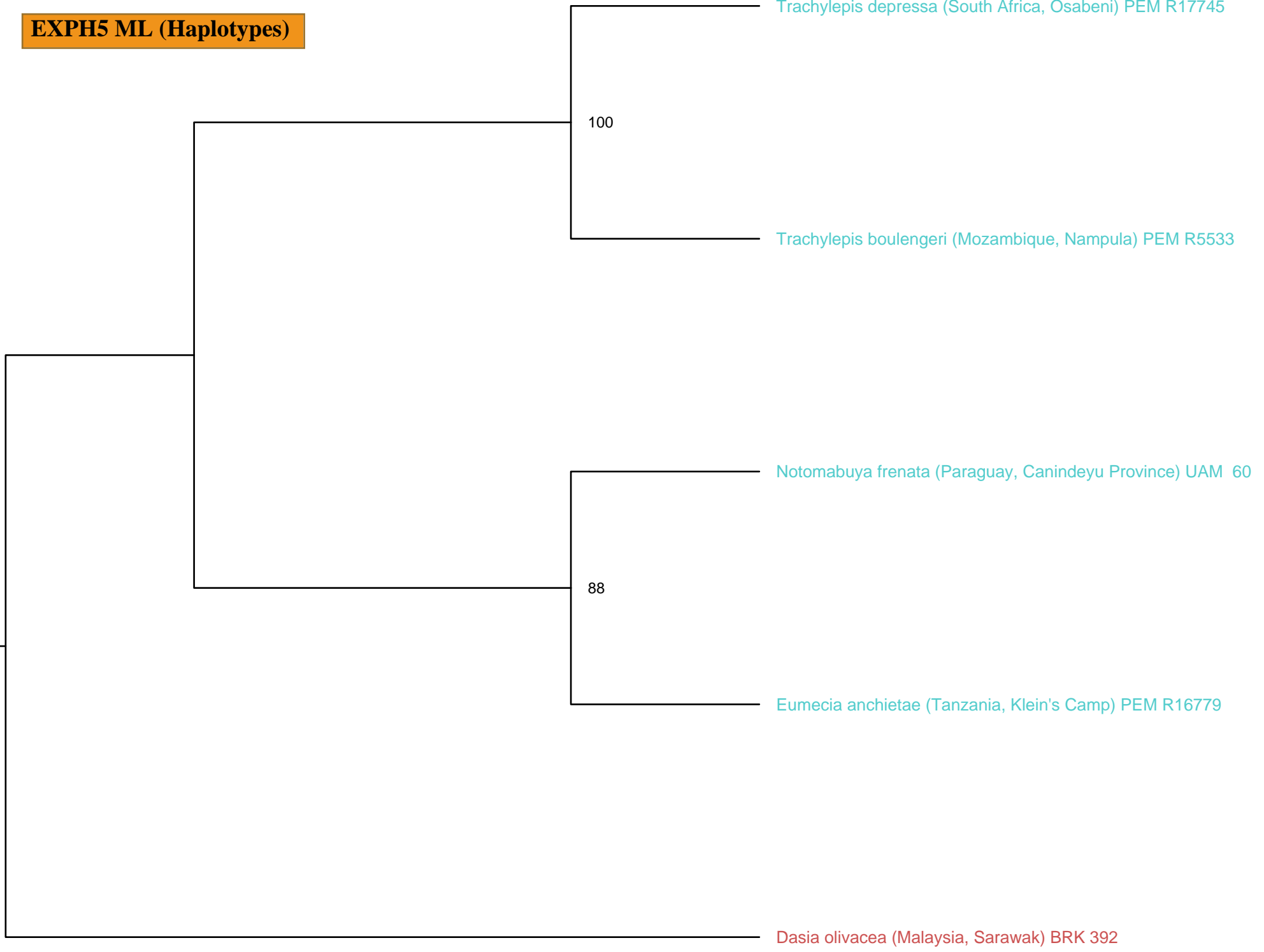
CytB ML - Midpoint Root Without Ambiguities



EXPH5 BI (Haplotypes)

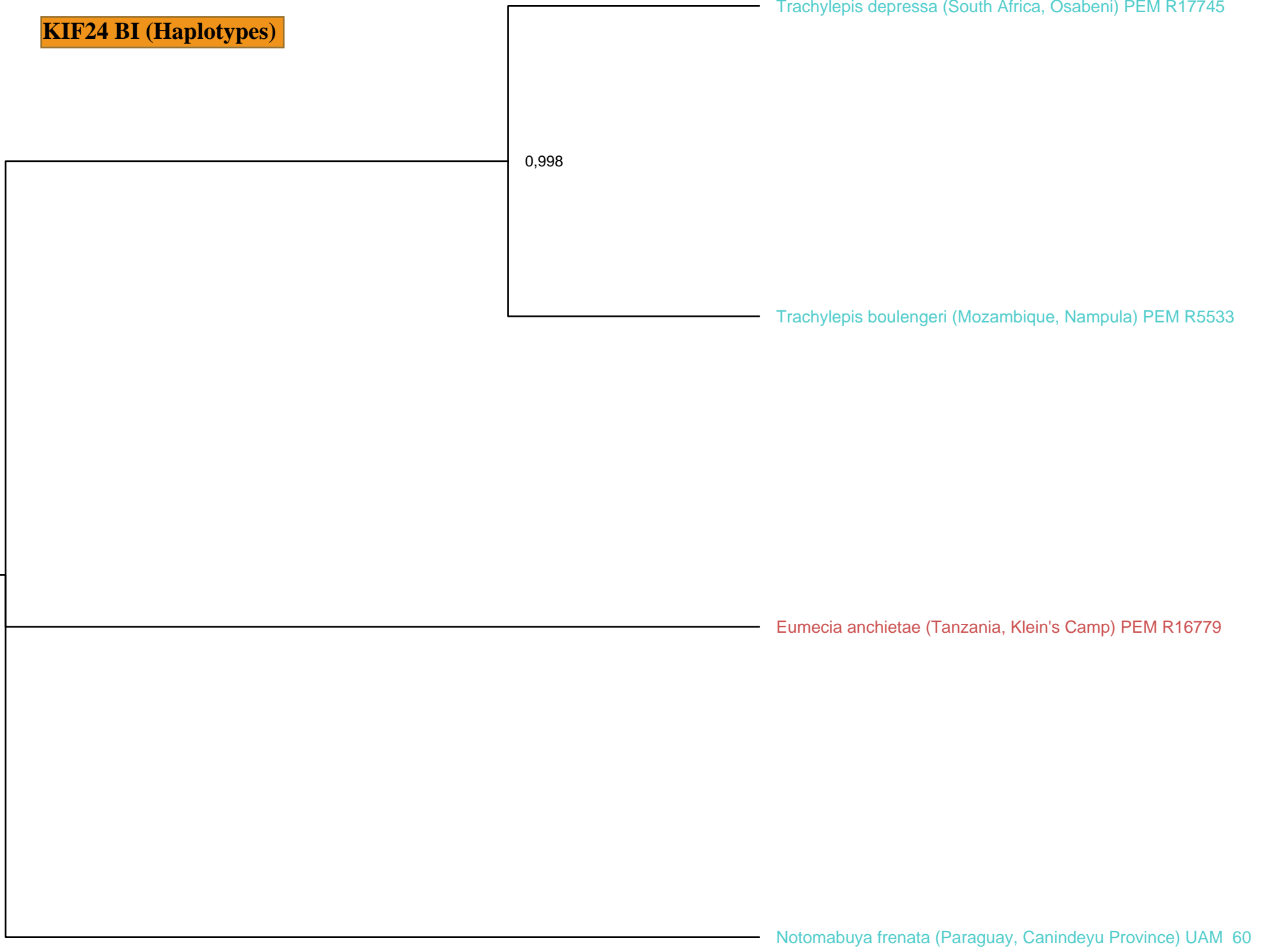


EXPH5 ML (Haplotypes)

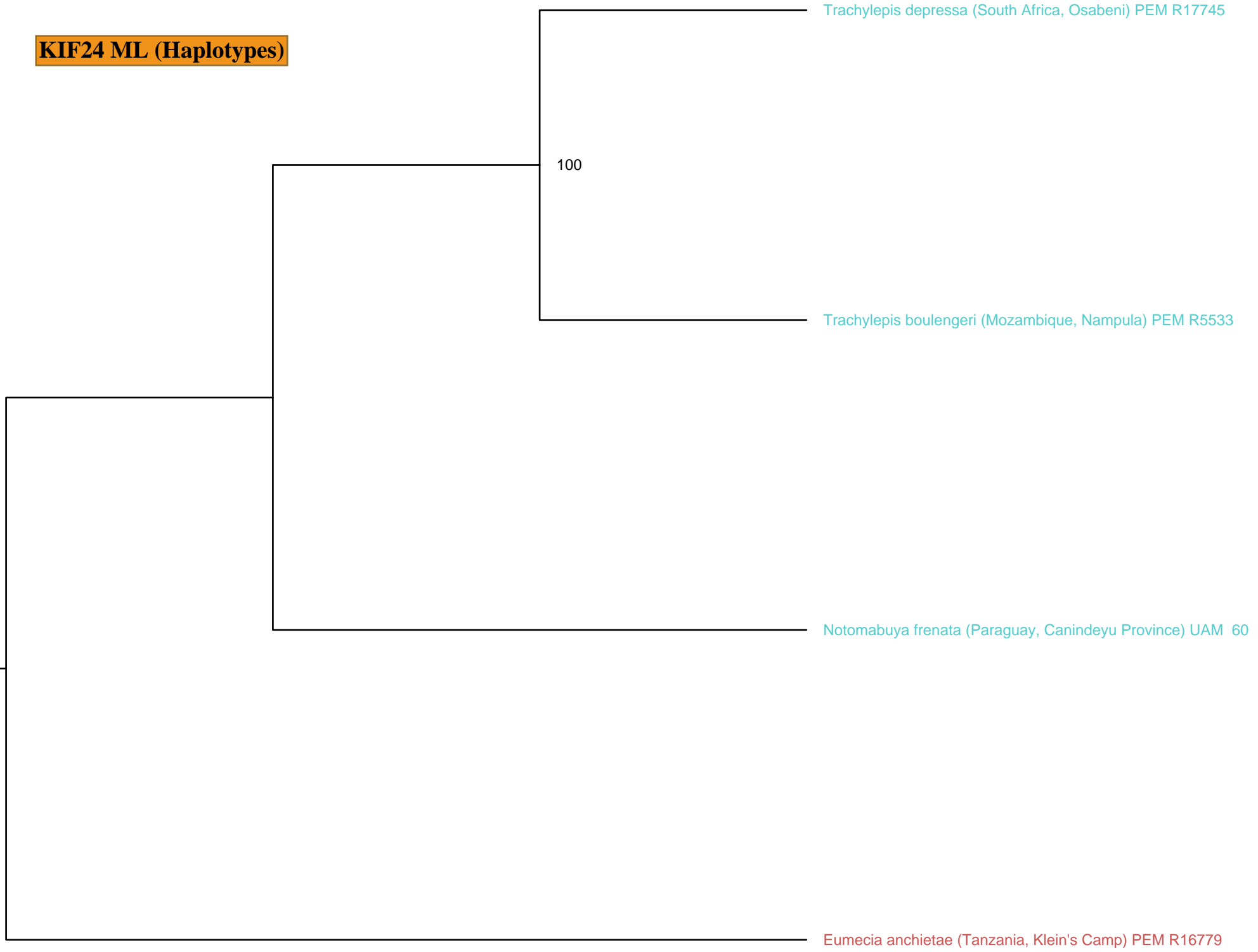


0.01

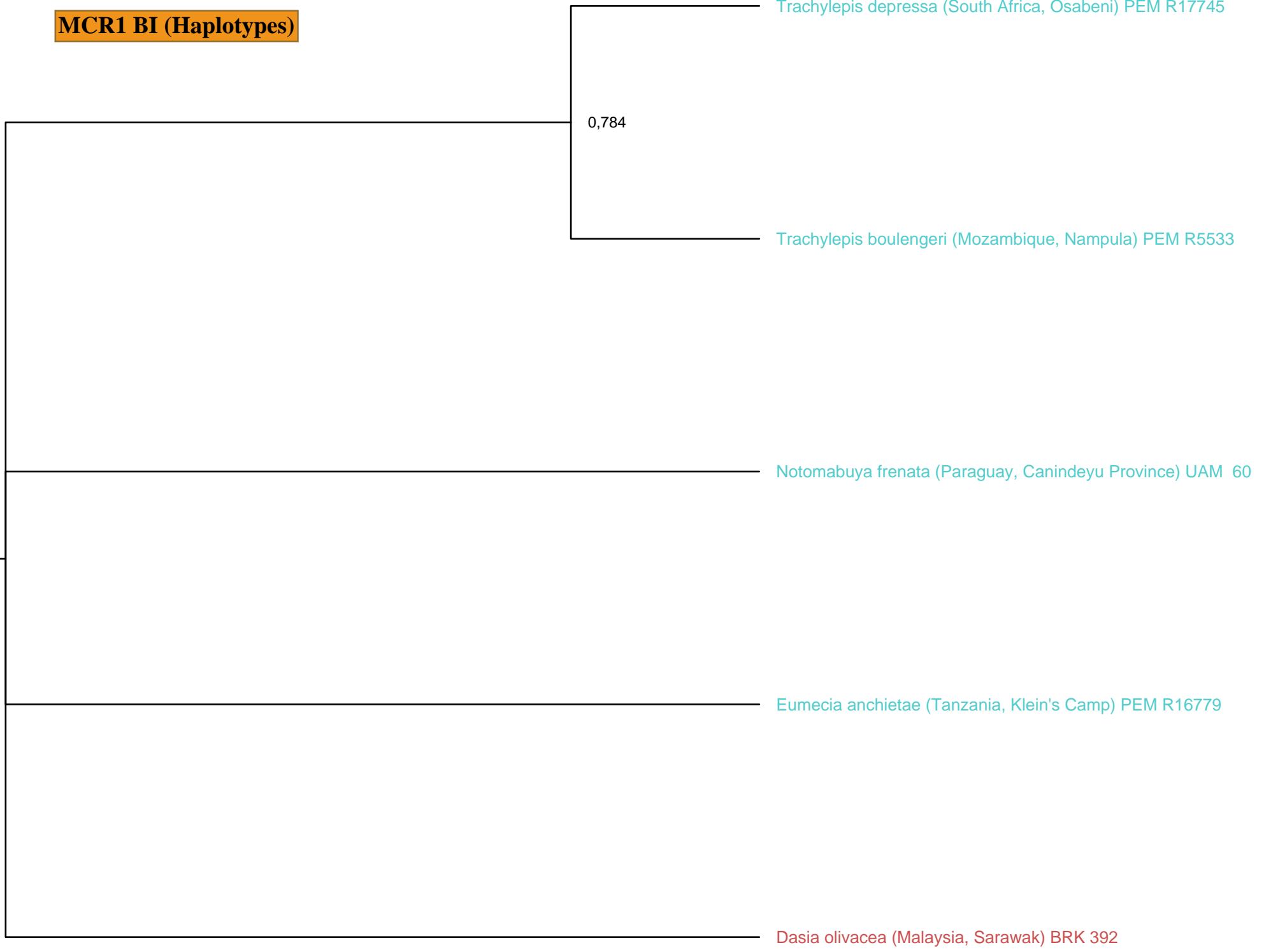
KIF24 BI (Haplotypes)



KIF24 ML (Haplotypes)

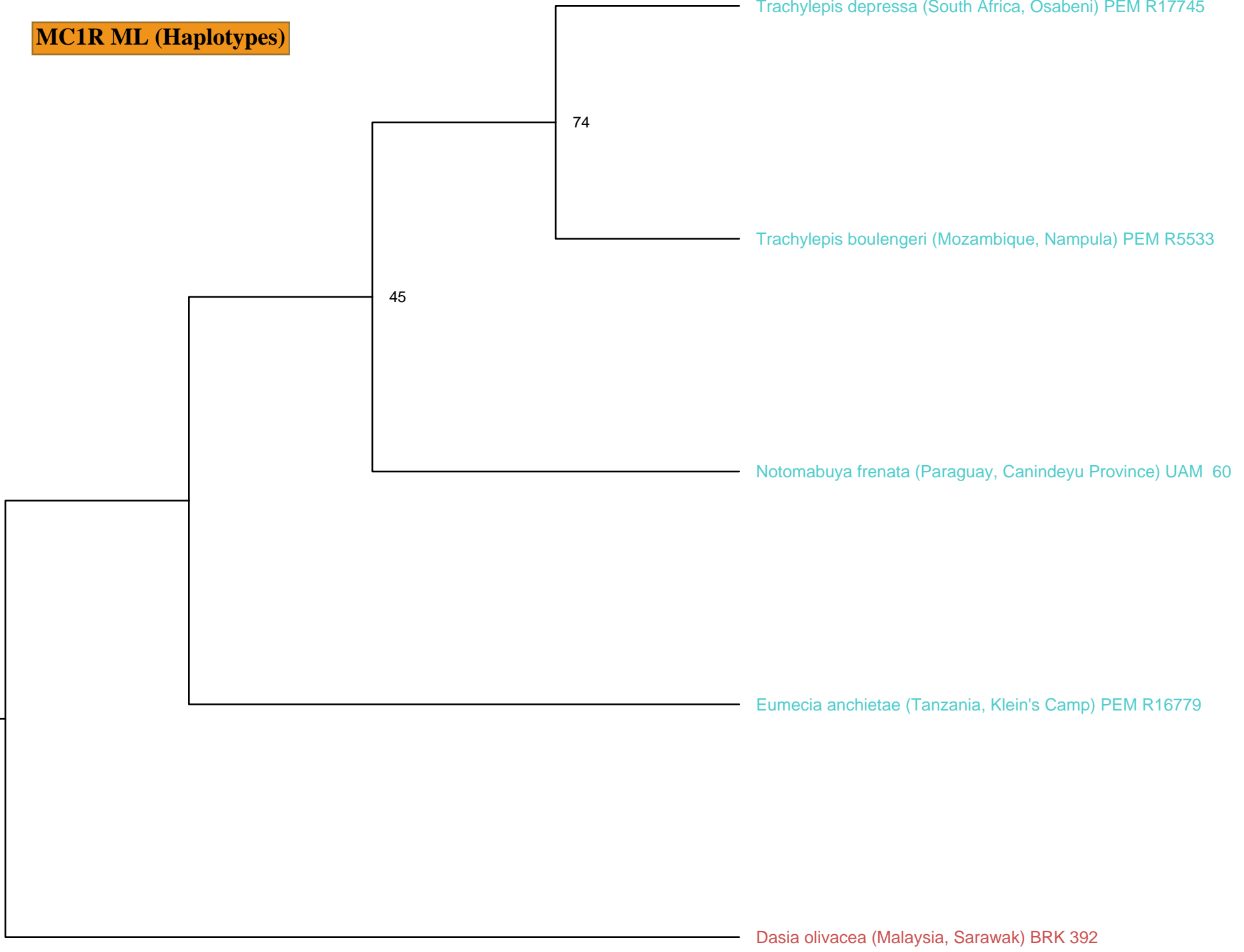


MCR1 BI (Haplotypes)



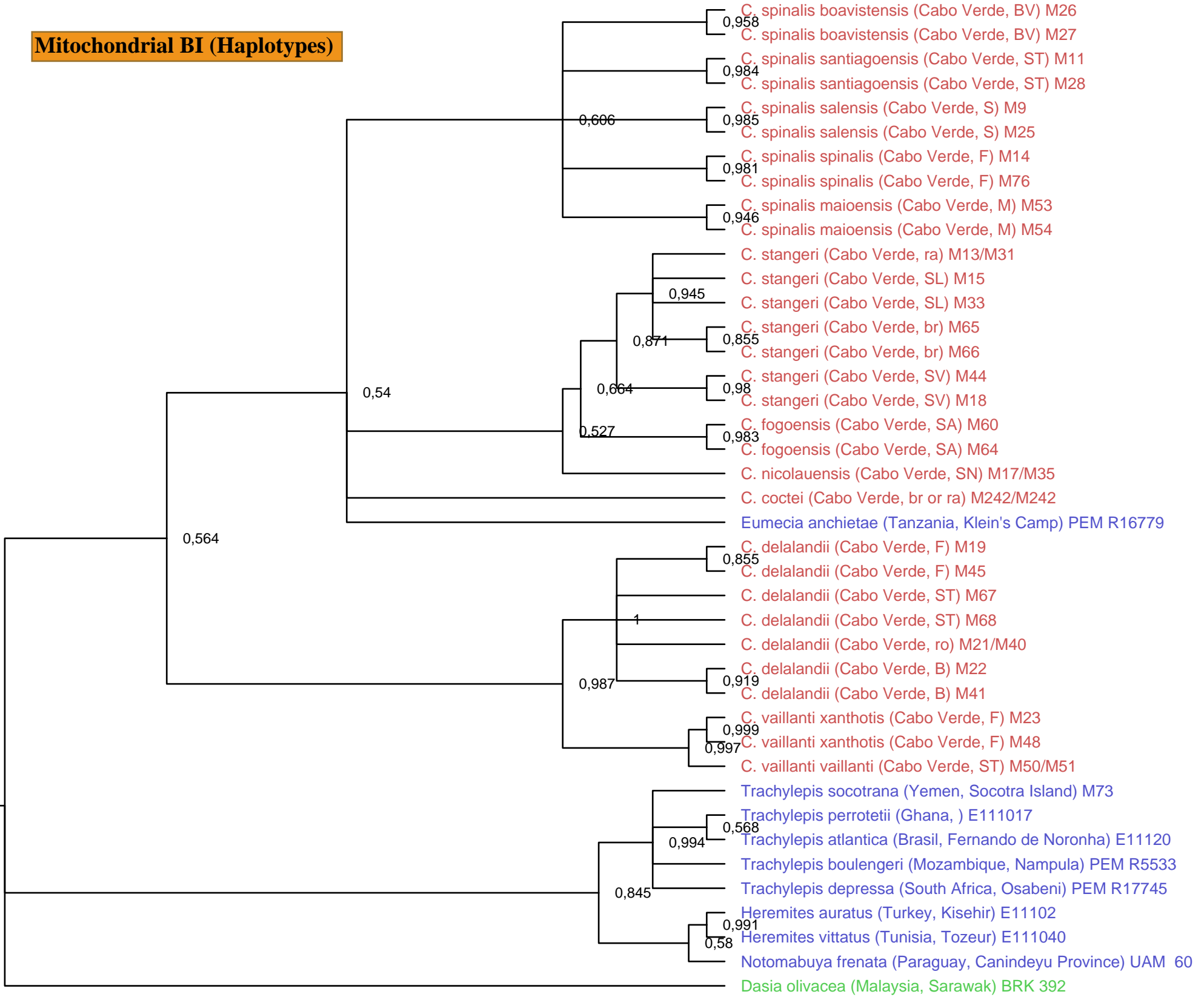
0.5

MC1R ML (Haplotypes)

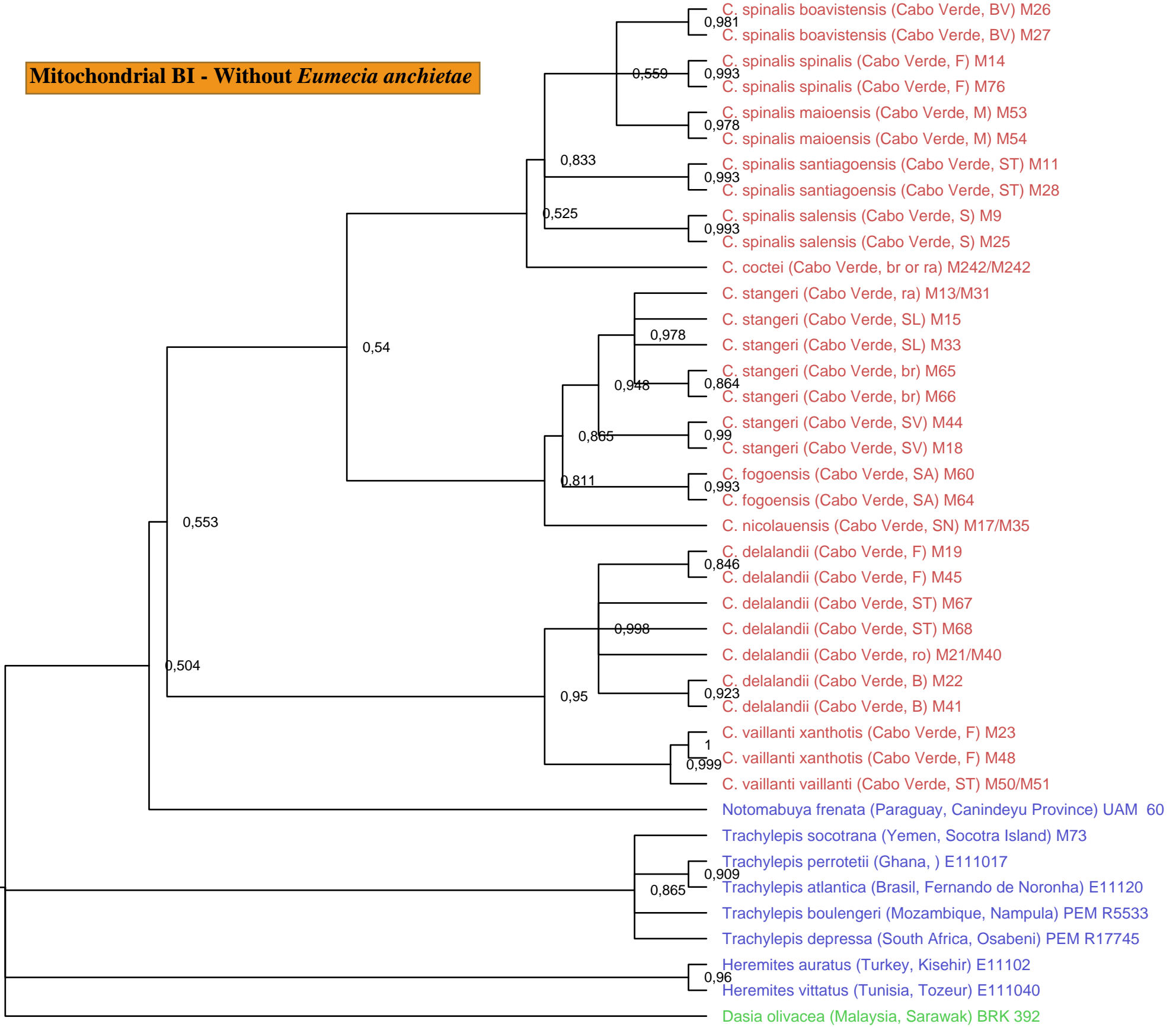


0.5

Mitochondrial BI (Haplotypes)

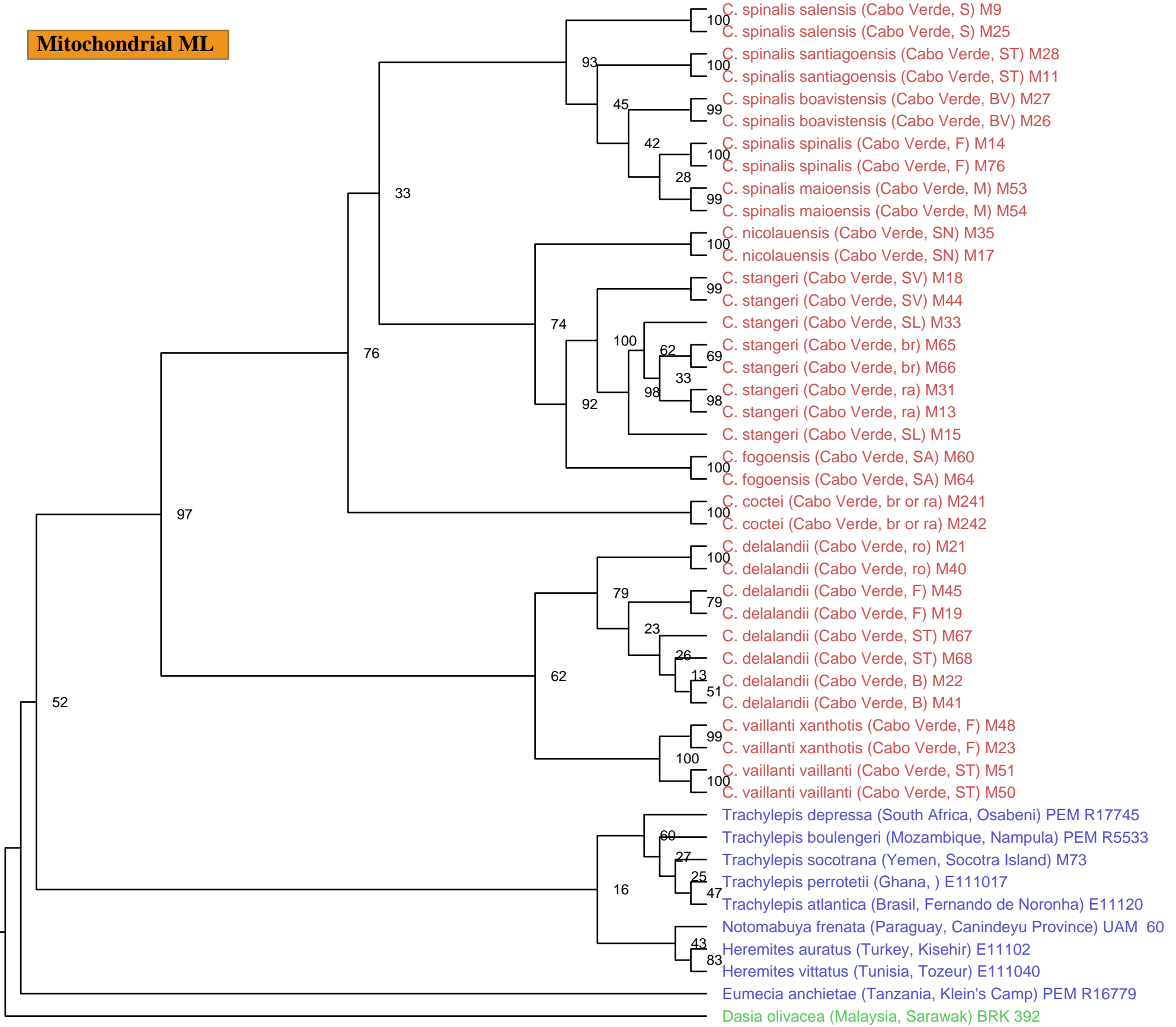


Mitochondrial BI - Without *Eumecia anchietae*

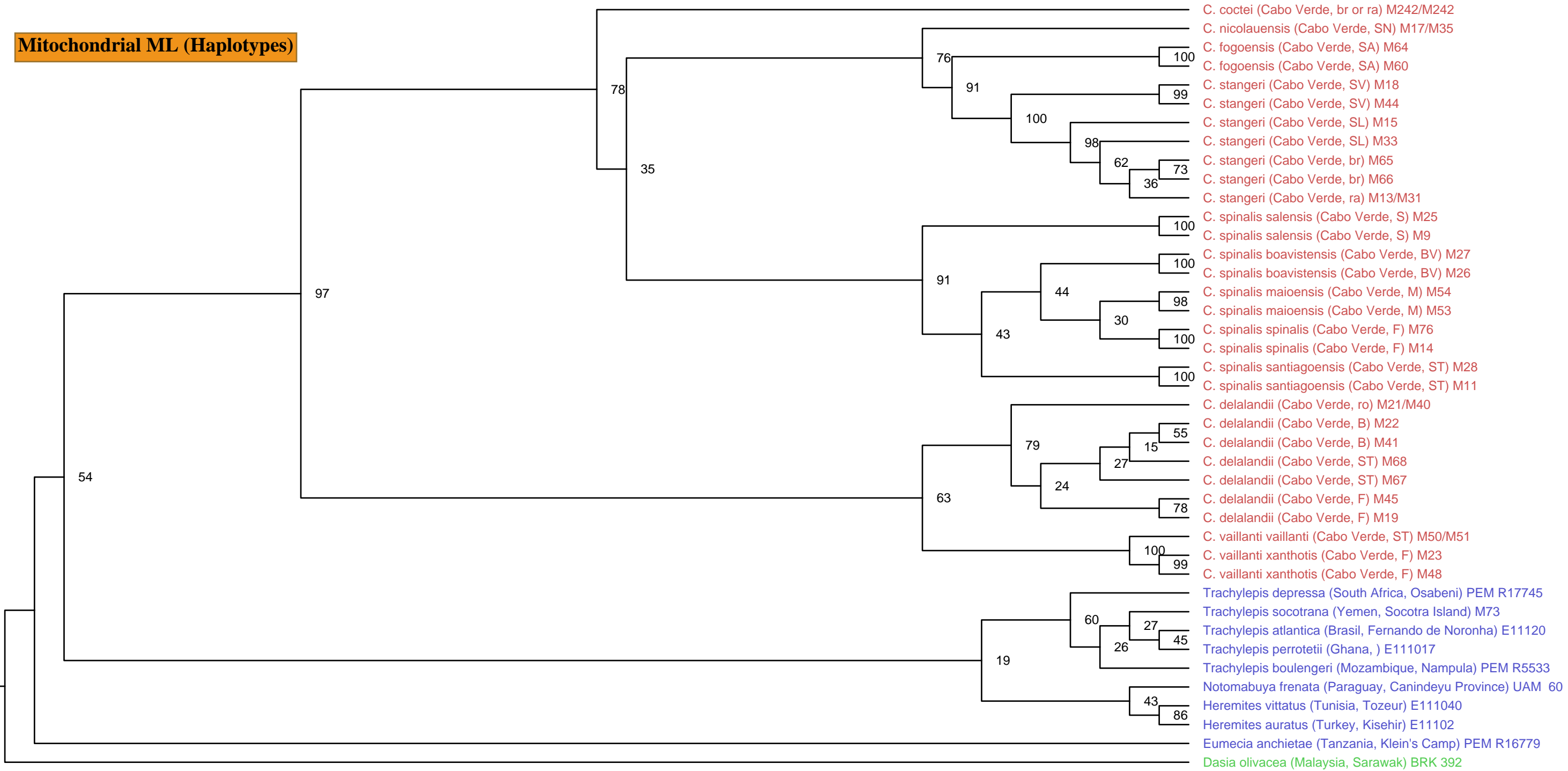


0.04

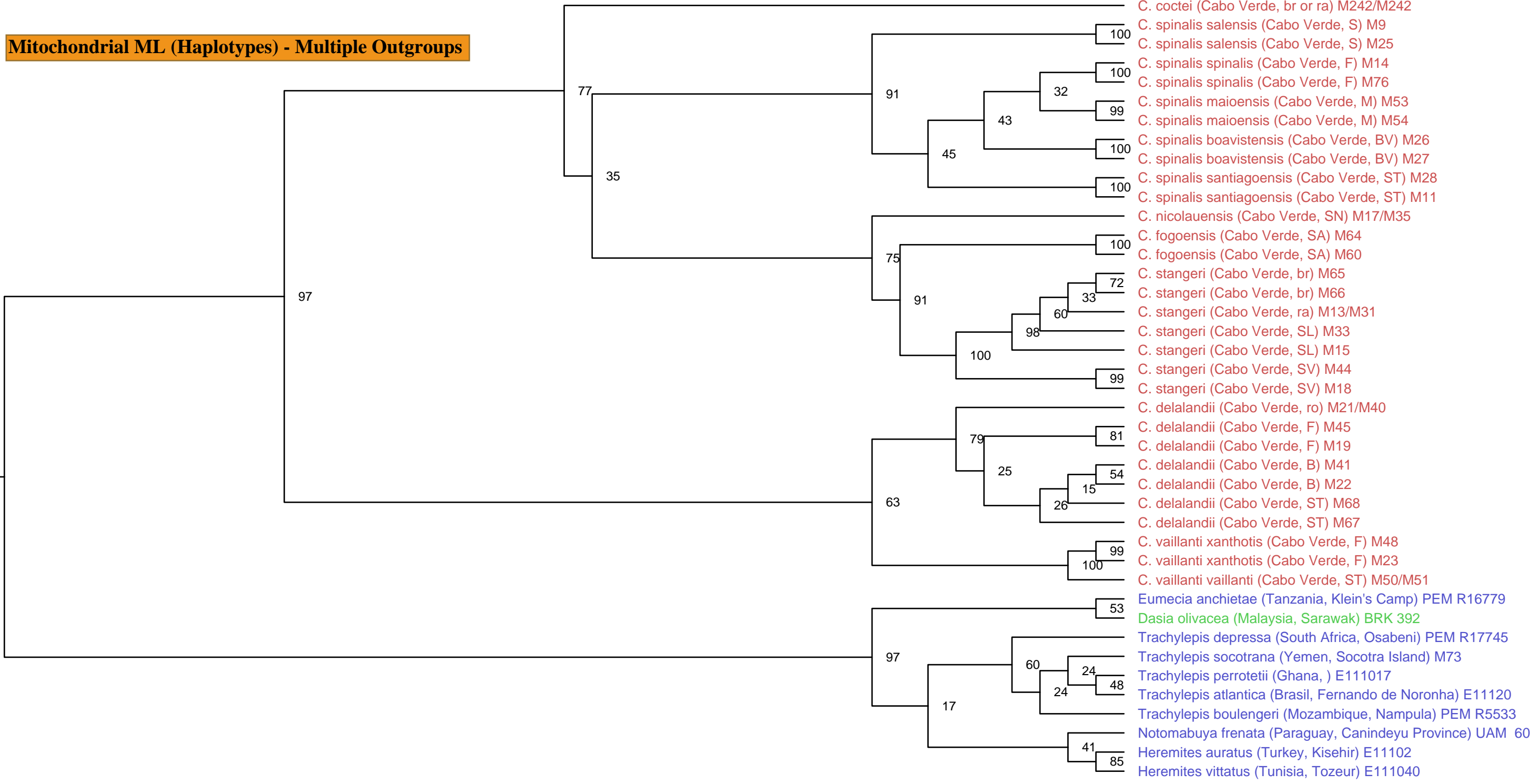
Mitochondrial ML



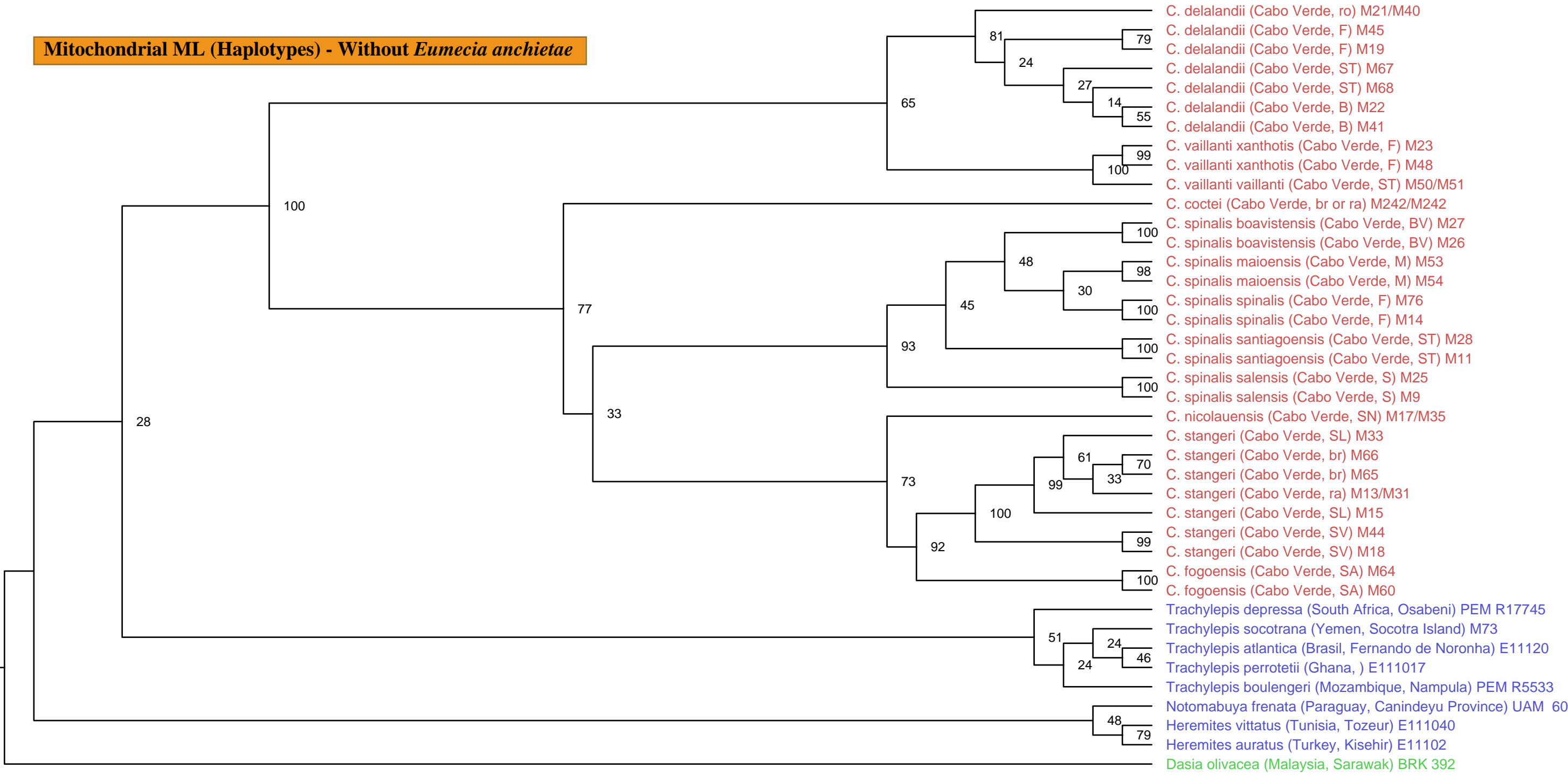
Mitochondrial ML (Haplotypes)



Mitochondrial ML (Haplotypes) - Multiple Outgroups

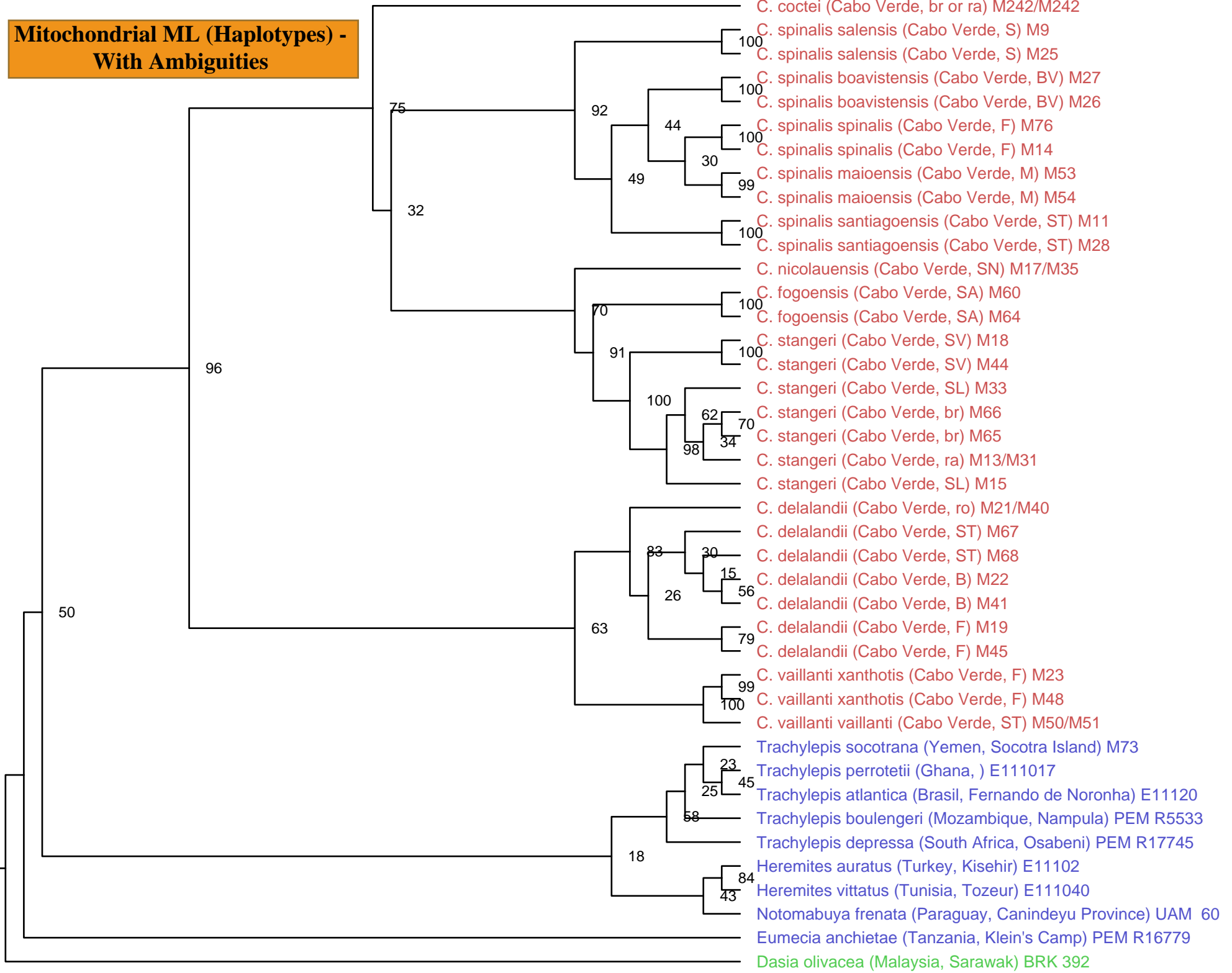


Mitochondrial ML (Haplotypes) - Without *Eumecia anchietae*



0.06

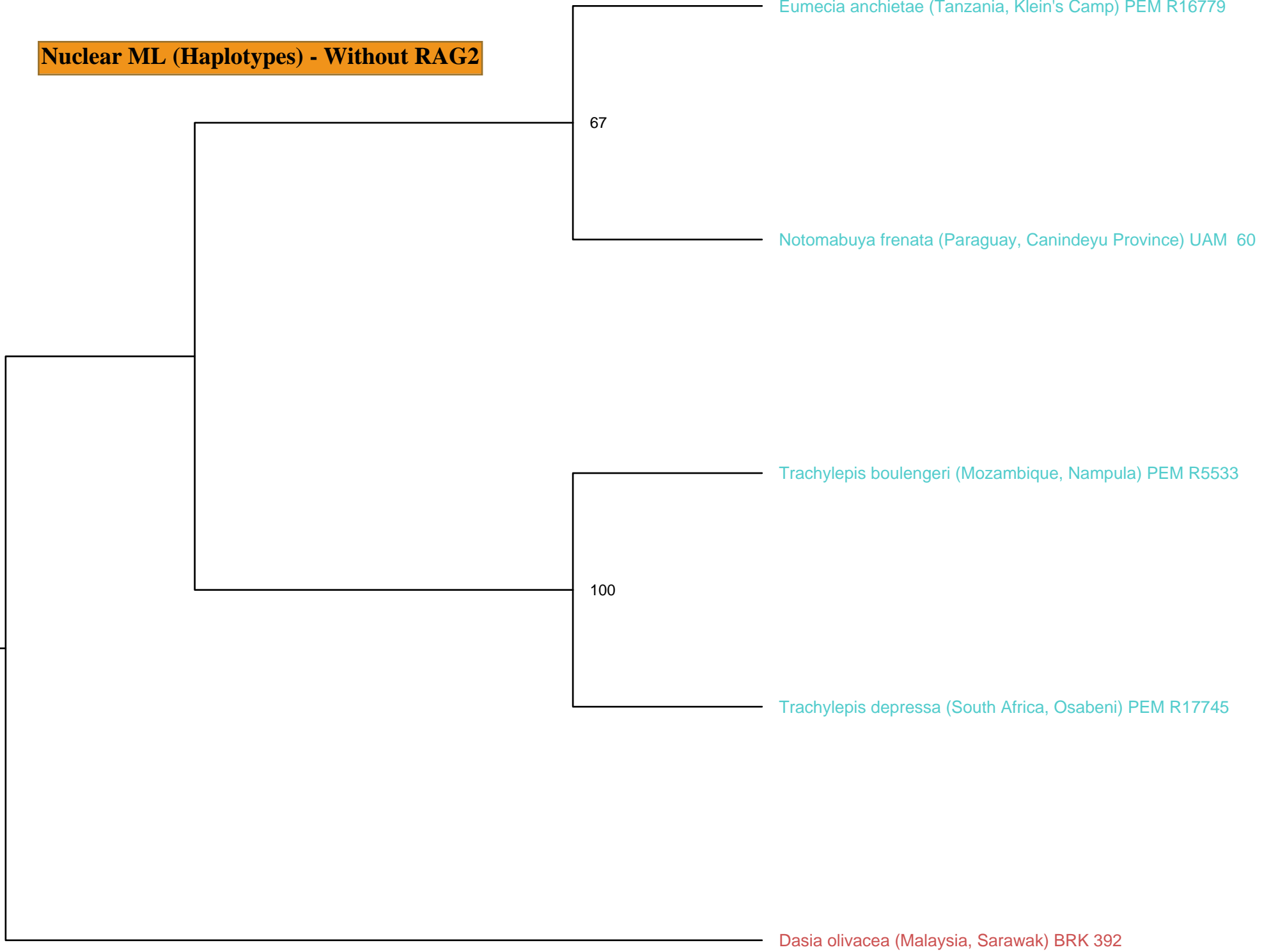
**Mitochondrial ML (Haplotypes) -
With Ambiguities**



Nuclear BI (Haplotypes)

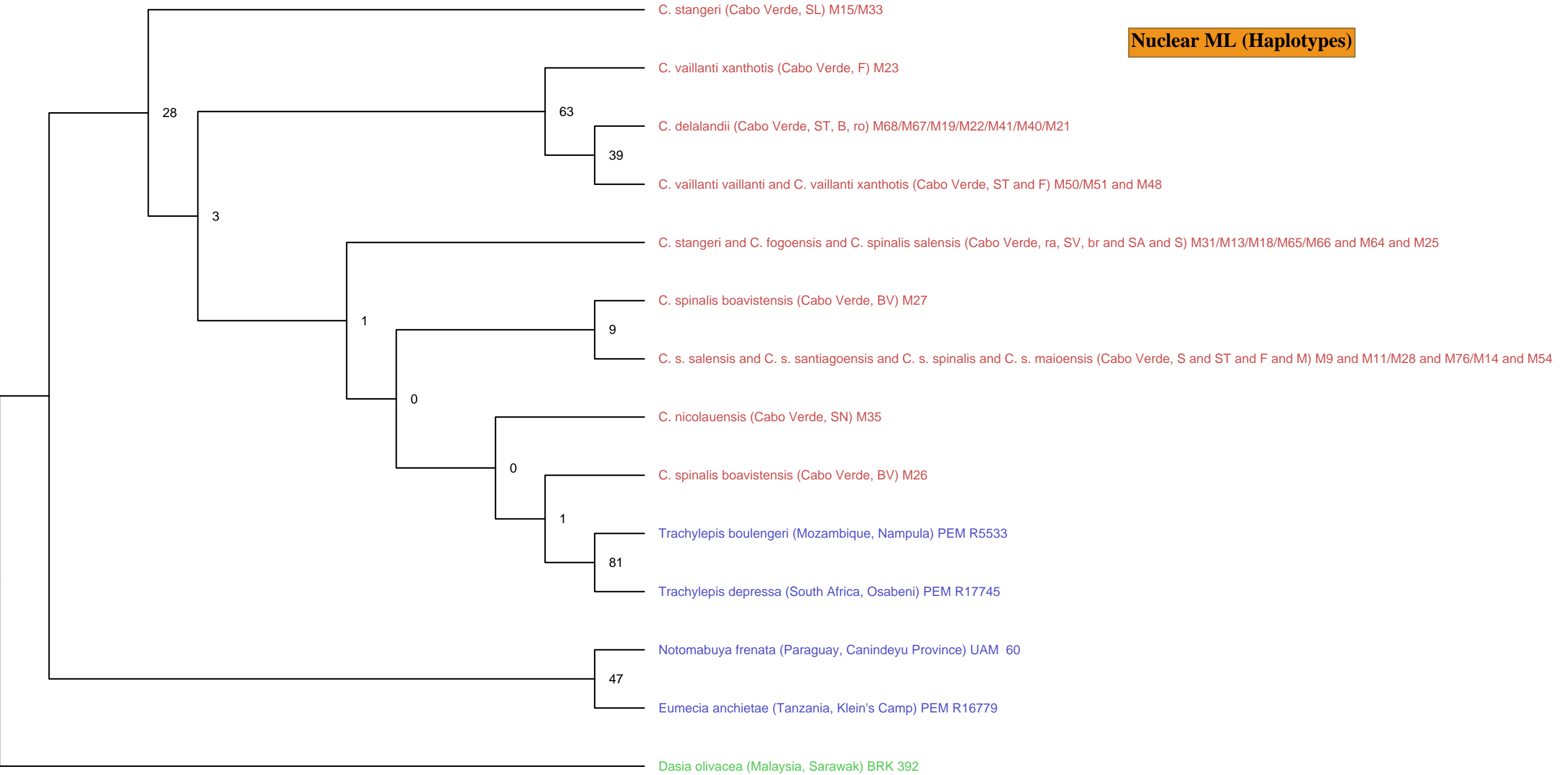


Nuclear ML (Haplotypes) - Without RAG2



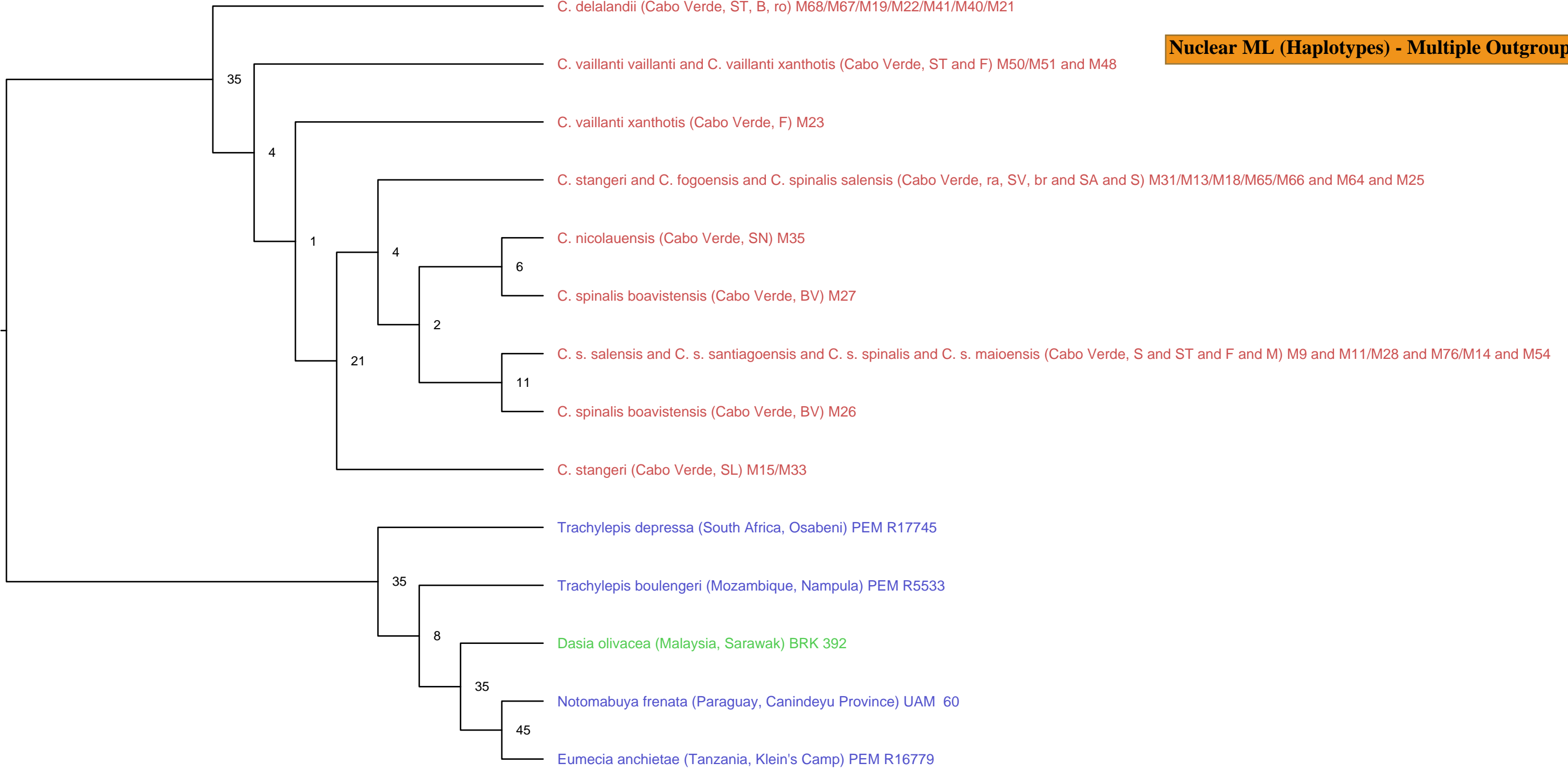
0.5

Nuclear ML (Haplotypes)



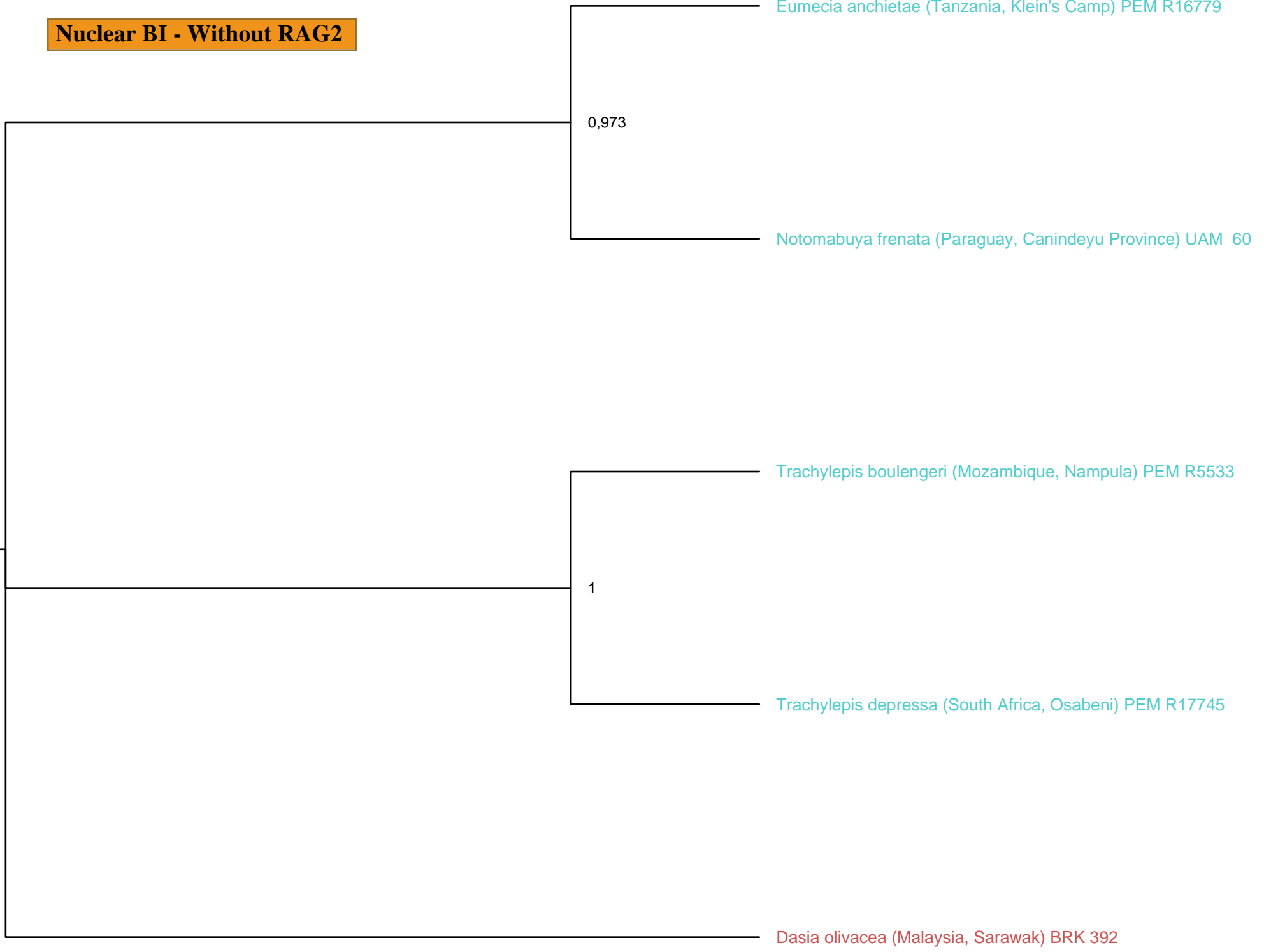
2.0

Nuclear ML (Haplotypes) - Multiple Outgroups

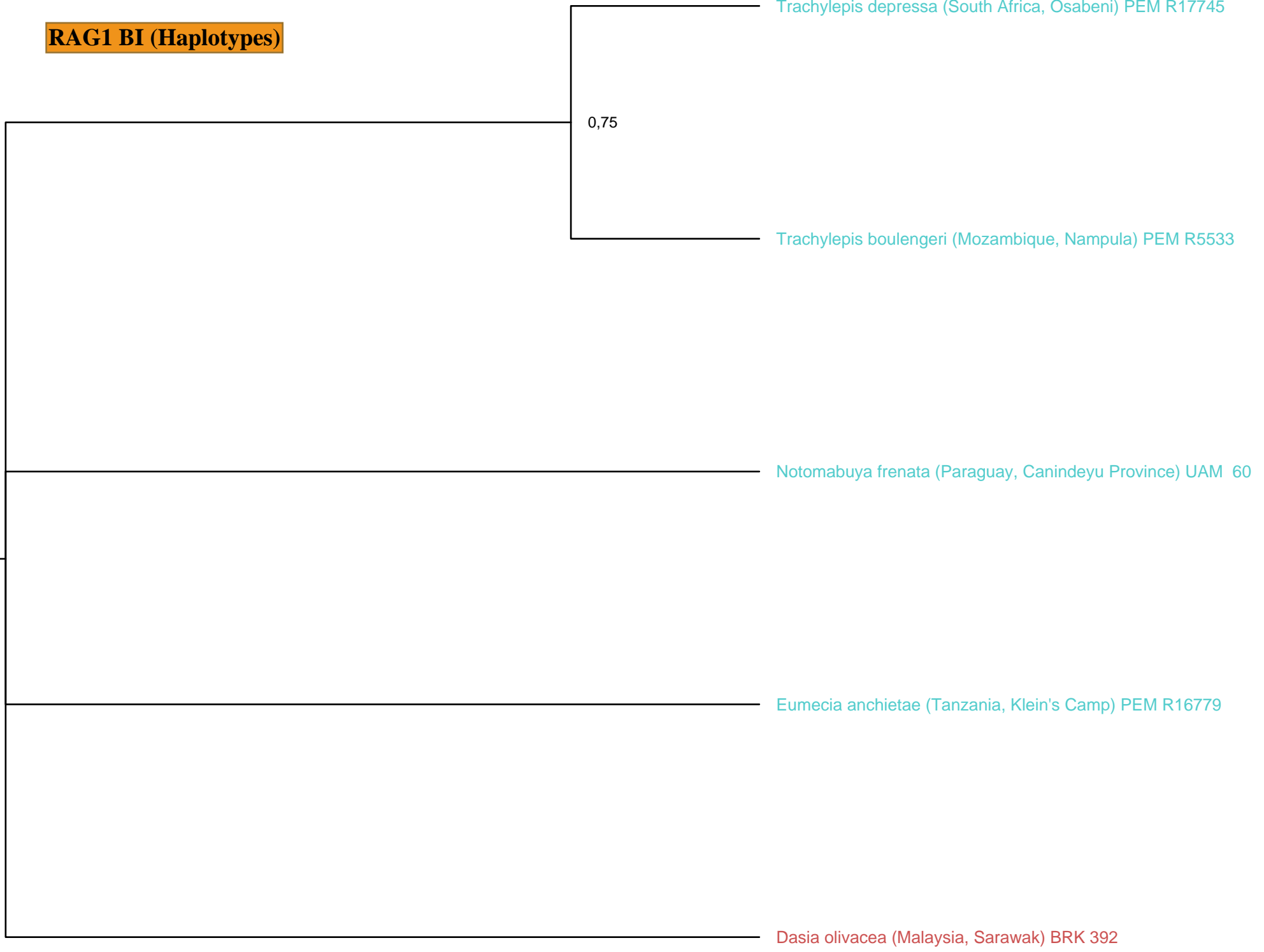


3.0

Nuclear BI - Without RAG2

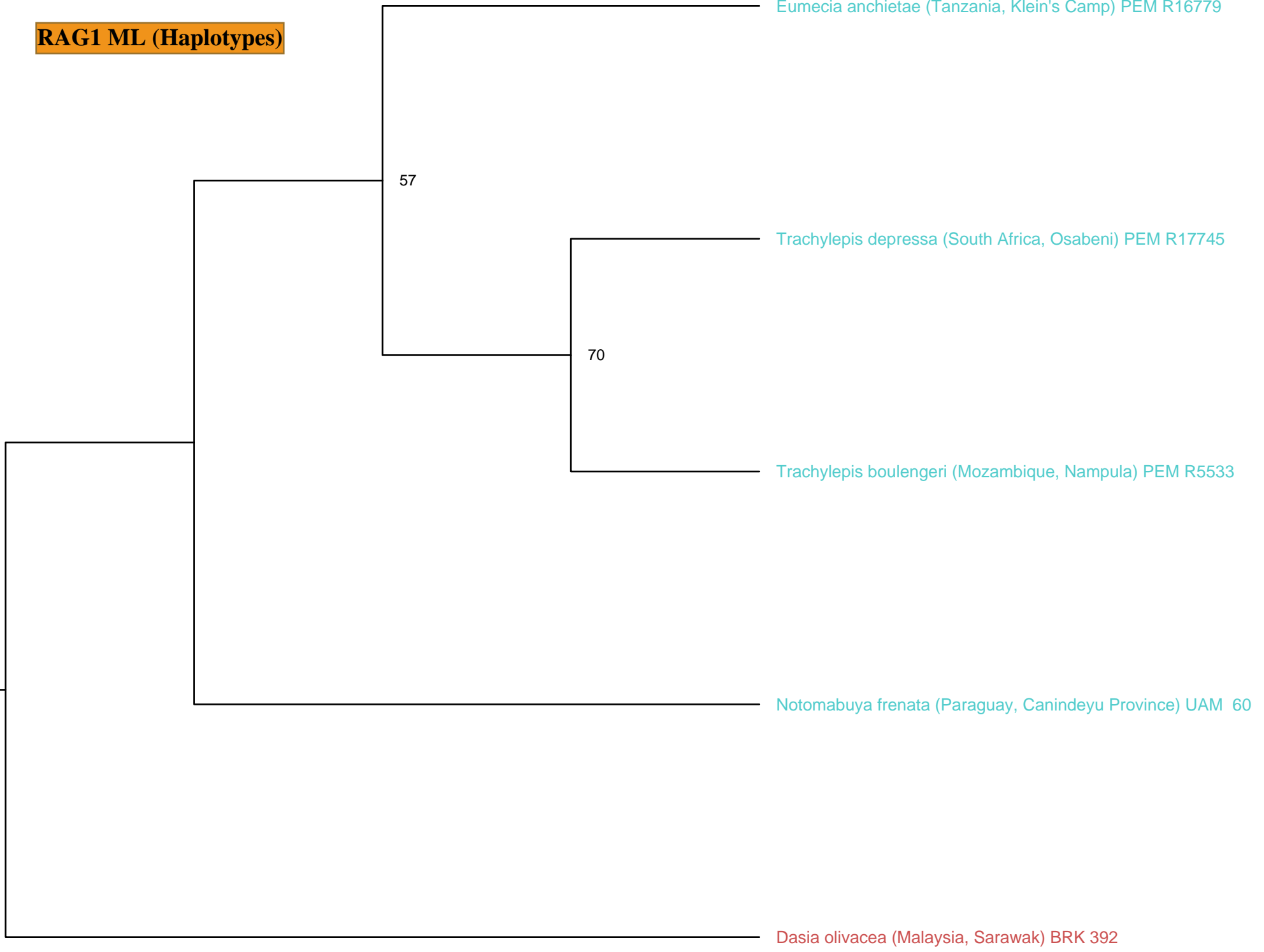


RAG1 BI (Haplotypes)



0.007

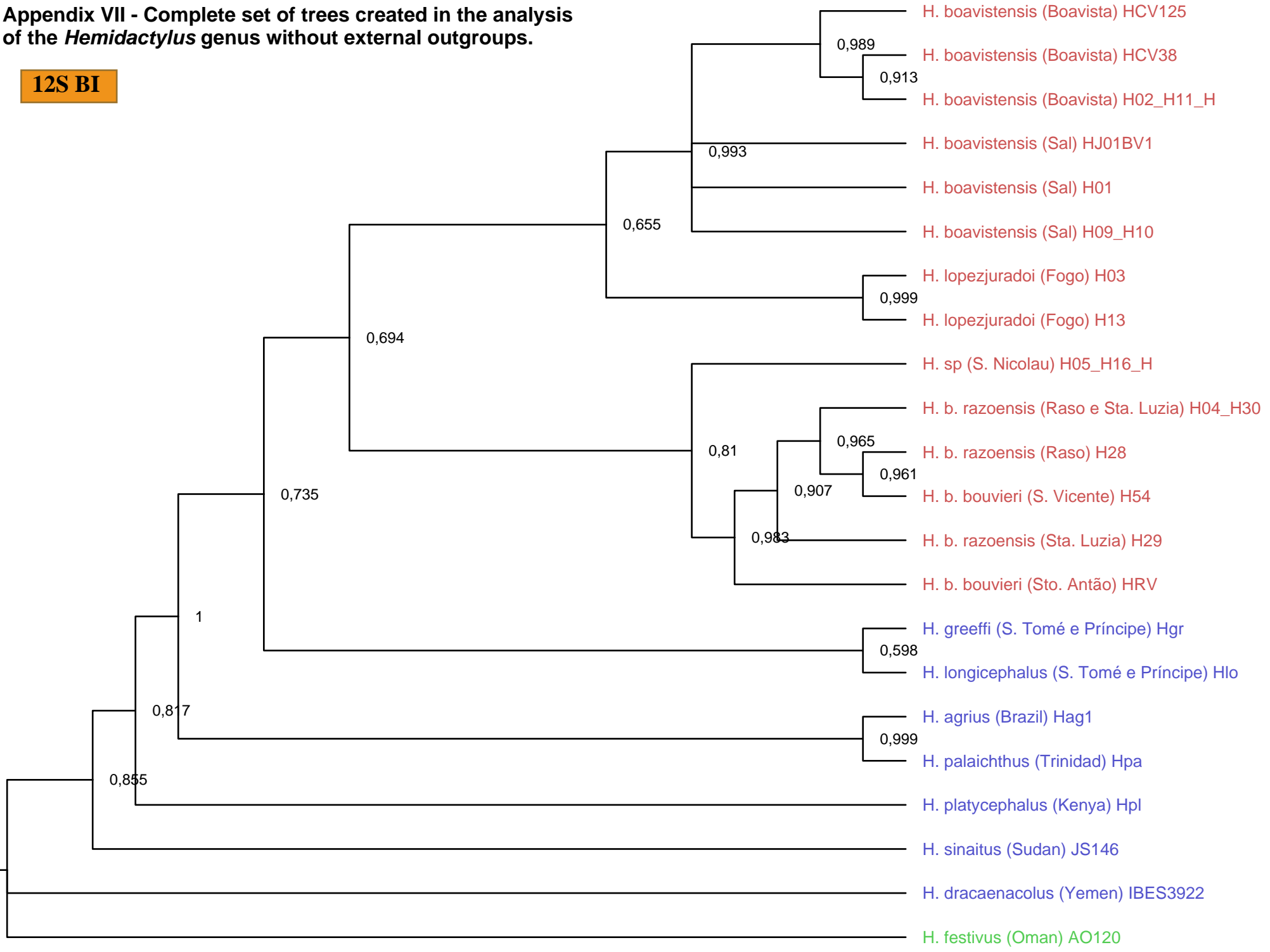
RAG1 ML (Haplotypes)



0.009

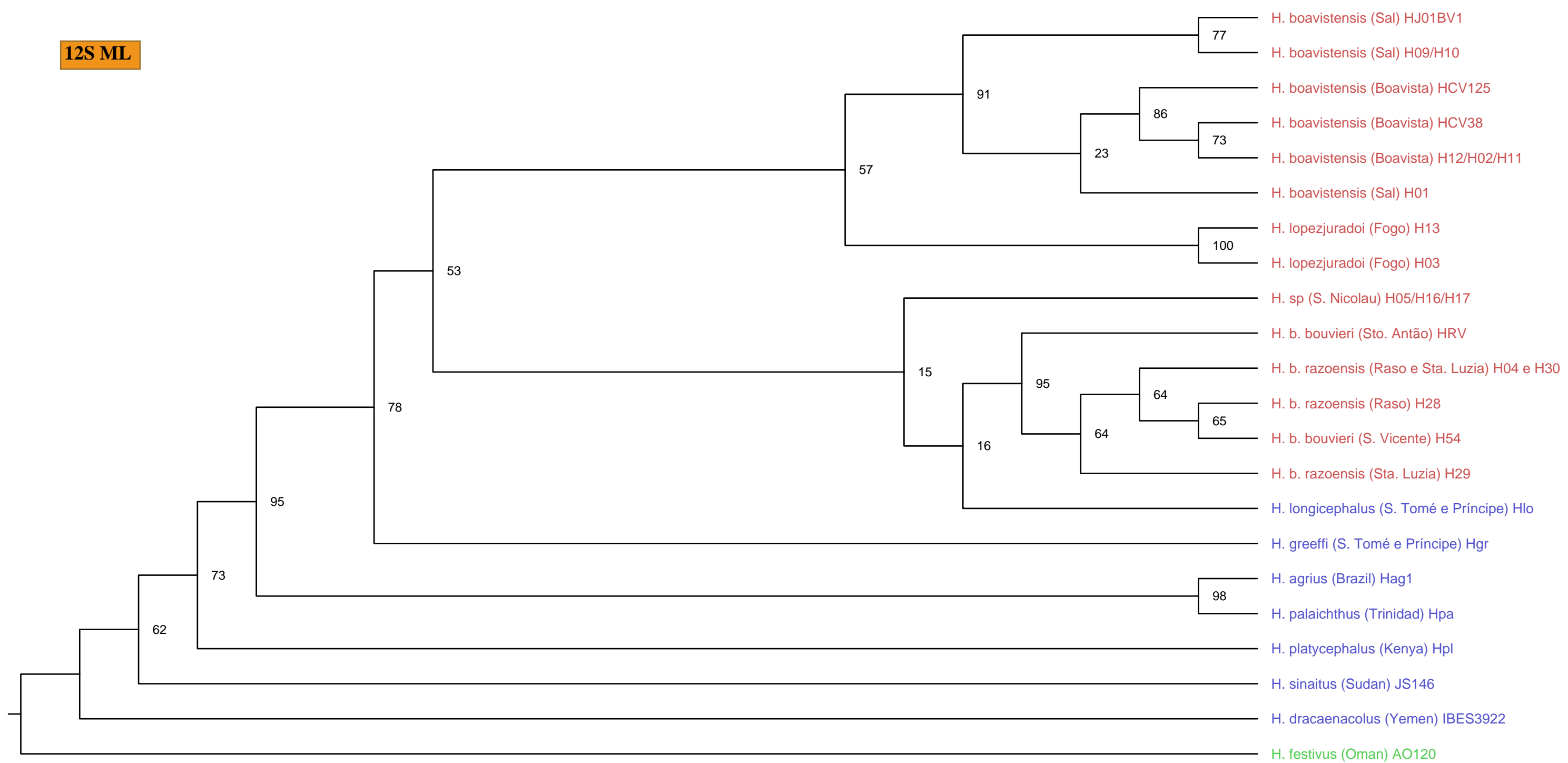
Appendix VII - Complete set of trees created in the analysis of the *Hemidactylus* genus without external outgroups.

12S BI



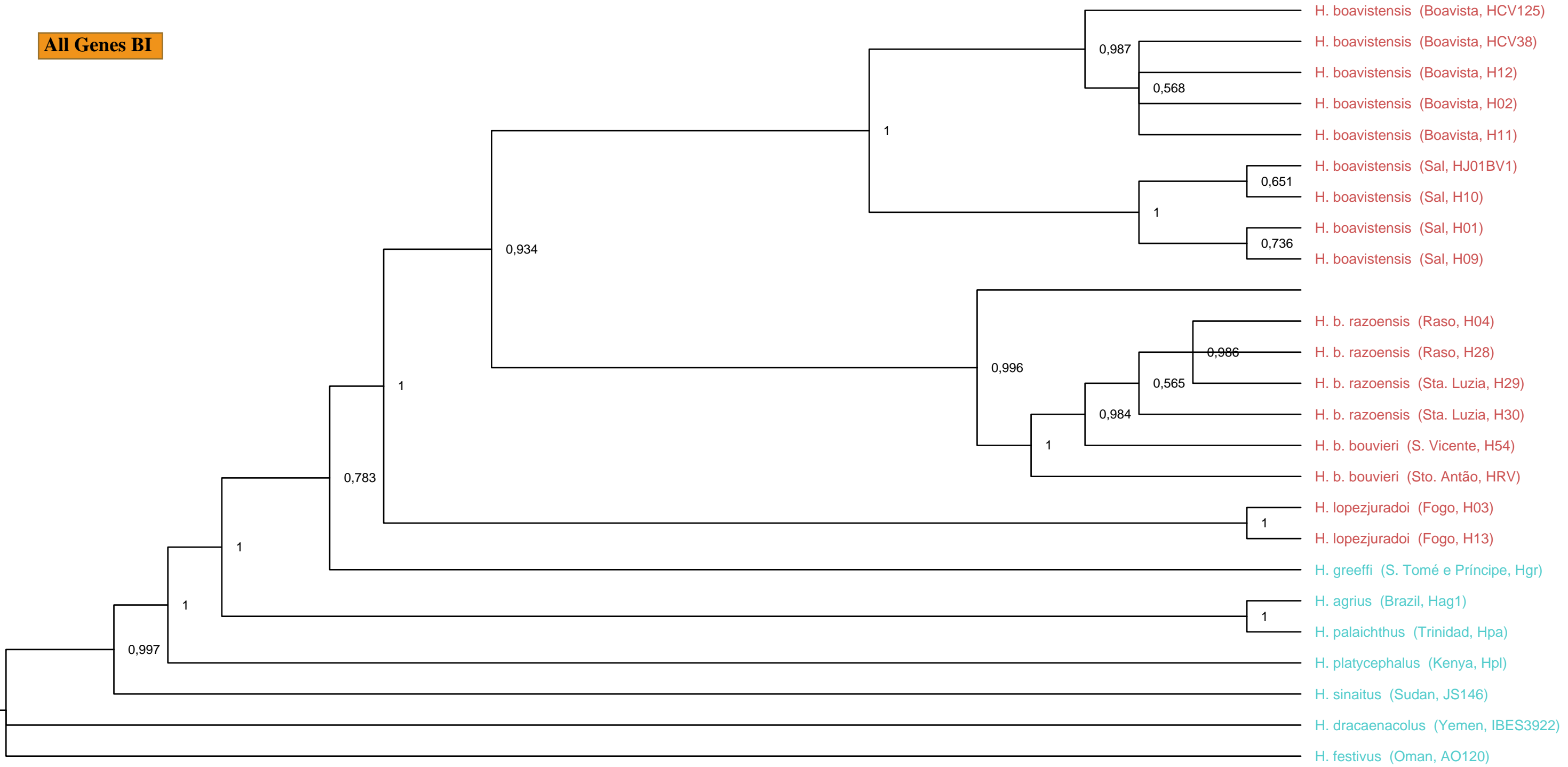
0.04

12S ML



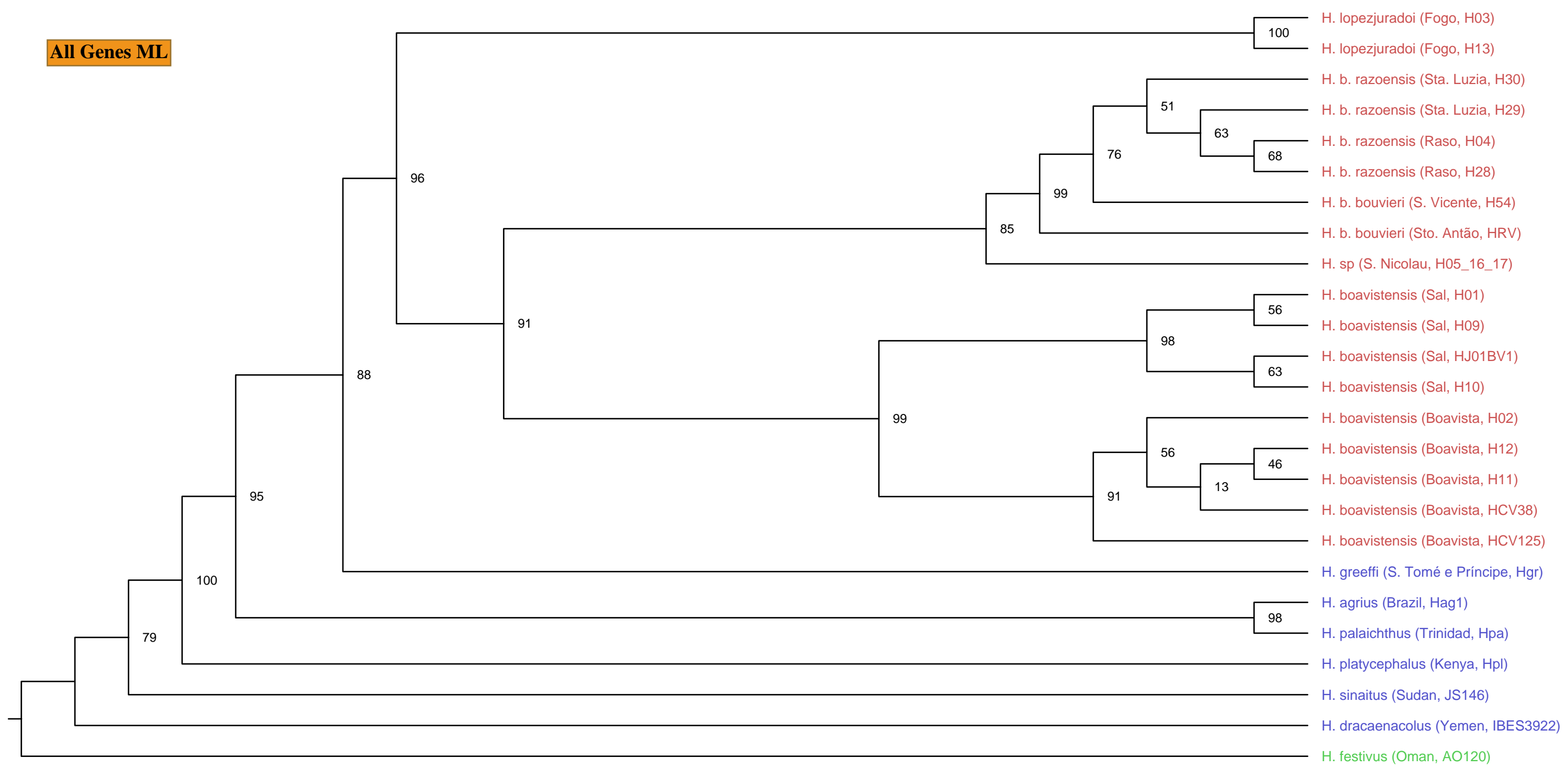
0.06

All Genes BI

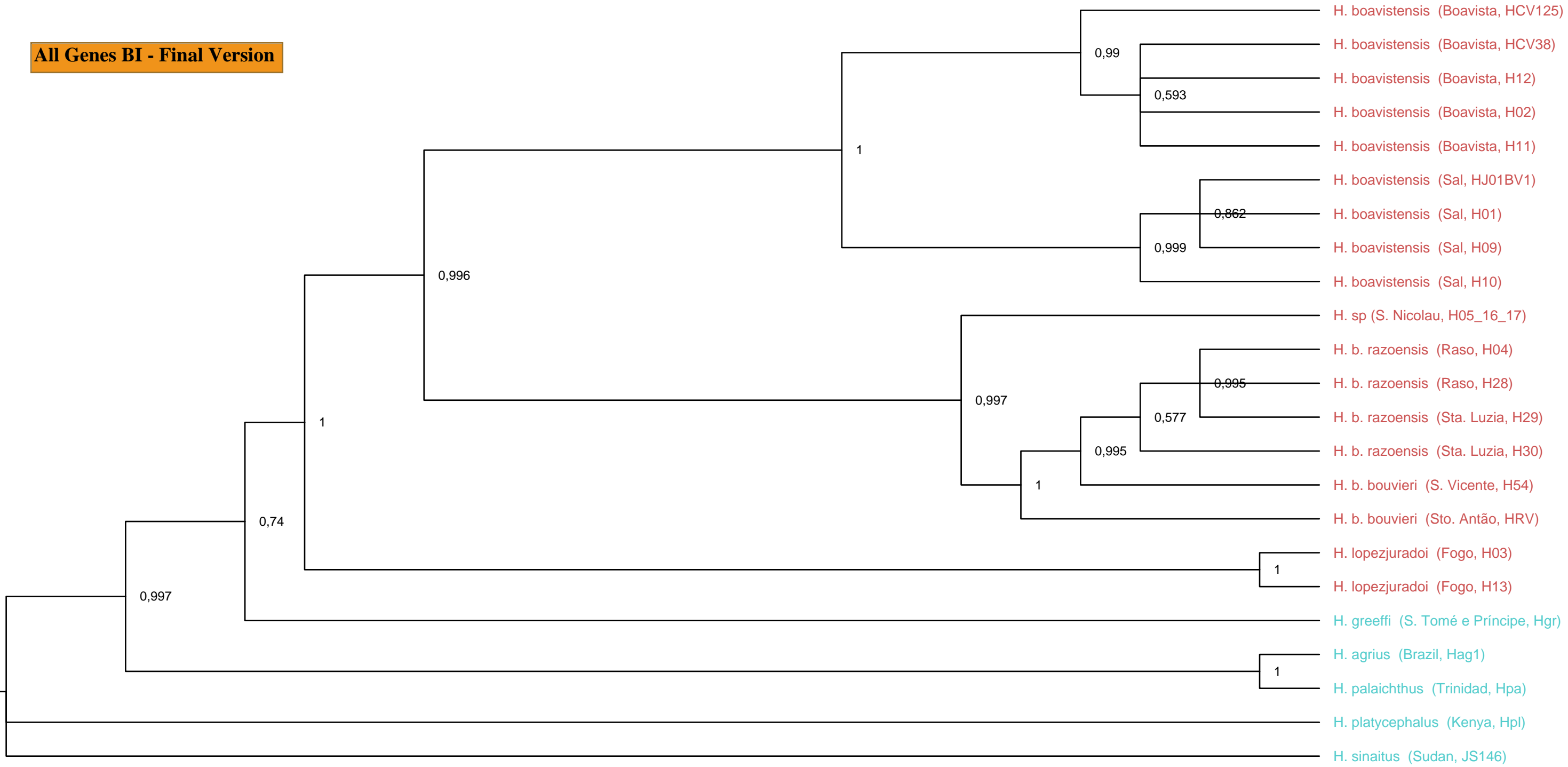


-
0.05

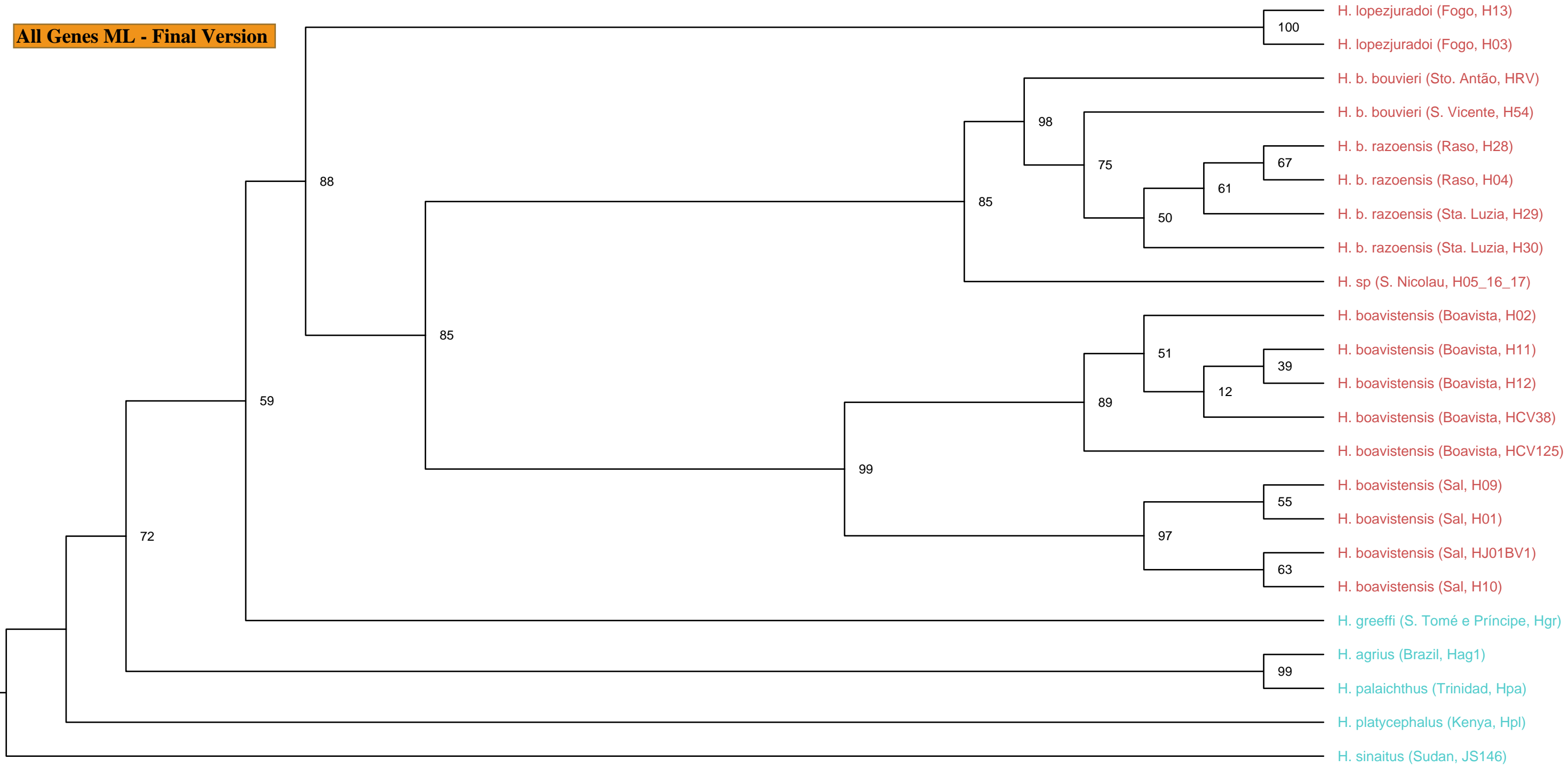
All Genes ML



All Genes BI - Final Version

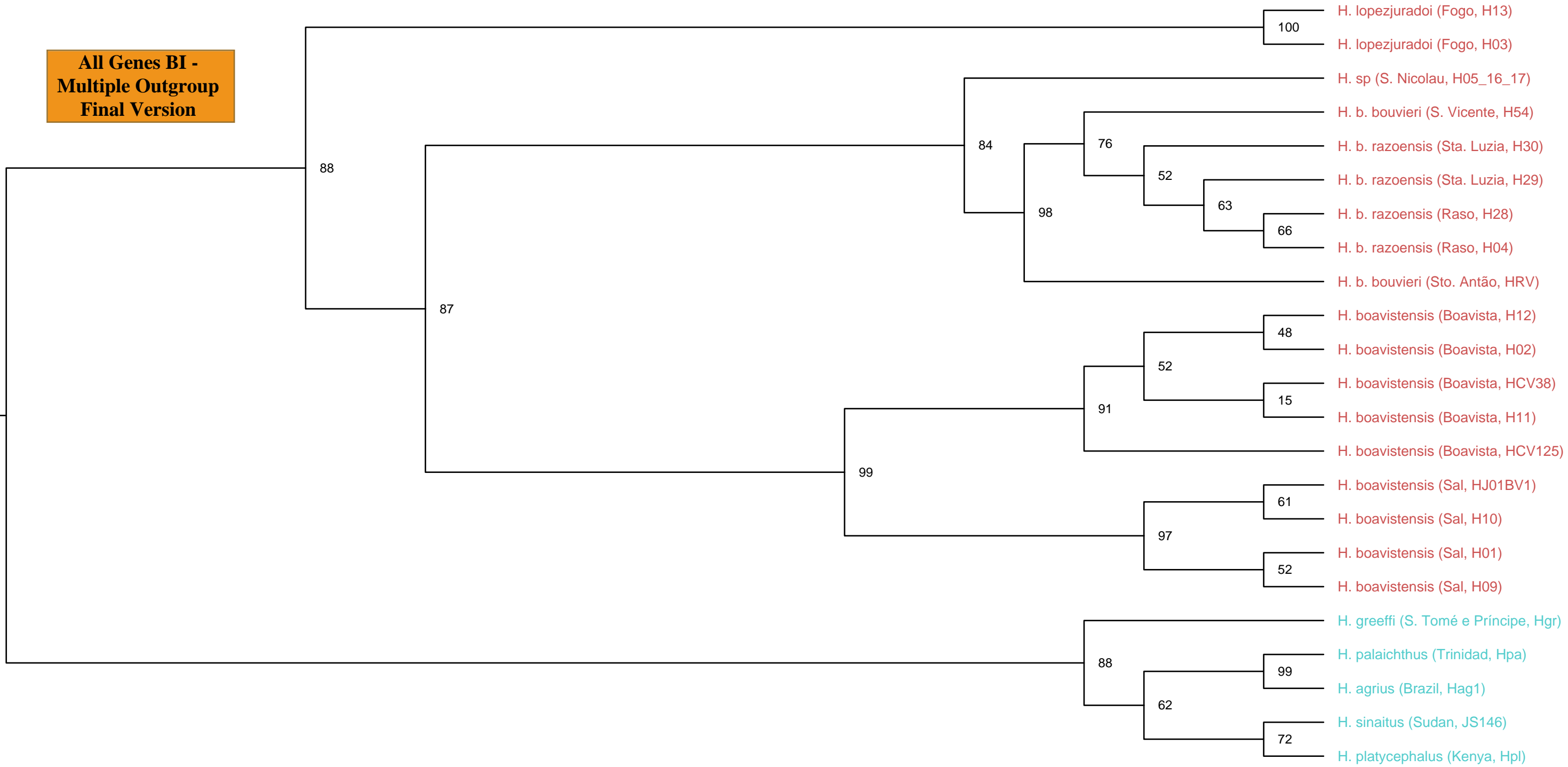


All Genes ML - Final Version

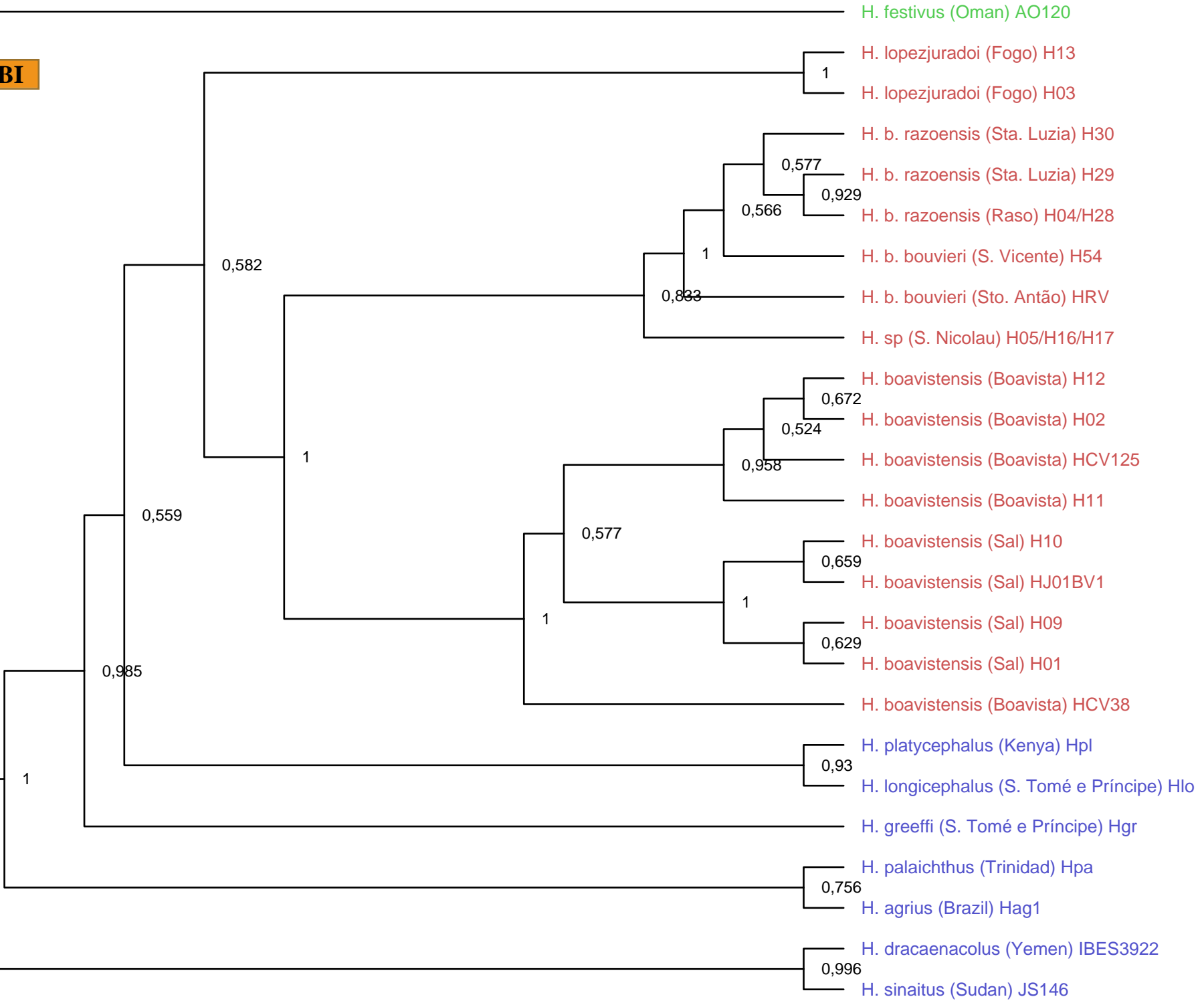


3.0

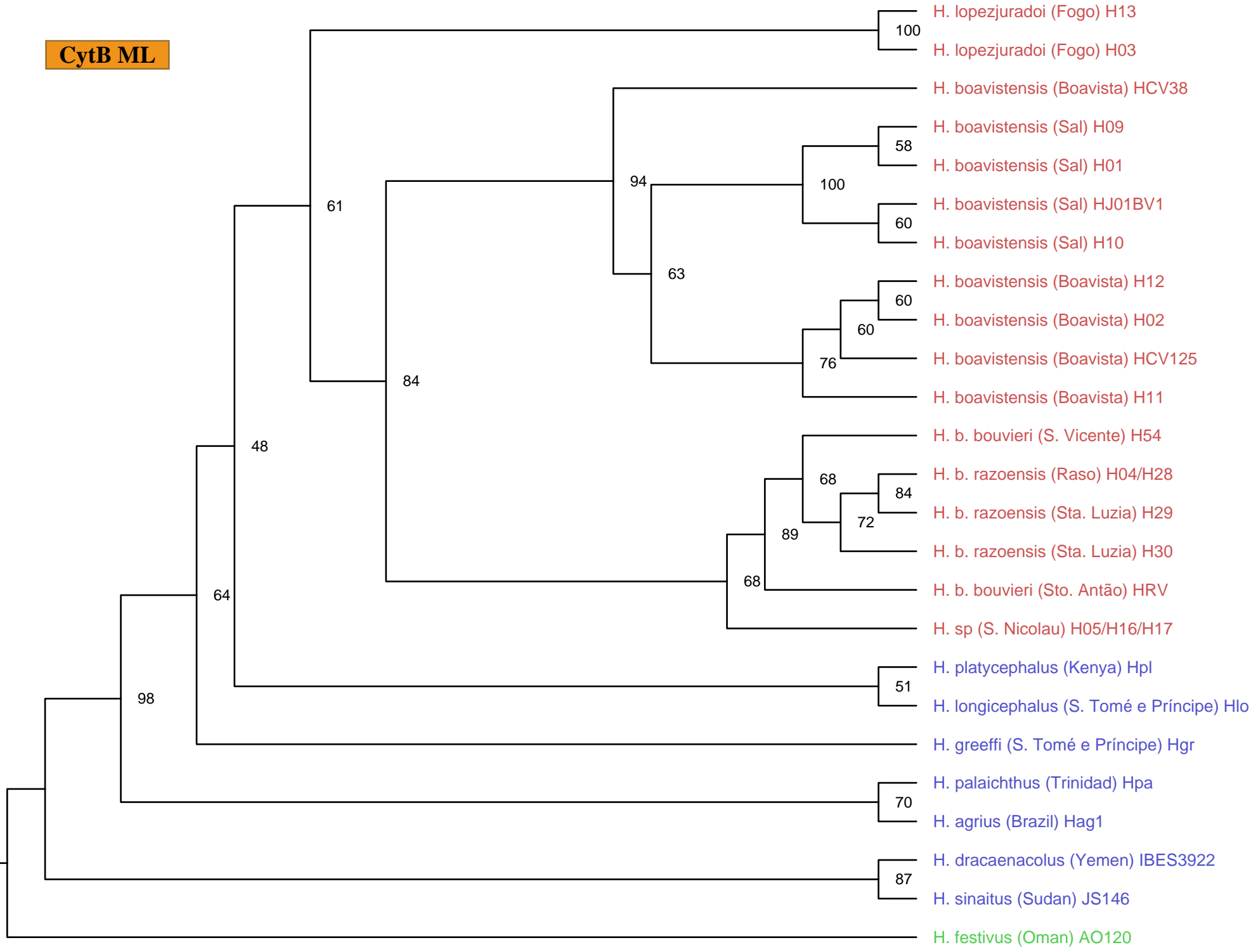
**All Genes BI -
Multiple Outgroup
Final Version**



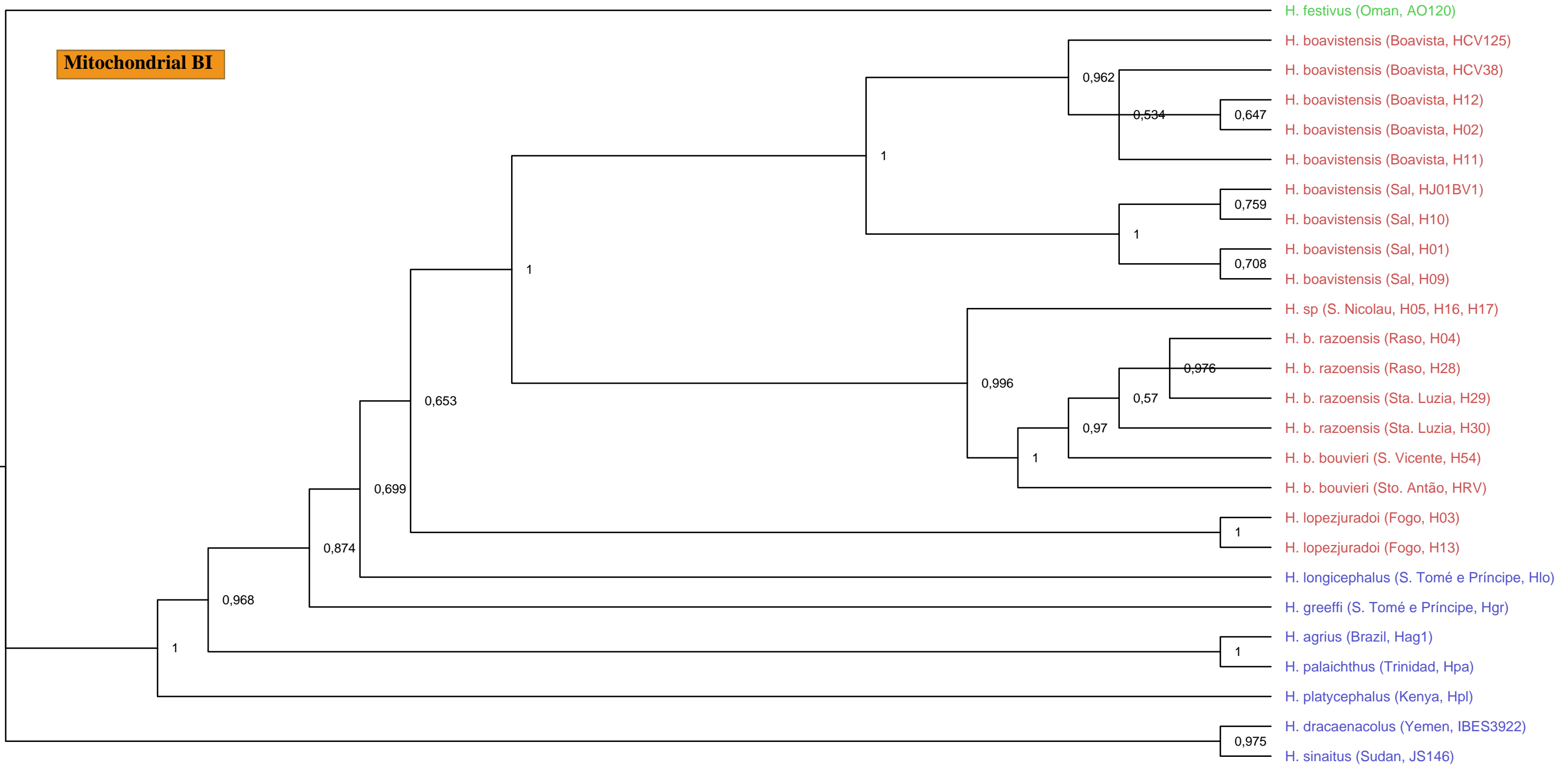
CytB BI



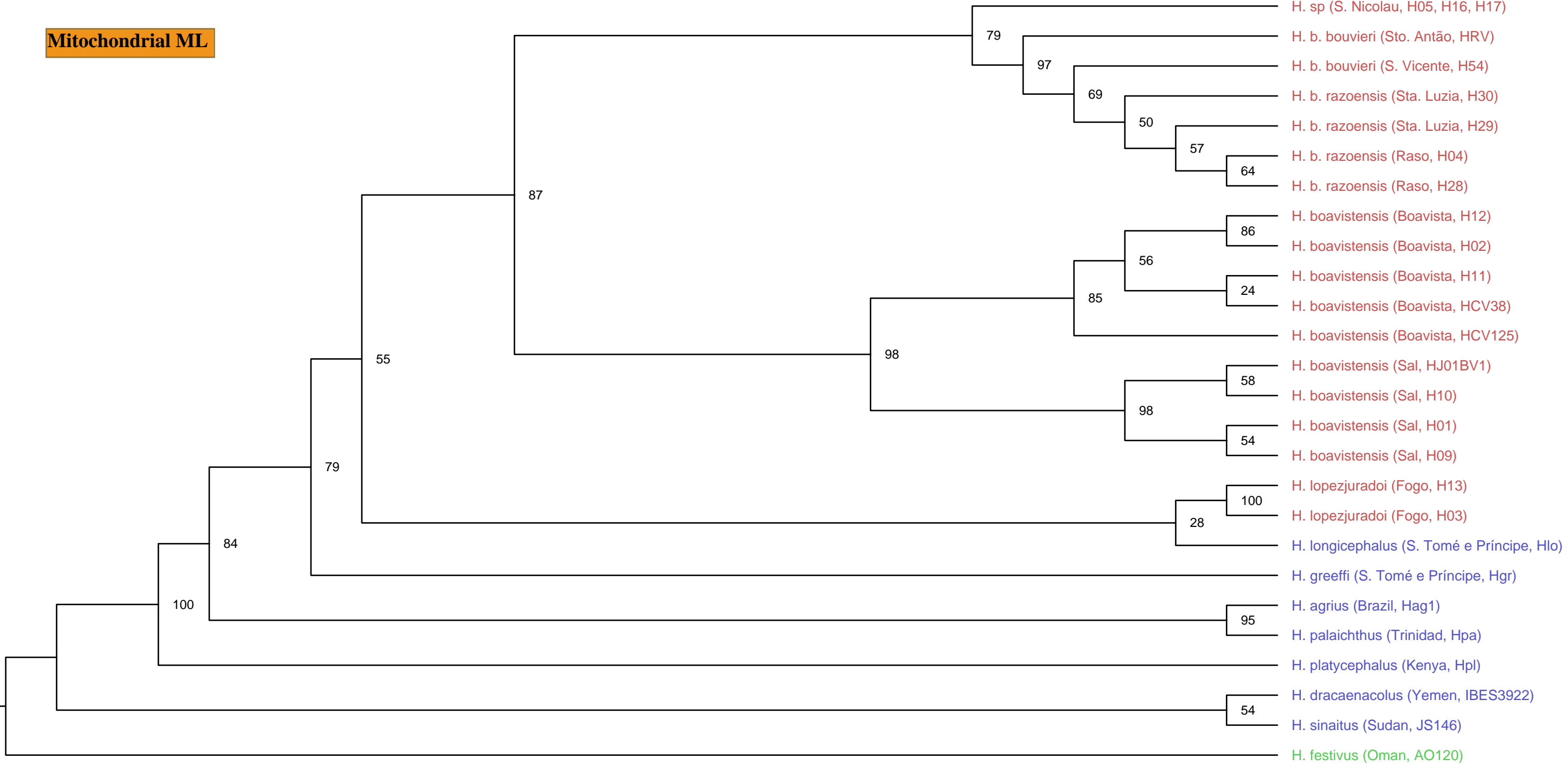
CytB ML



Mitochondrial BI

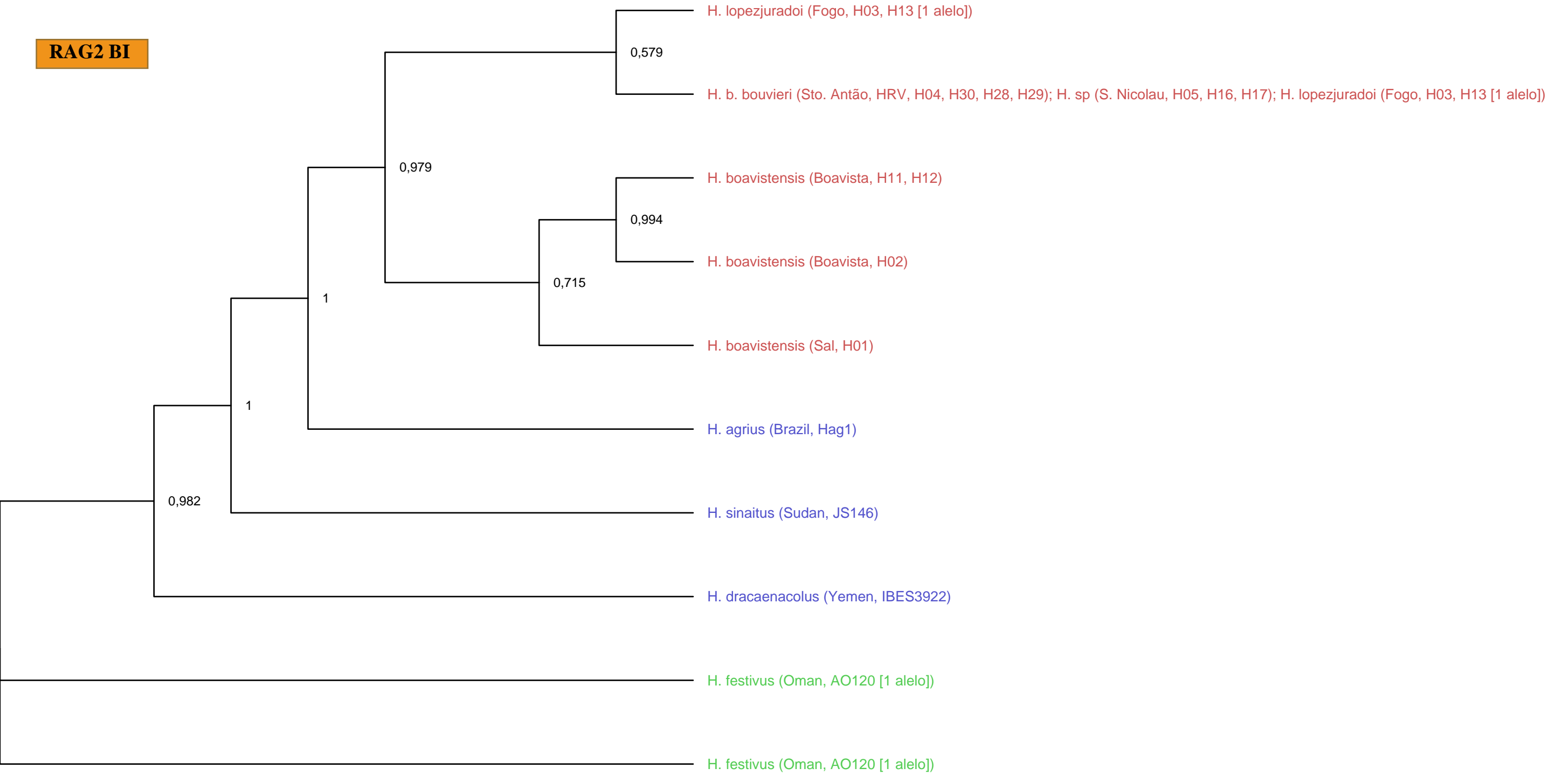


Mitochondrial ML



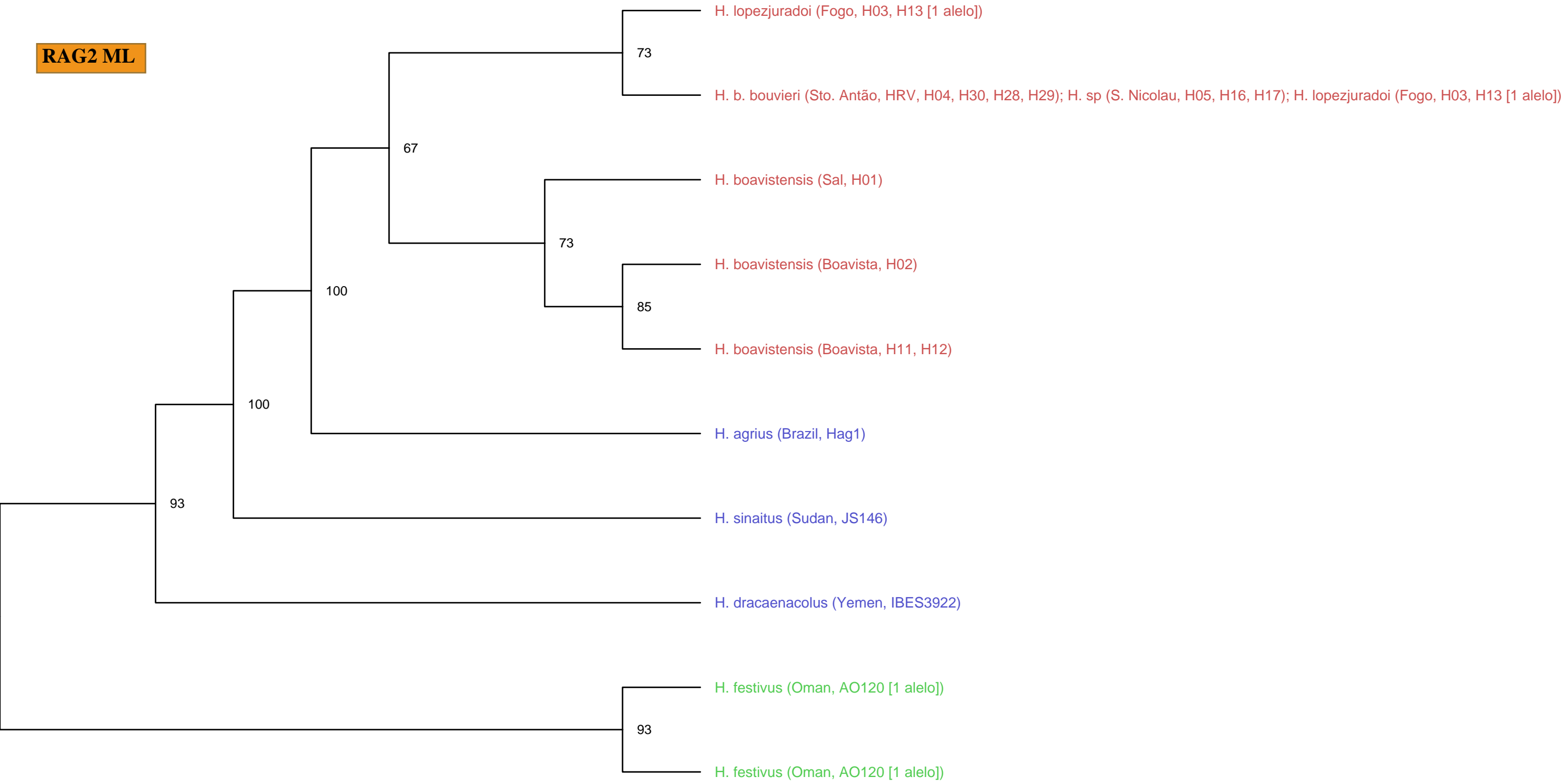
-
0.09

RAG2 BI



0.003

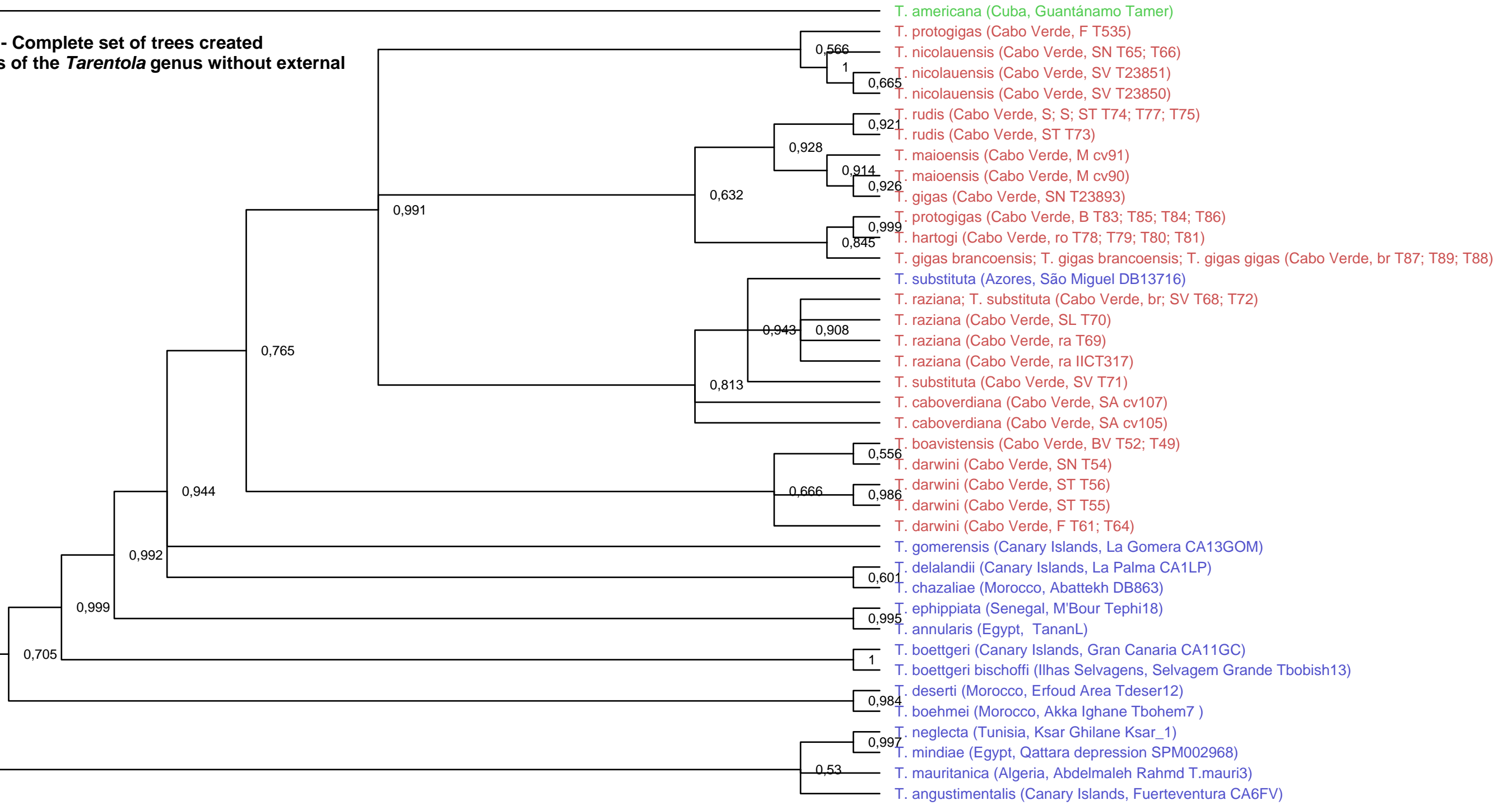
RAG2 ML



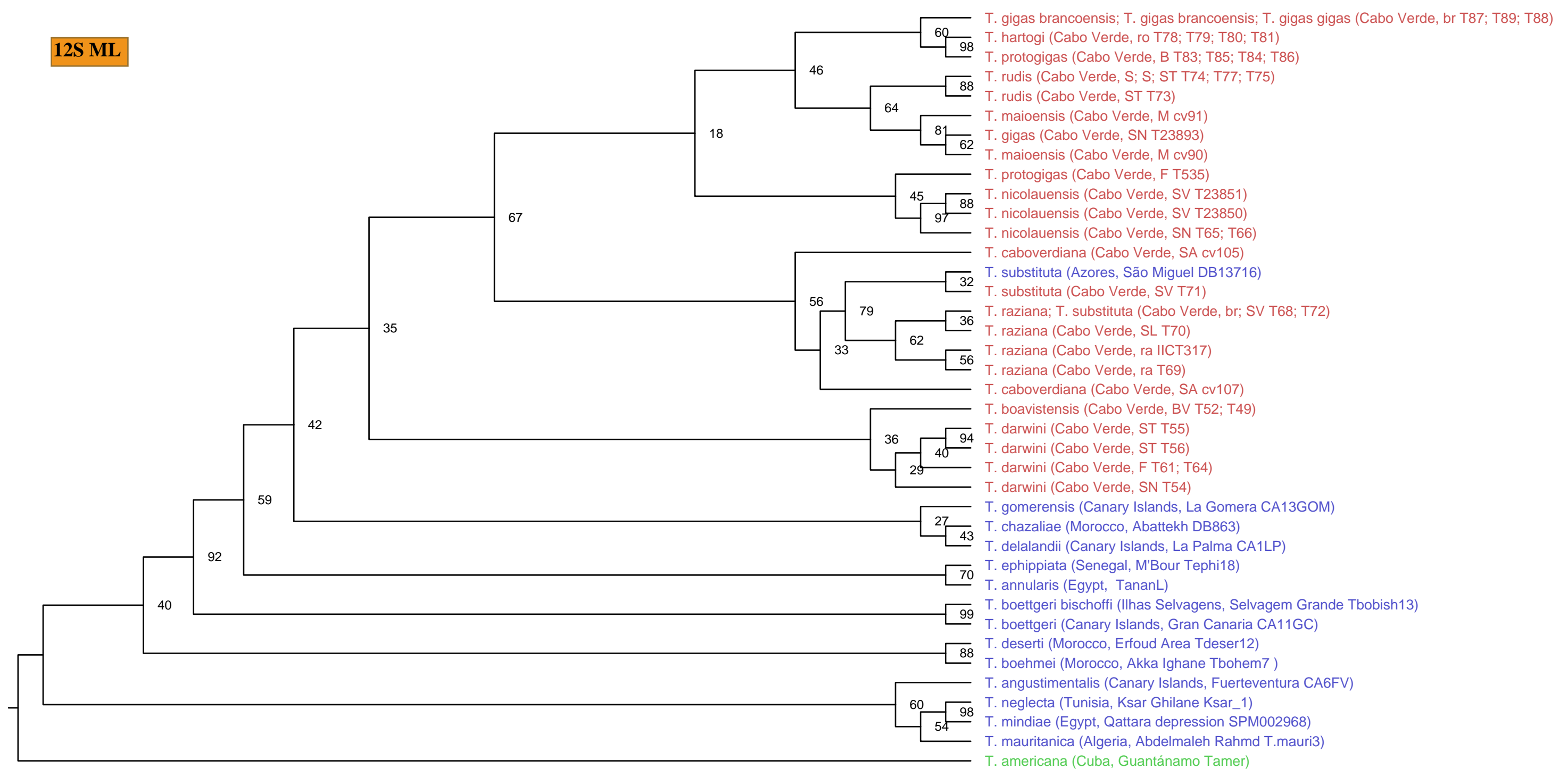
0.004

Appendix VIII - Complete set of trees created in the analysis of the *Tarentola* genus without external outgroups.

12S BI



12S ML



ACM4 BI

T. chazaliae (Morocco, Abattekh DB863)

Multiple CV Species (all except T. boavistensis) (Cabo Verde, Multiple Islands)

T. boavistensis (Cabo Verde, BV T661;T665)

T. gigas (Cabo Verde, JB45)

T. fogoensis (Cabo Verde, F T533;T537) [1 alelo]

1

T. raziana (Cabo Verde, SL T144)

T. bocagei (Cabo Verde, SN T349) [1 alelo]

T. bocagei (Cabo Verde, SN T304) [1 alelo]

T. bocagei (Cabo Verde, SN T302) [1 alelo]

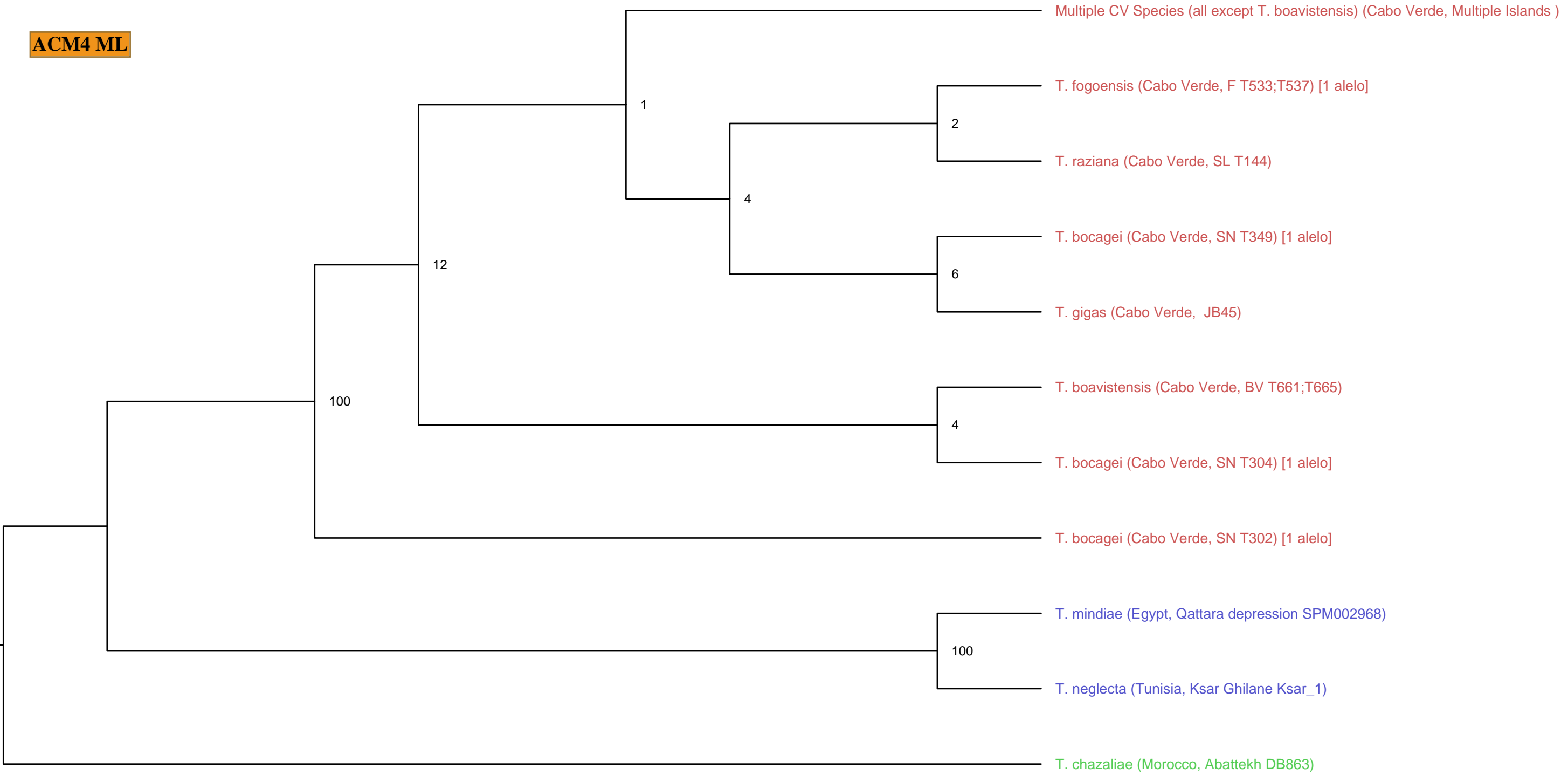
T. neglecta (Tunisia, Ksar Ghilane Ksar_1)

0,999

T. mindiae (Egypt, Qattara depression SPM002968)

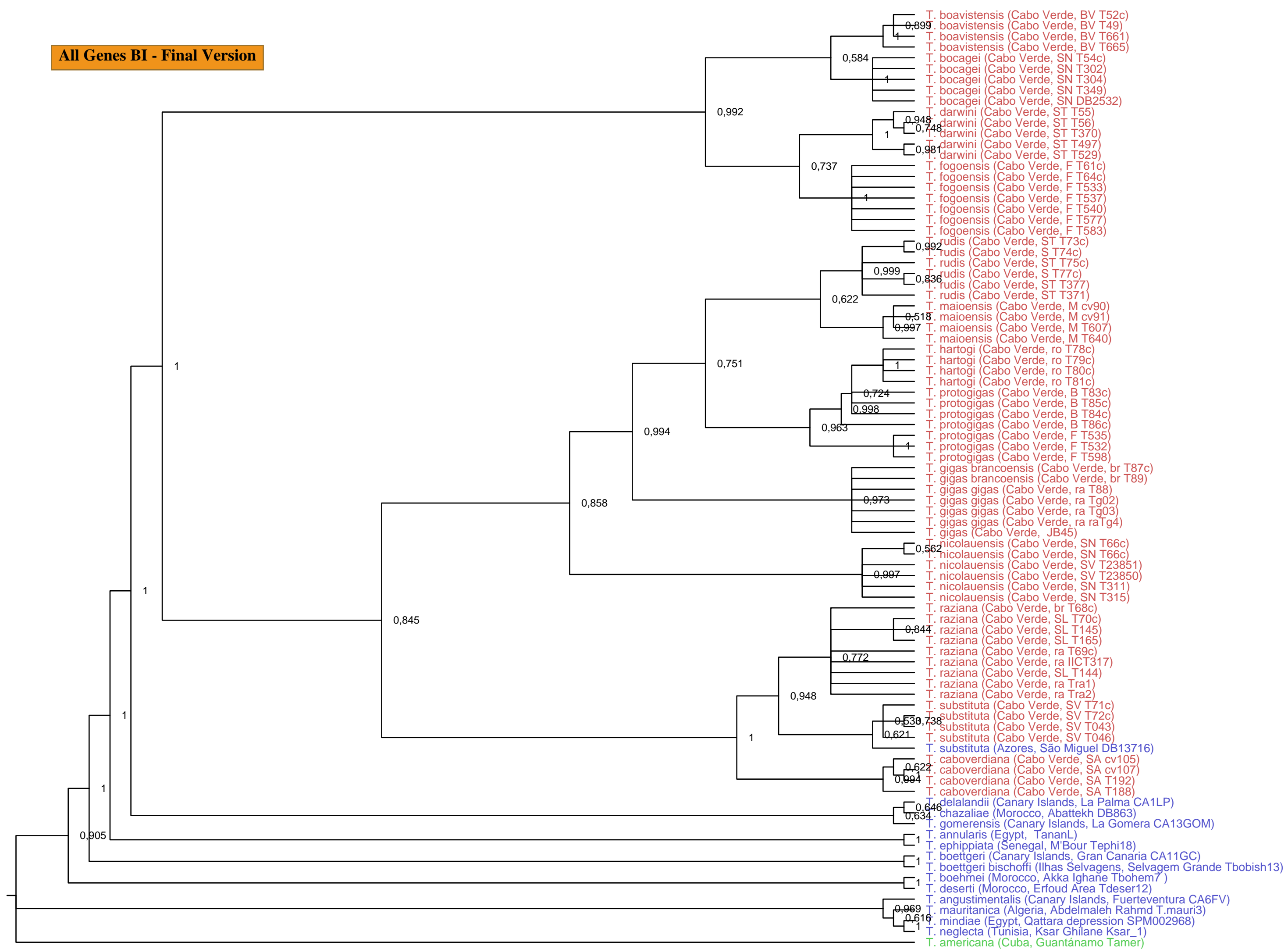
6.0E-4

ACM4 ML



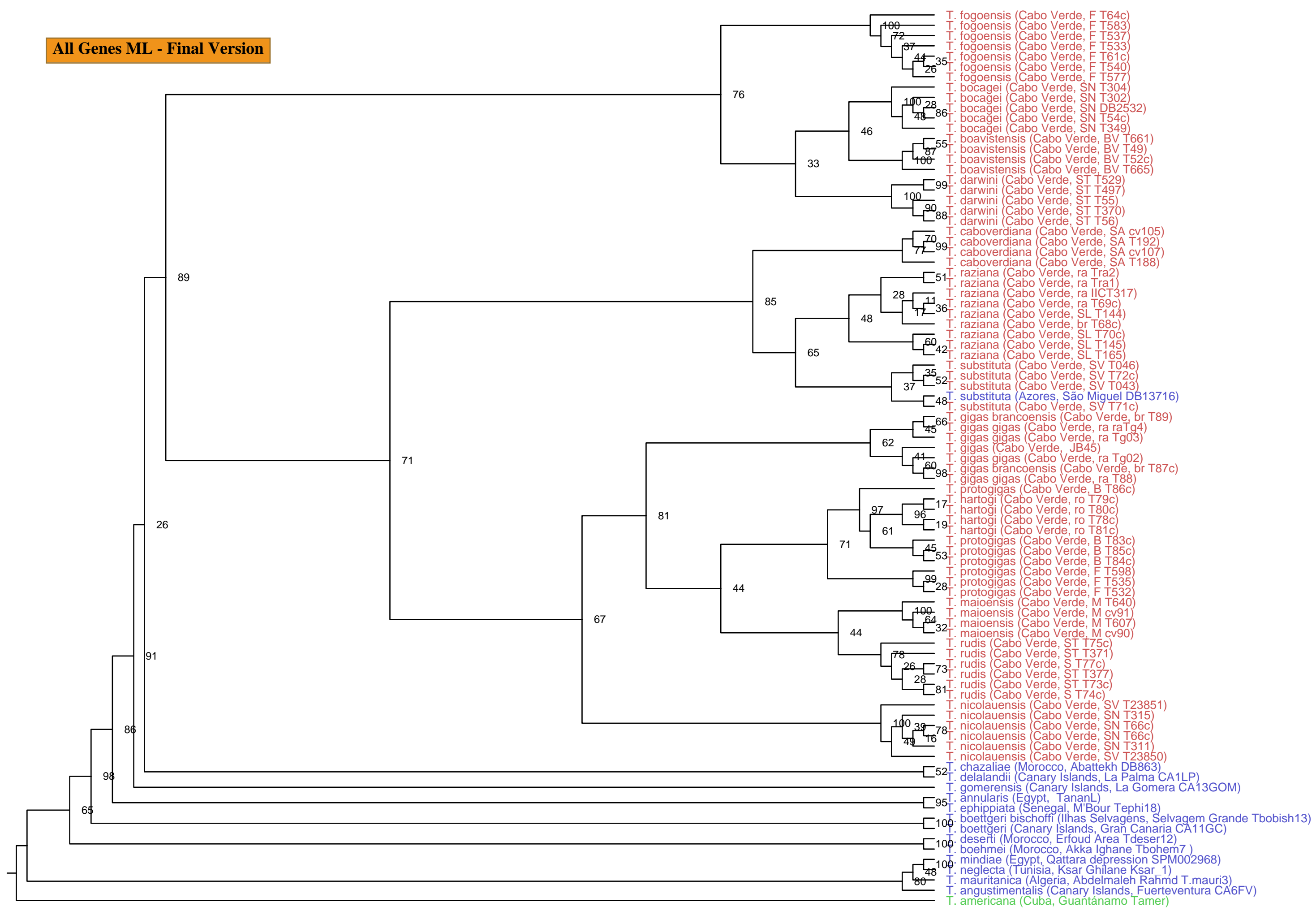
0.002

All Genes BI - Final Version

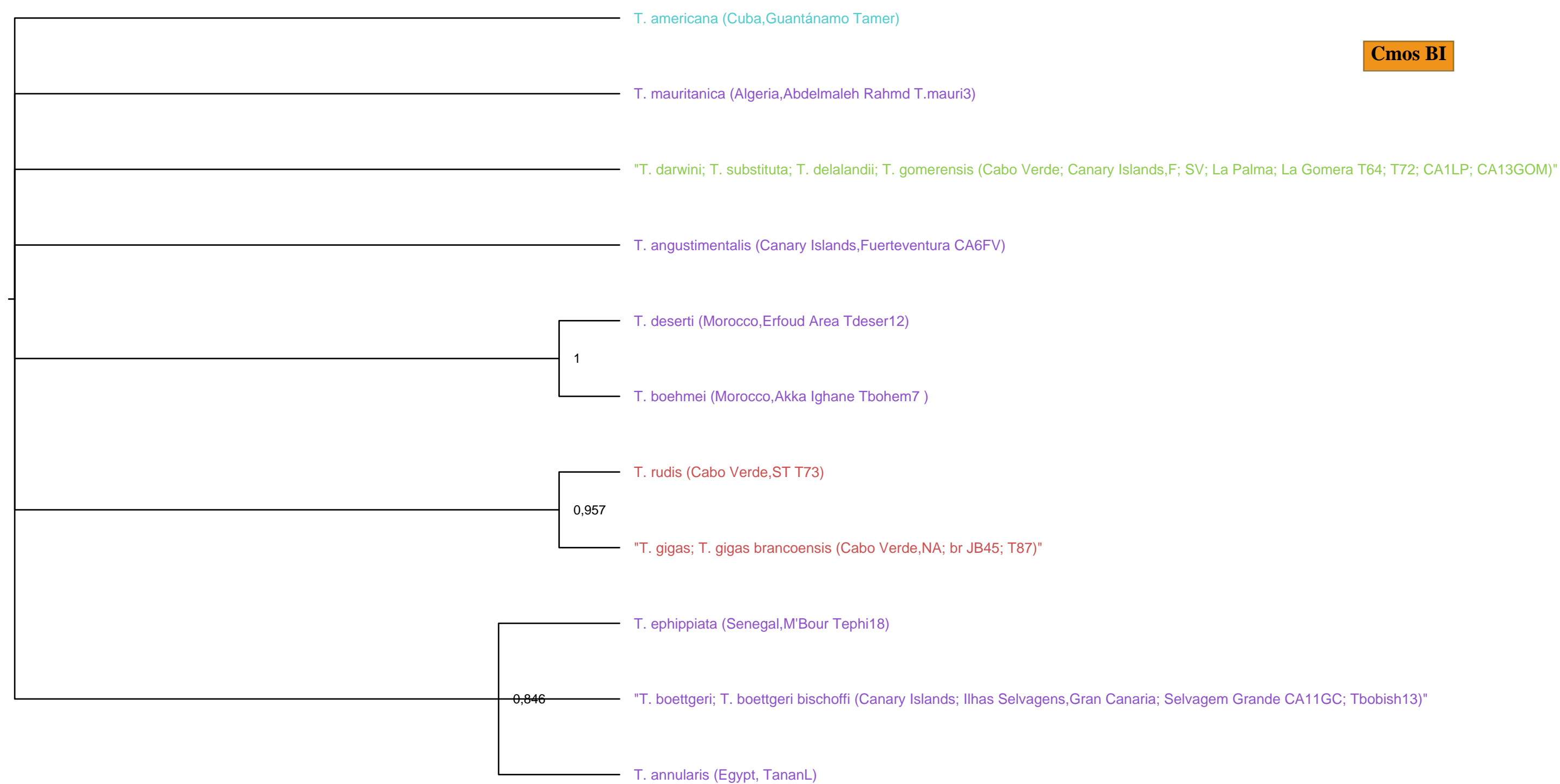


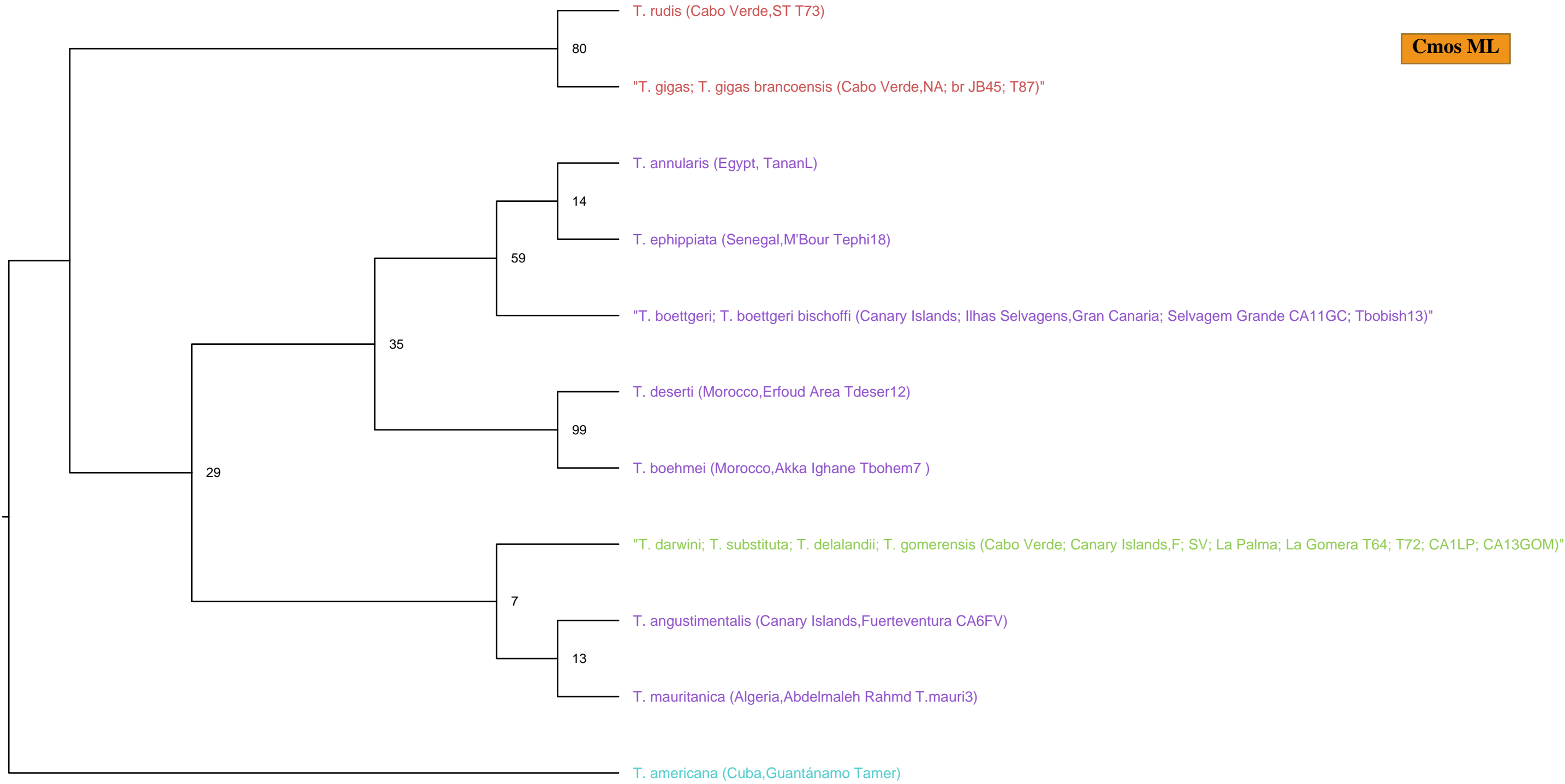
9.0

All Genes ML - Final Version



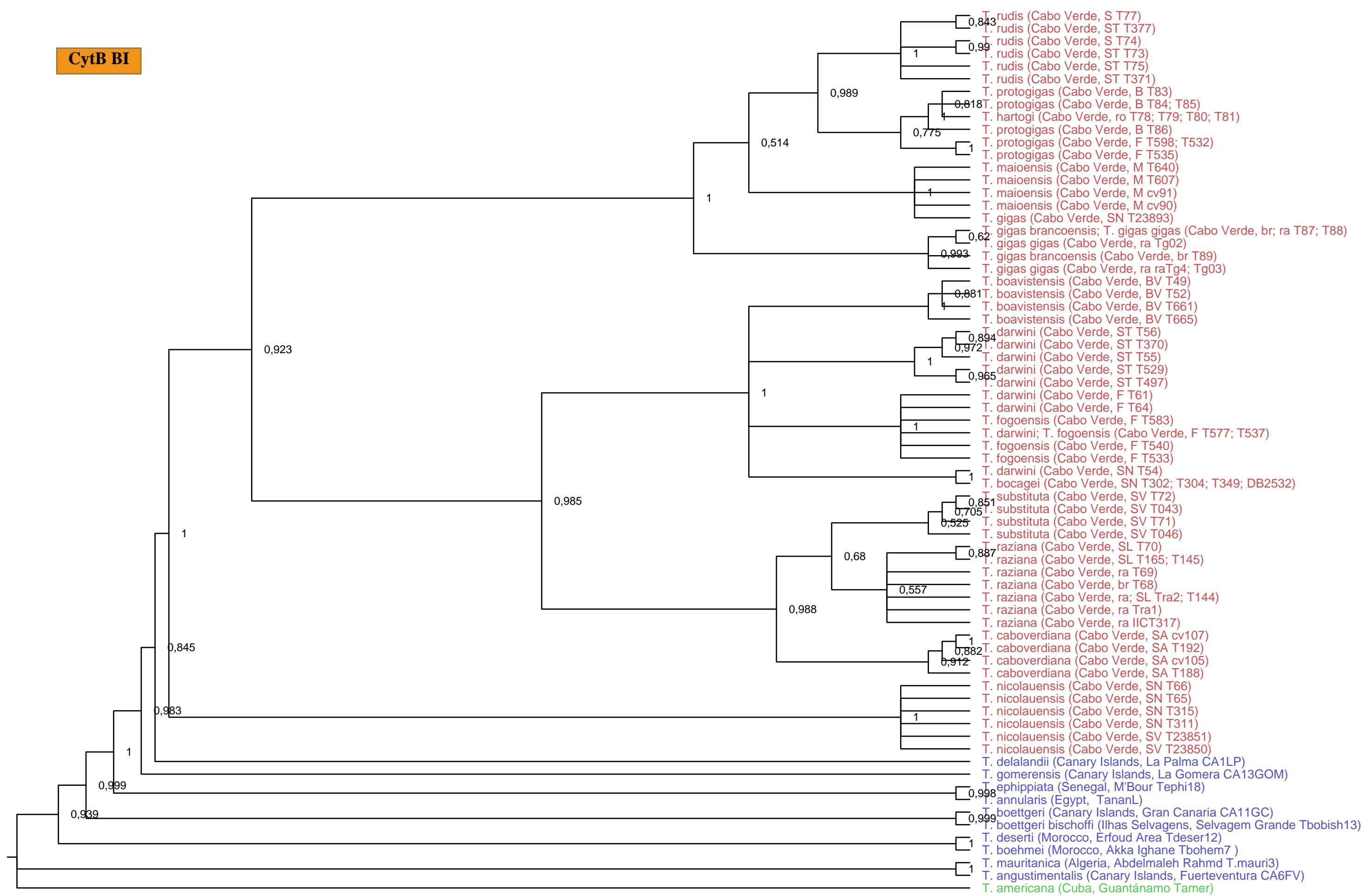
9.0





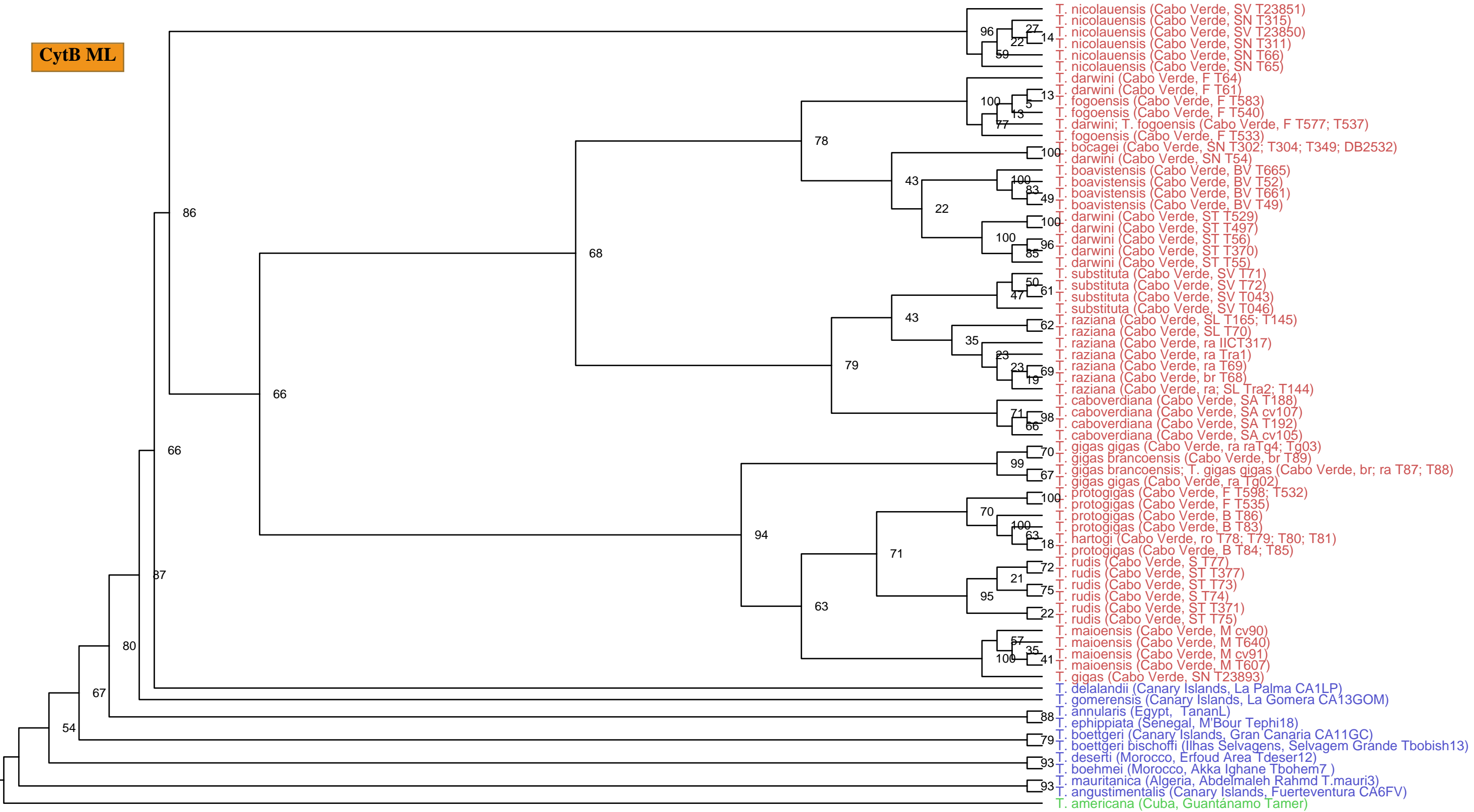
0.003

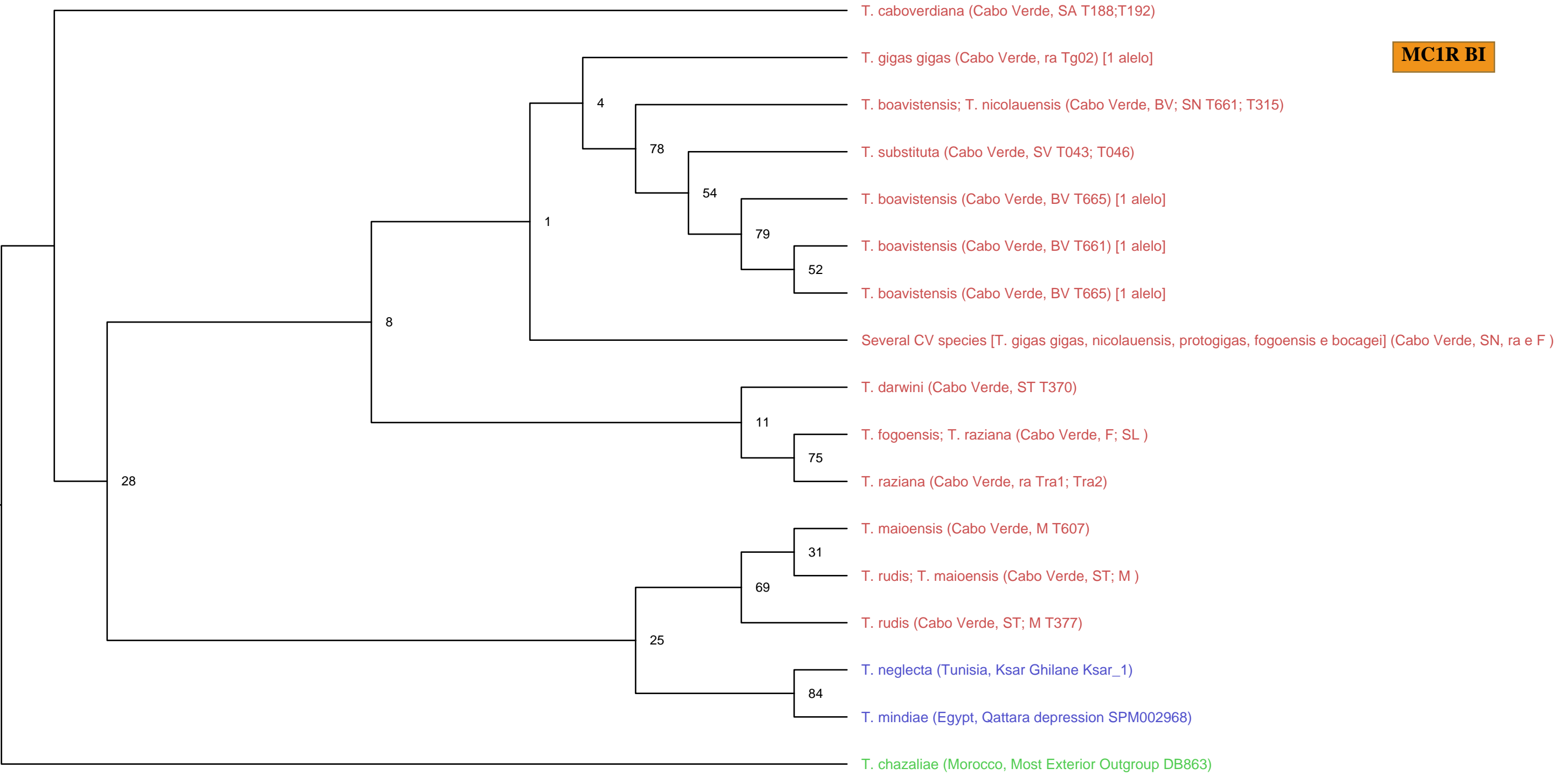
CytB BI



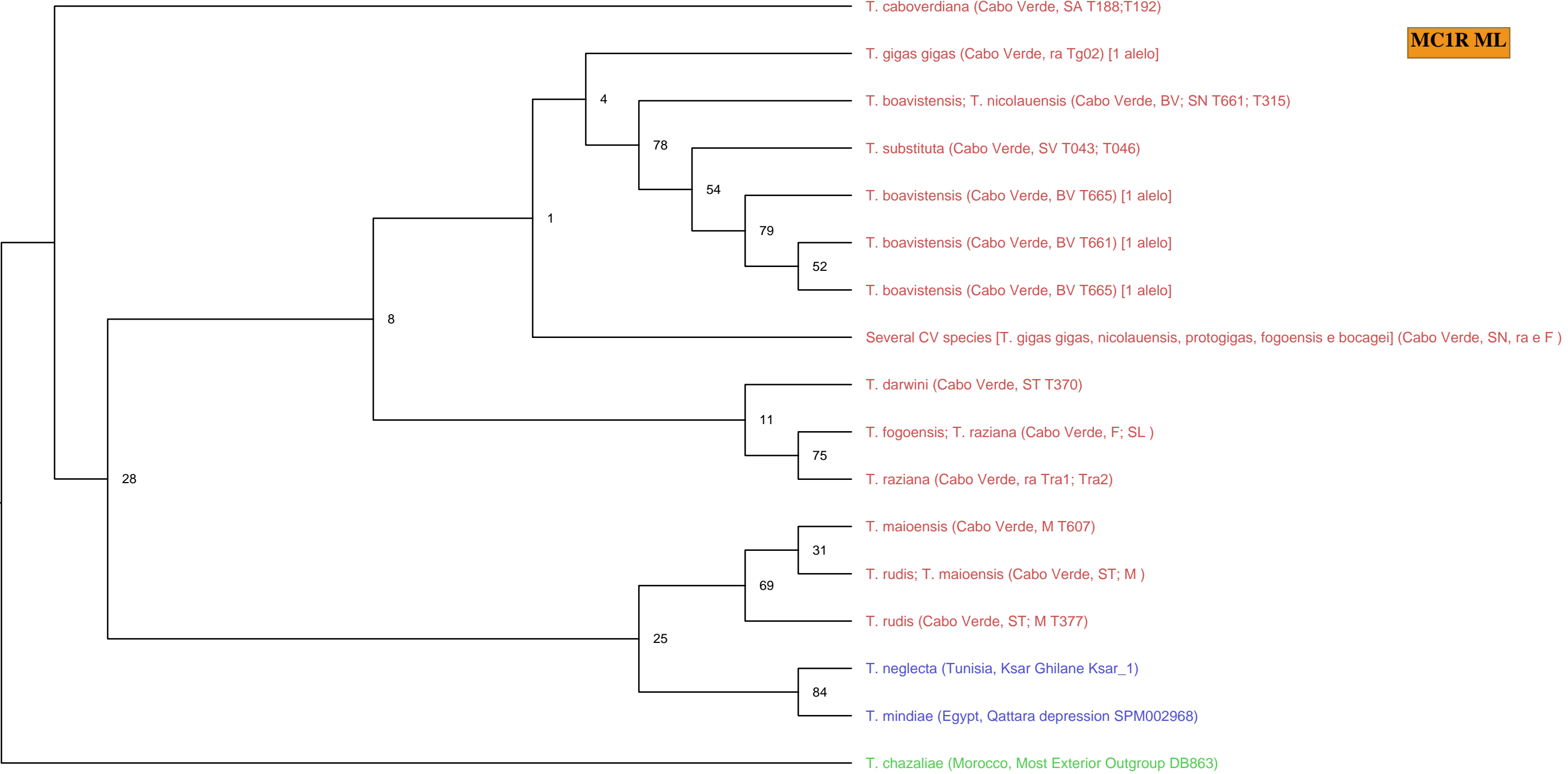
7.0

CytB ML

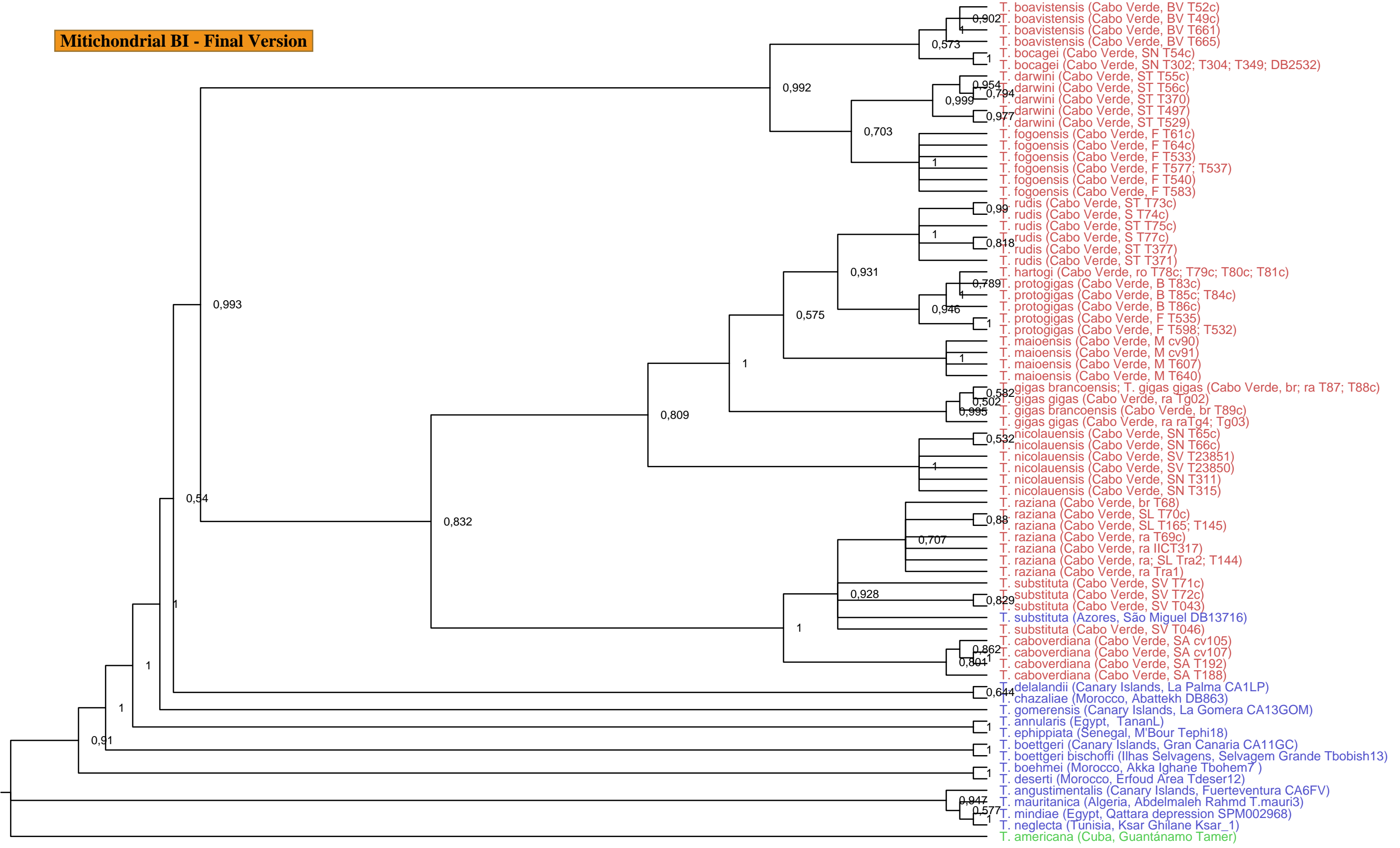




0.003

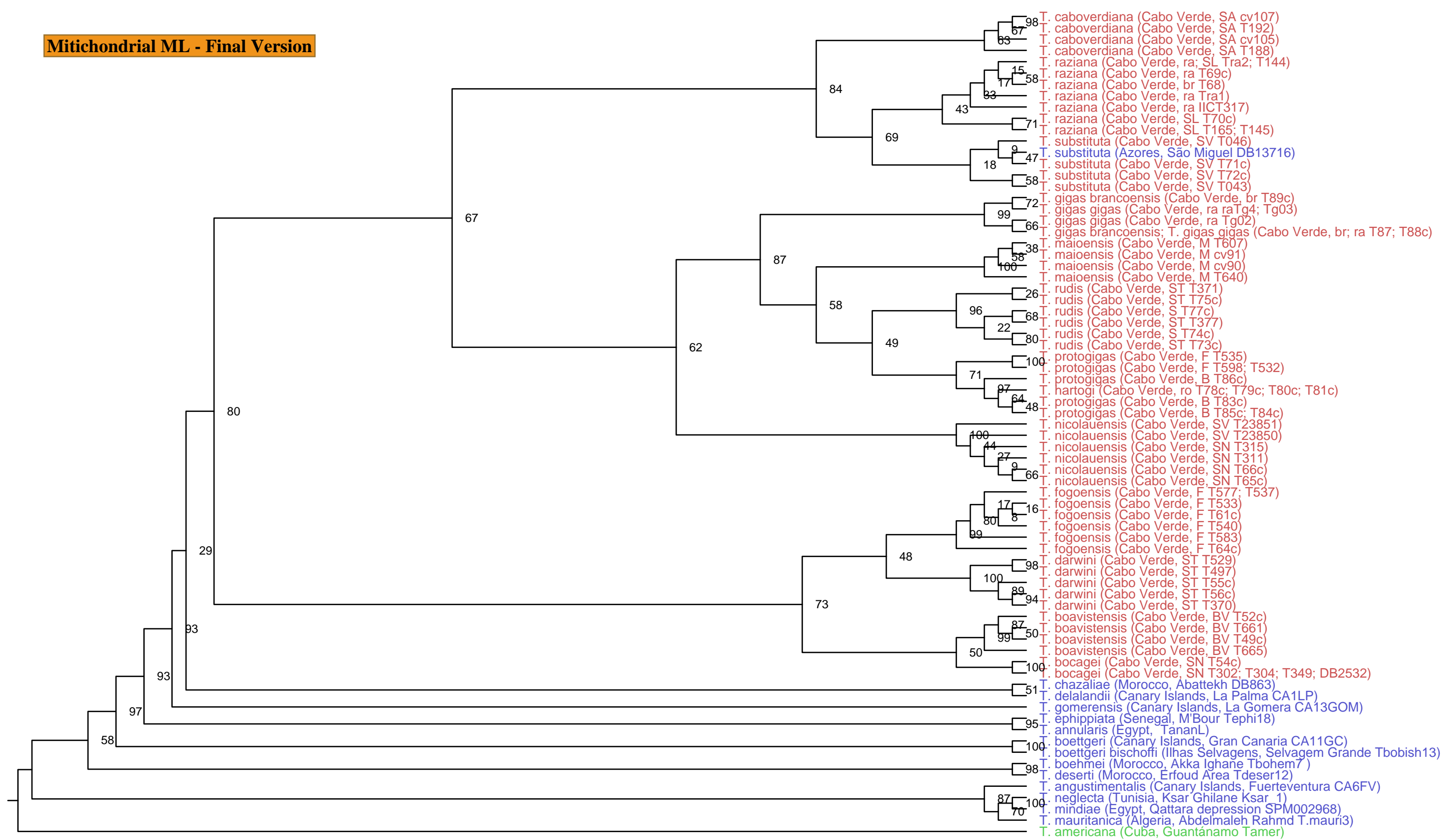


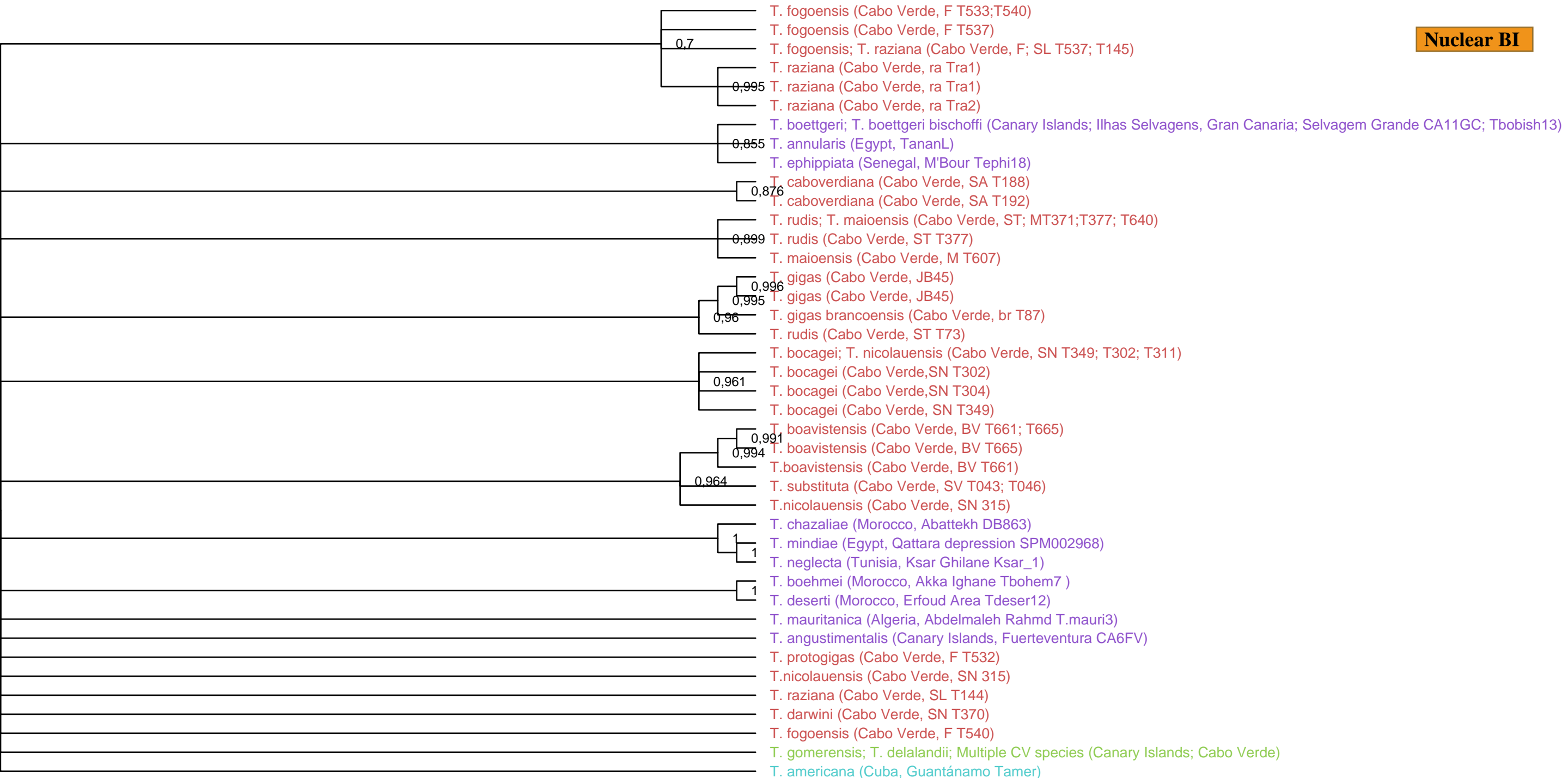
Mitichondrial BI - Final Version



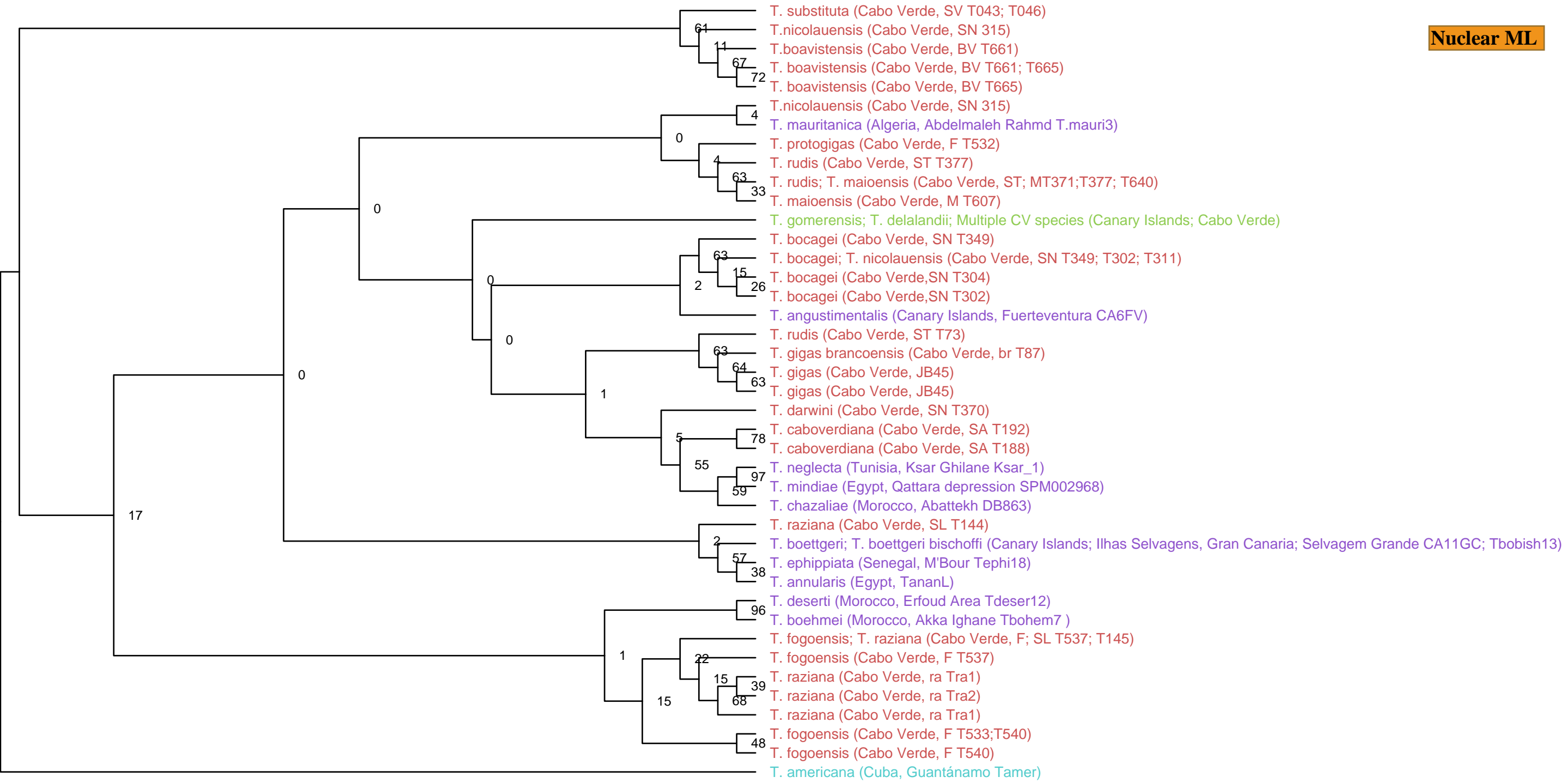
8.0

Mitichondrial ML - Final Version



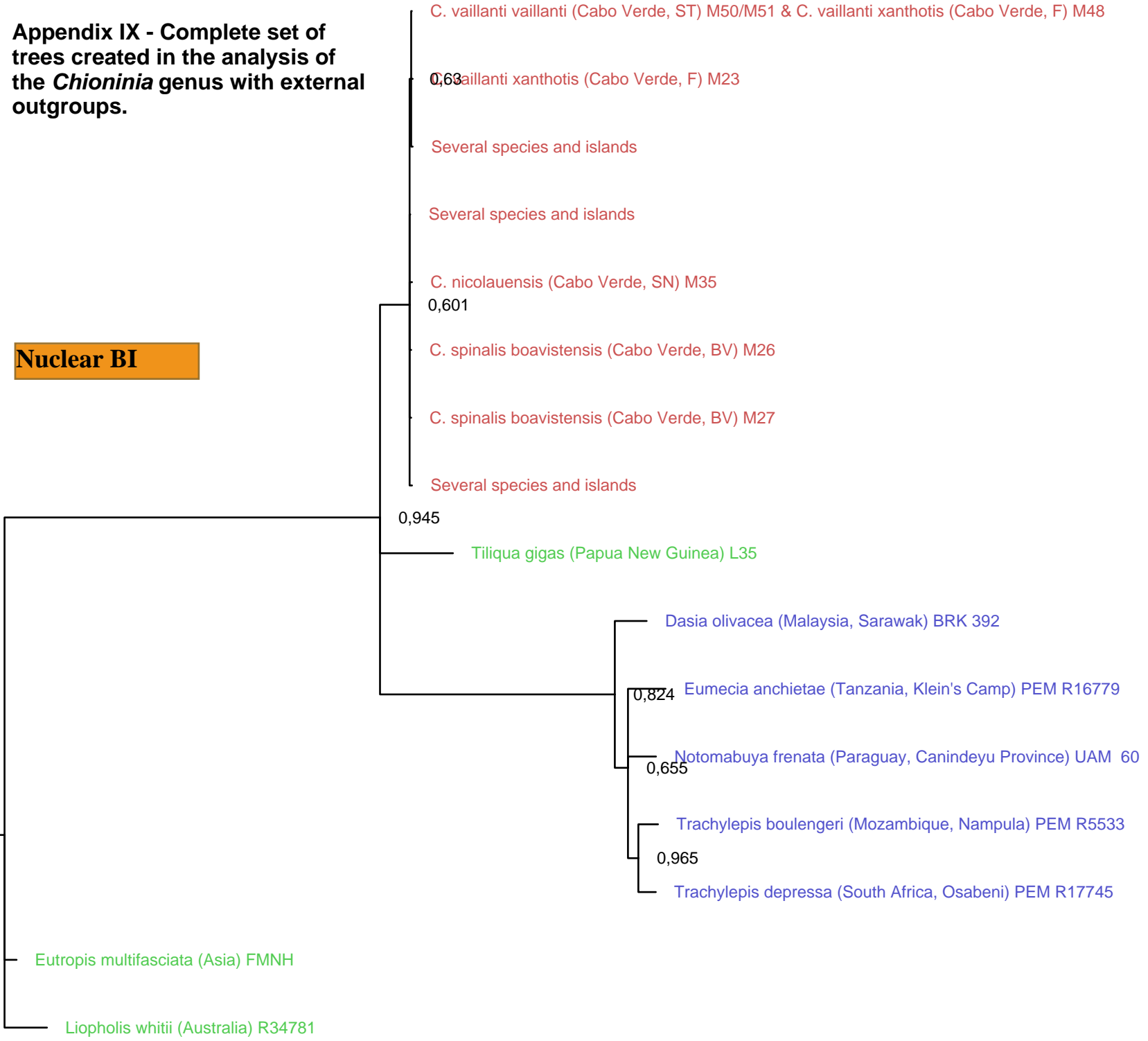


5.0

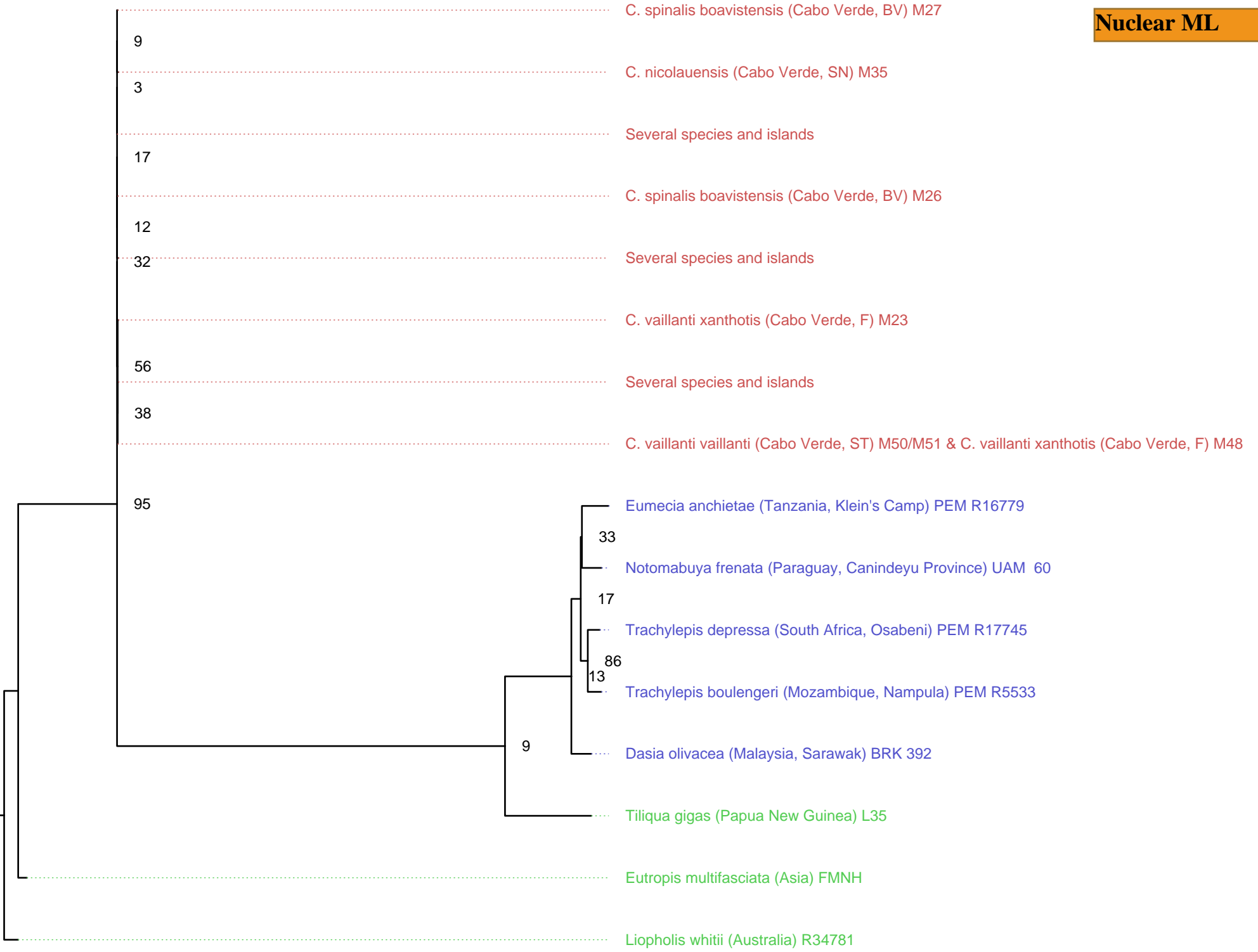


Appendix IX - Complete set of trees created in the analysis of the *Chioninia* genus with external outgroups.

Nuclear BI

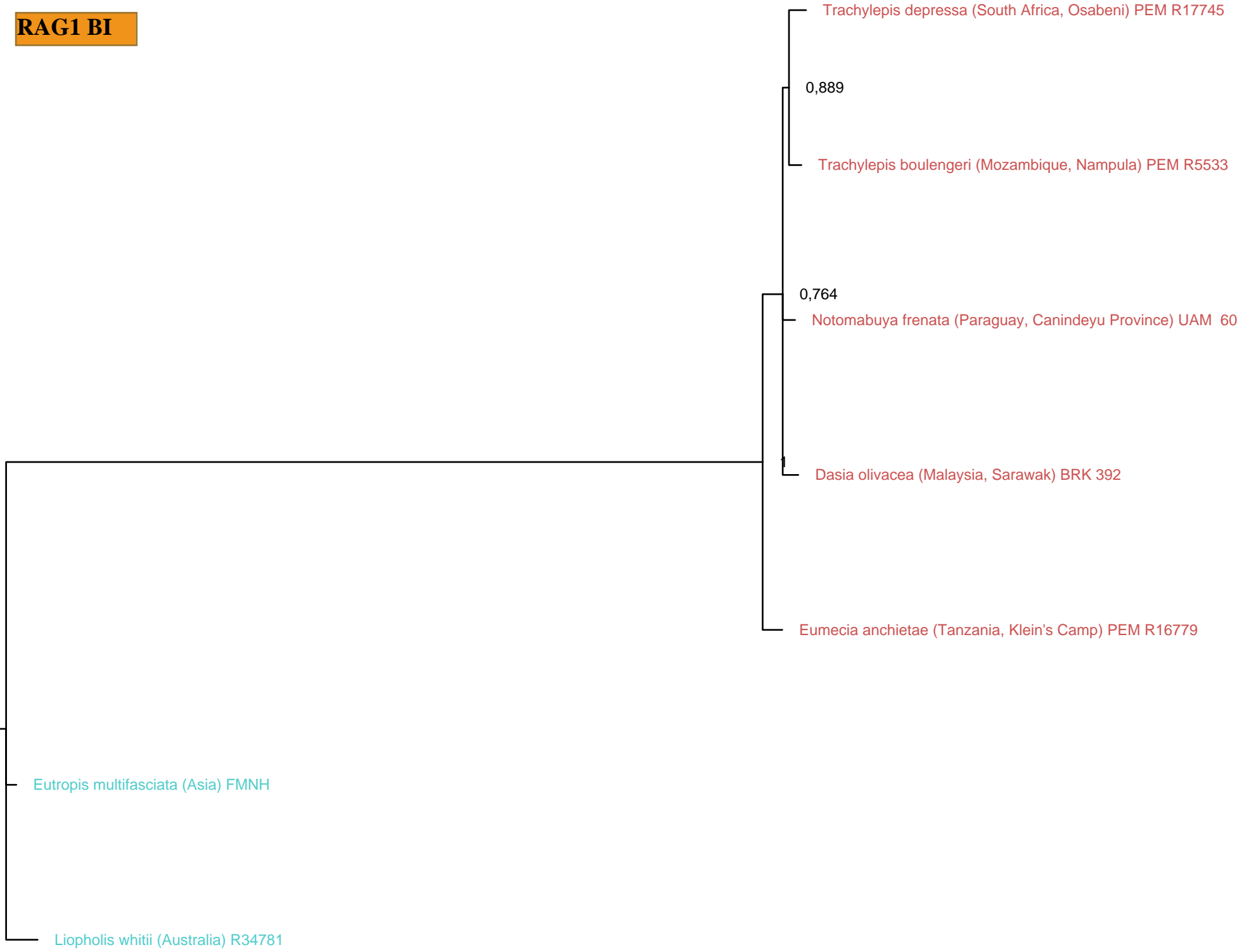


0.09



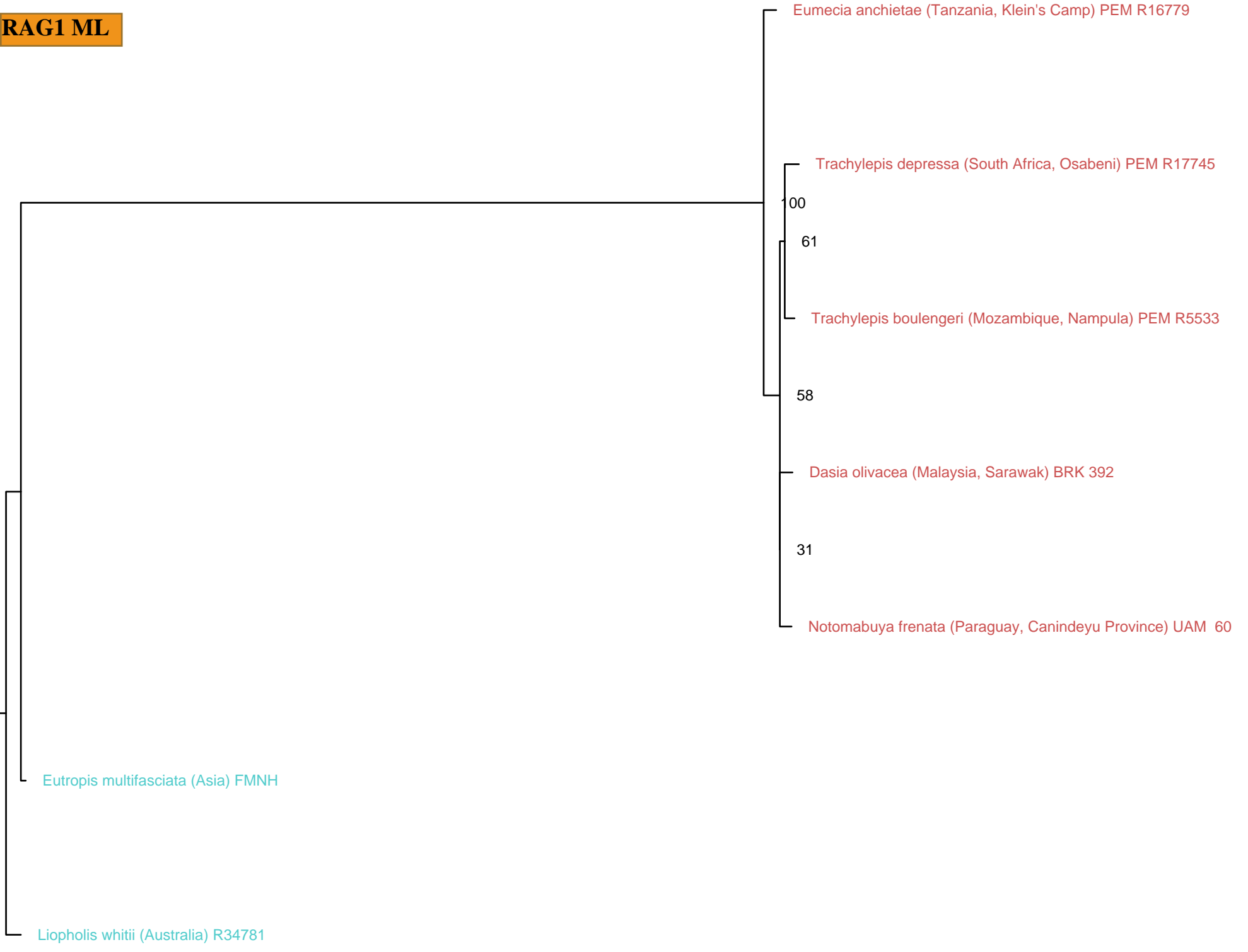
0.2

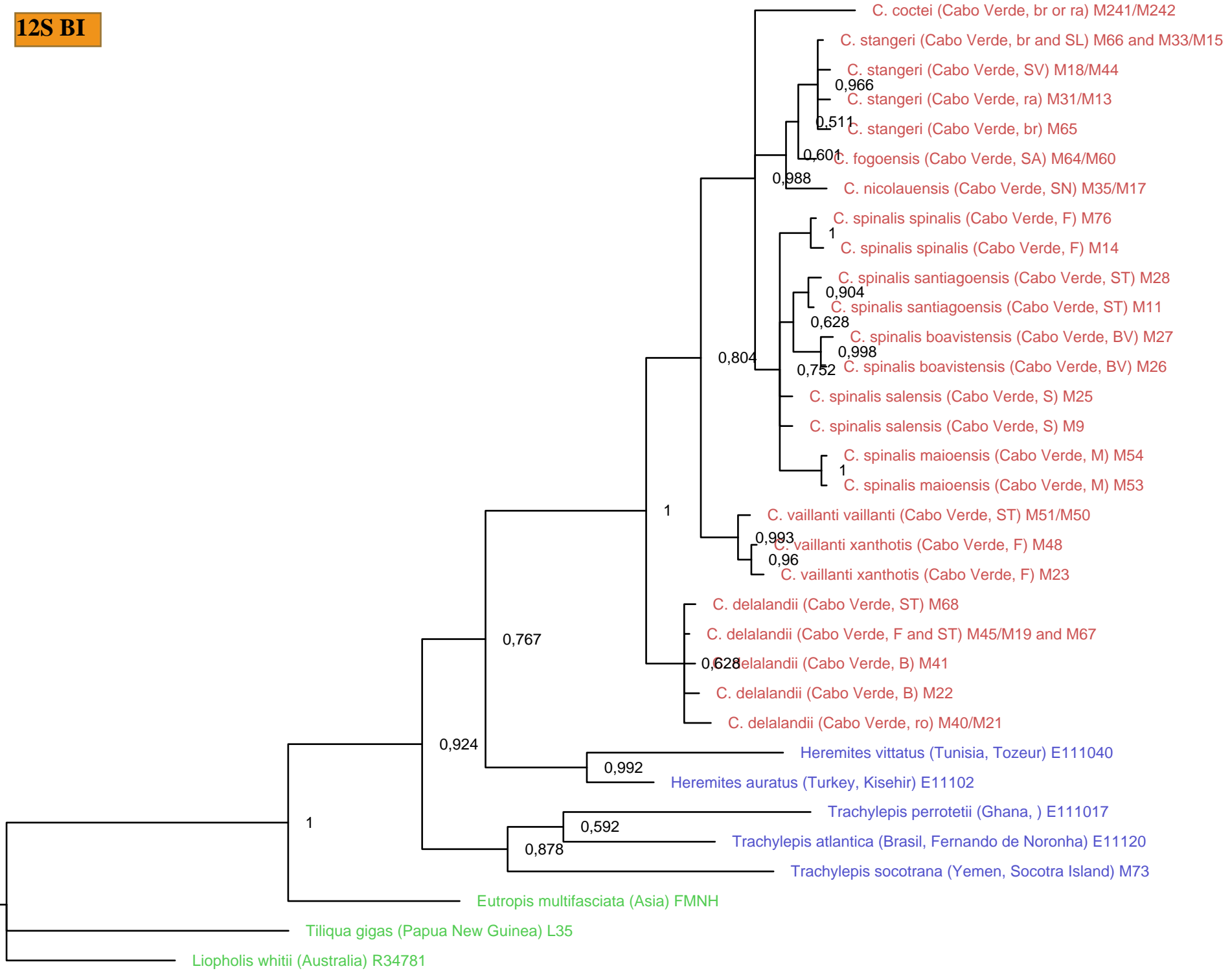
RAG1 BI



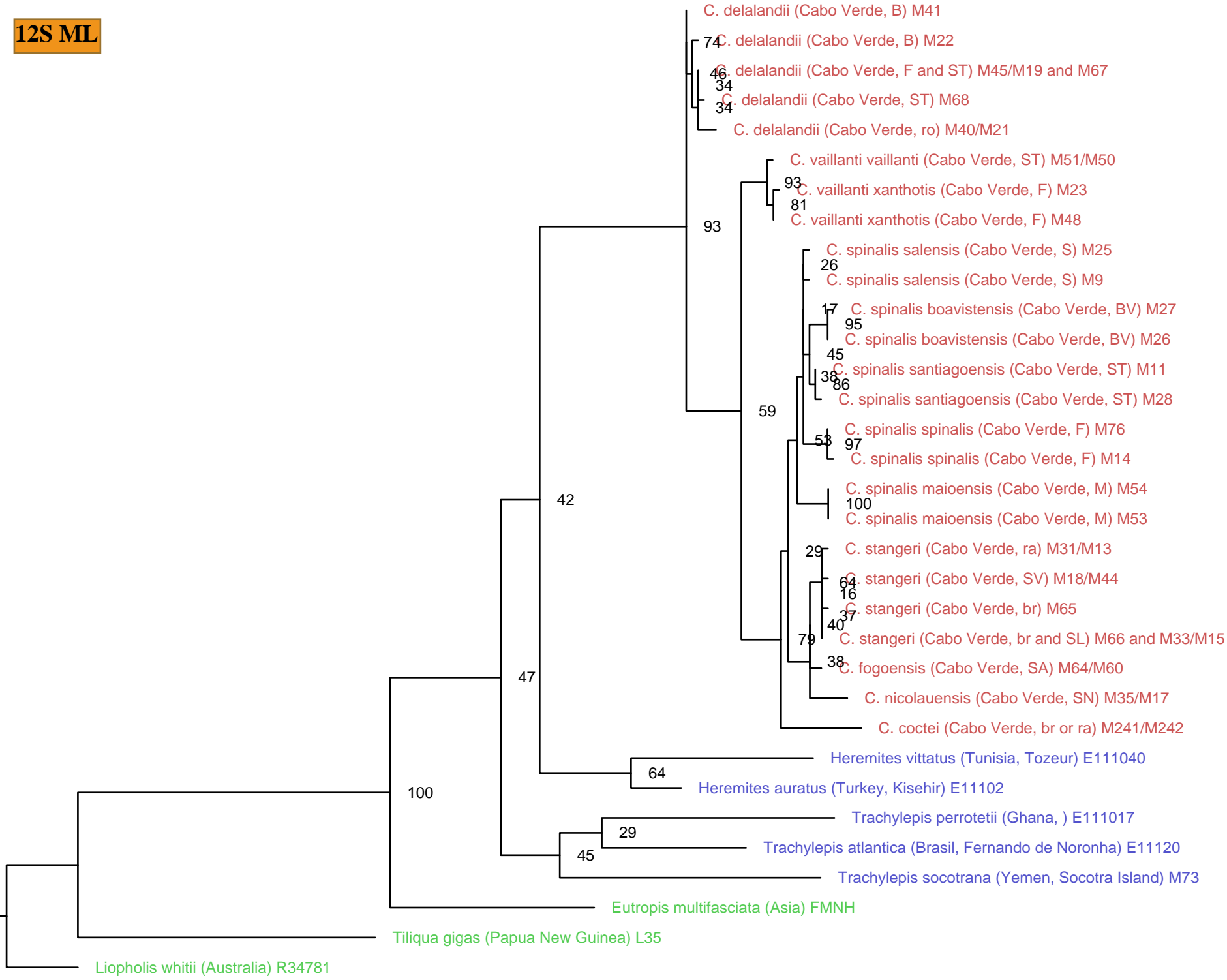
0.2

RAG1 ML

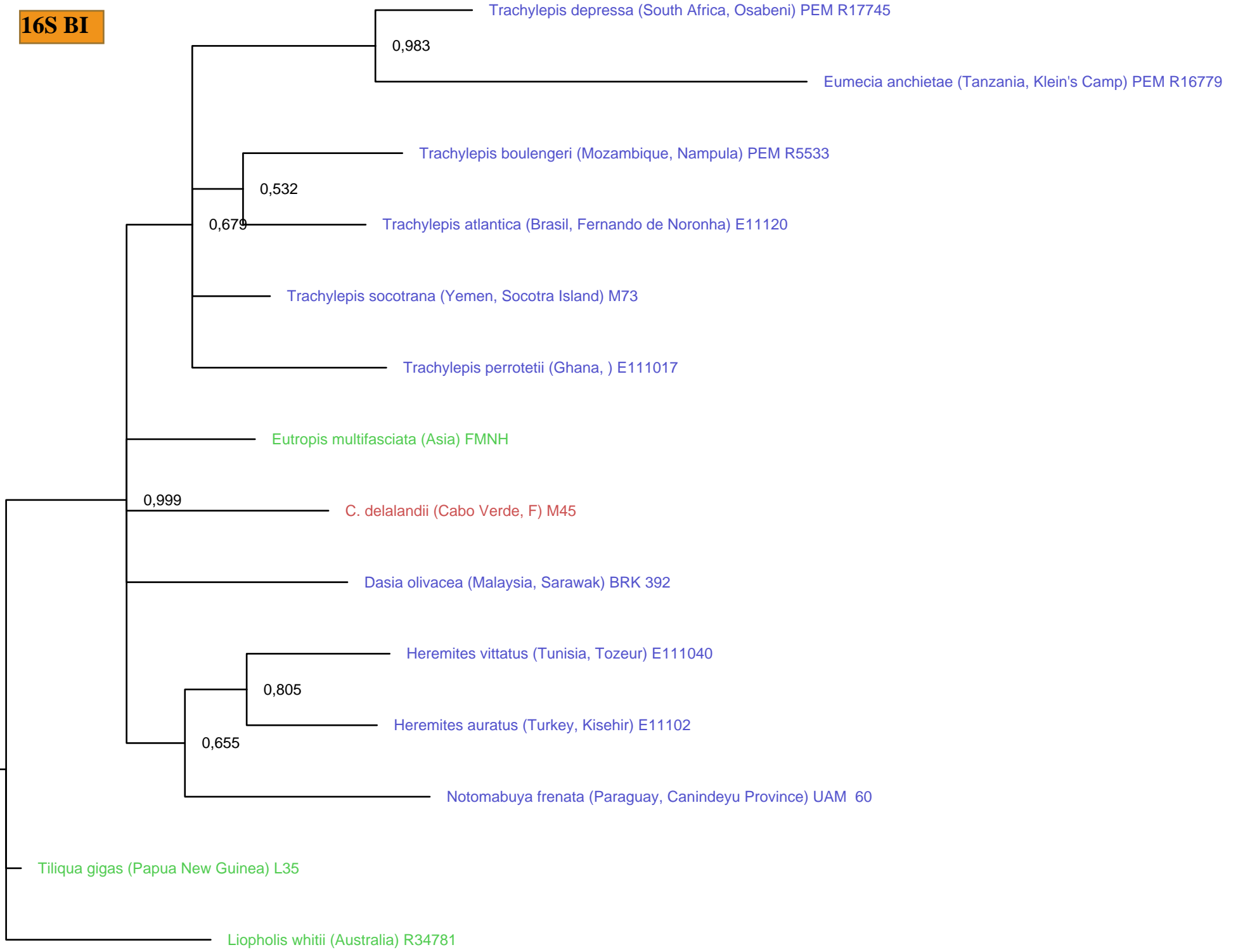




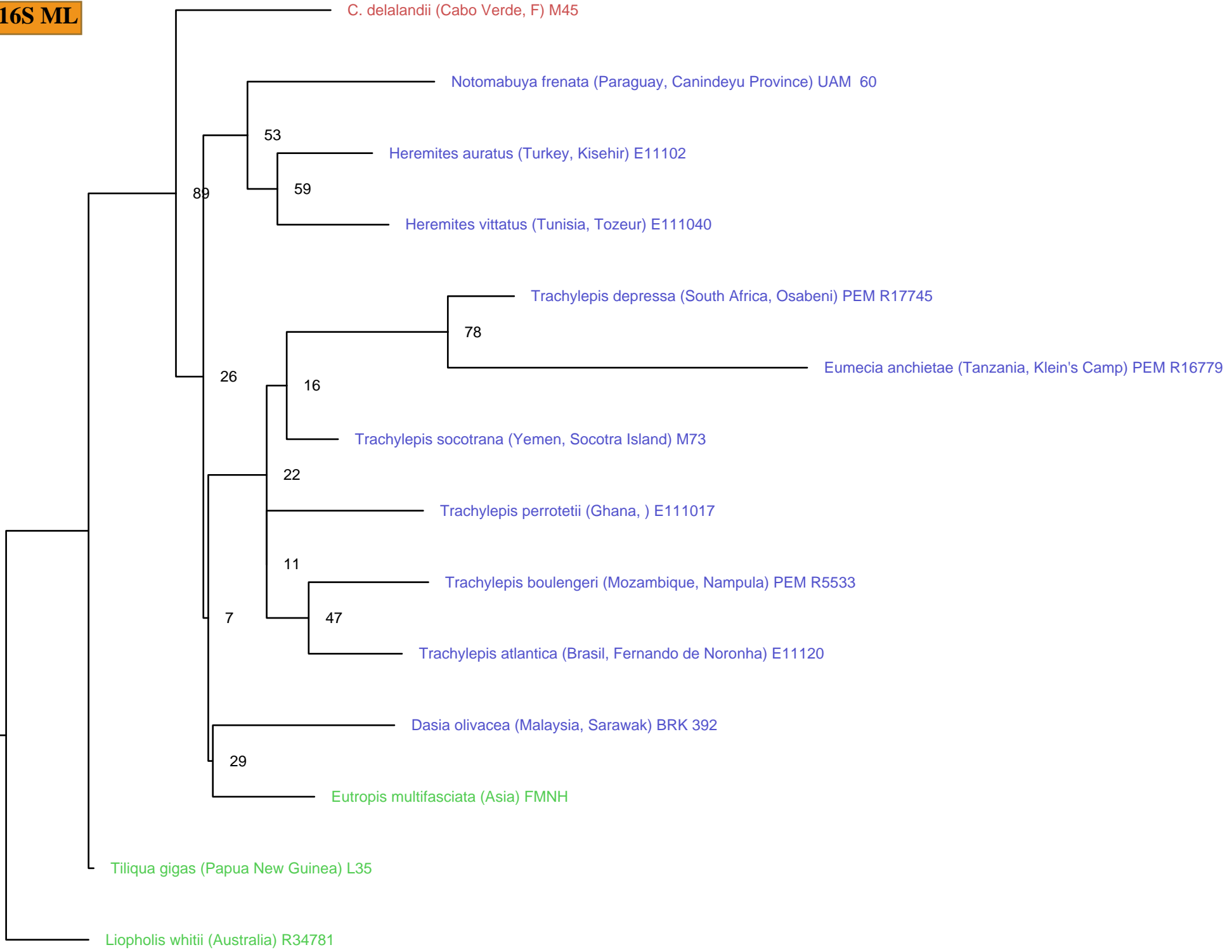
0.05



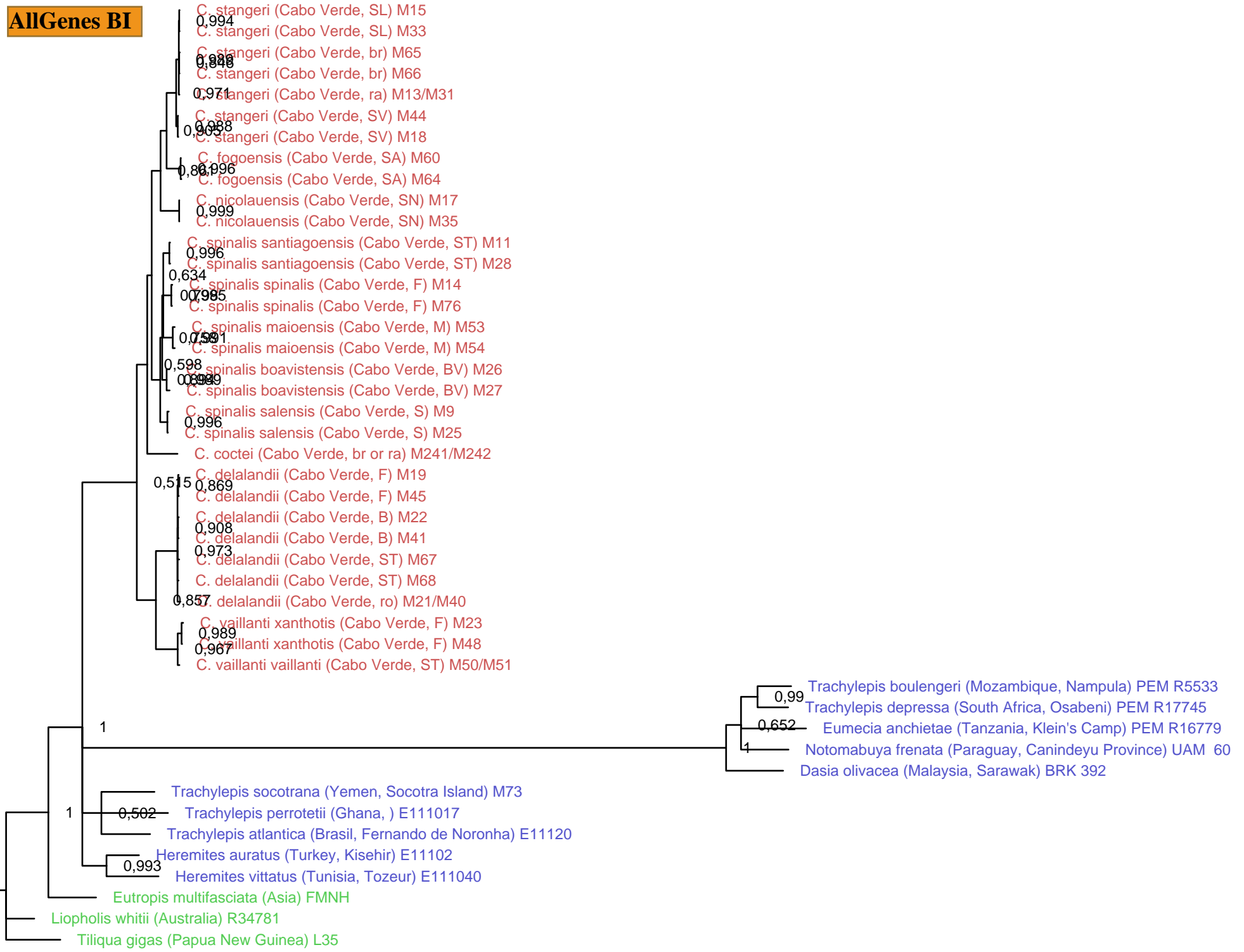
0.04



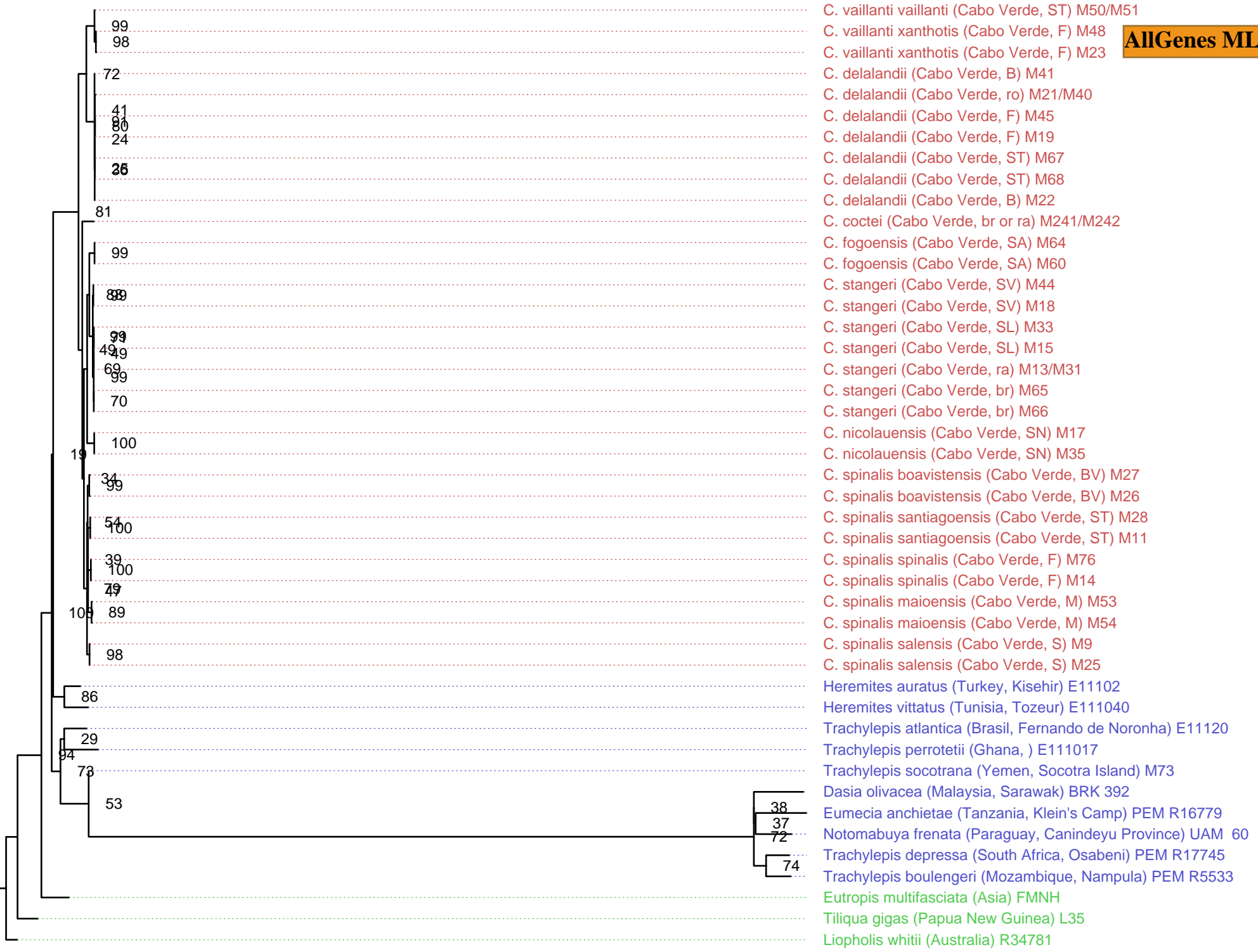
0.08



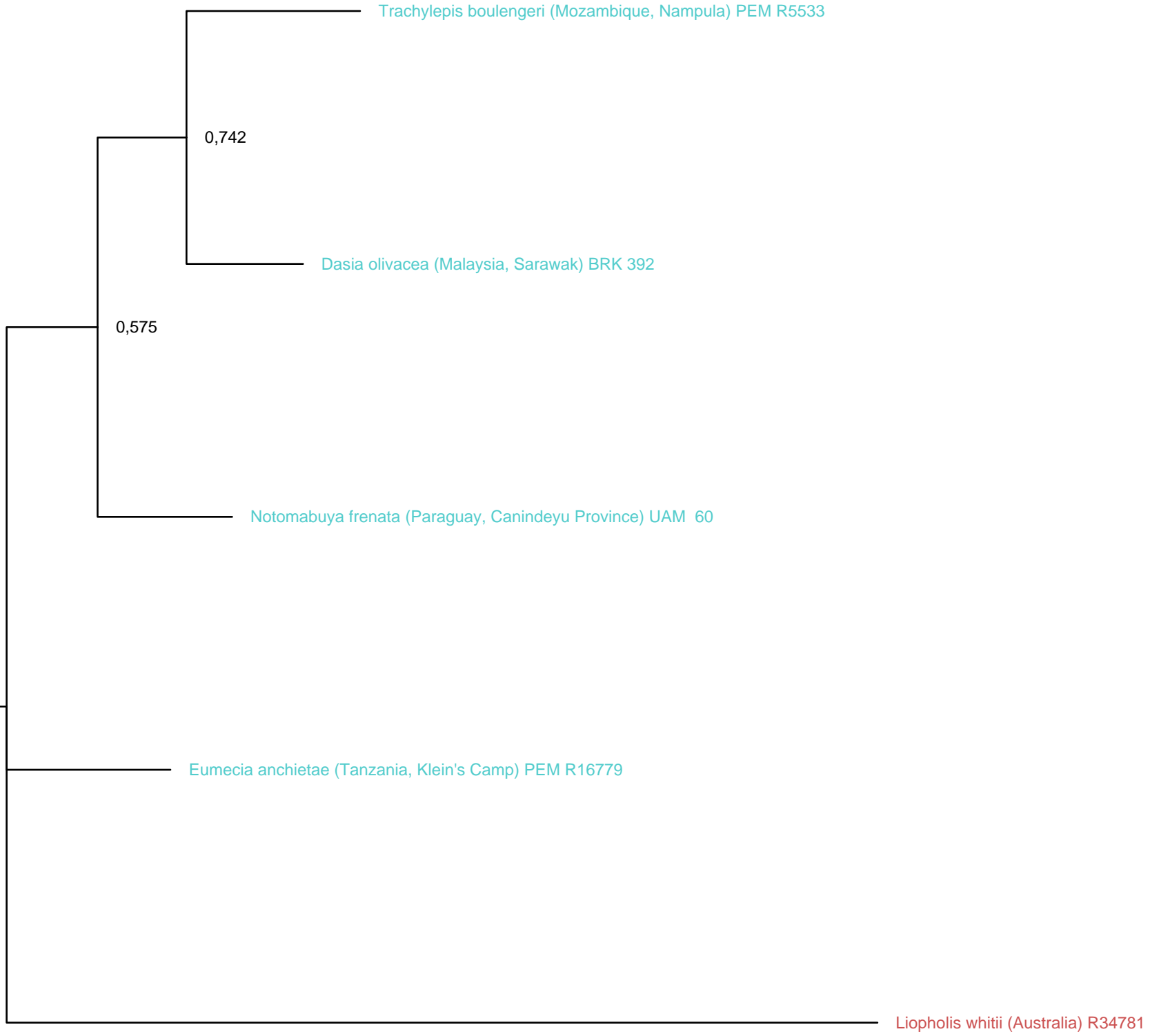
0.1



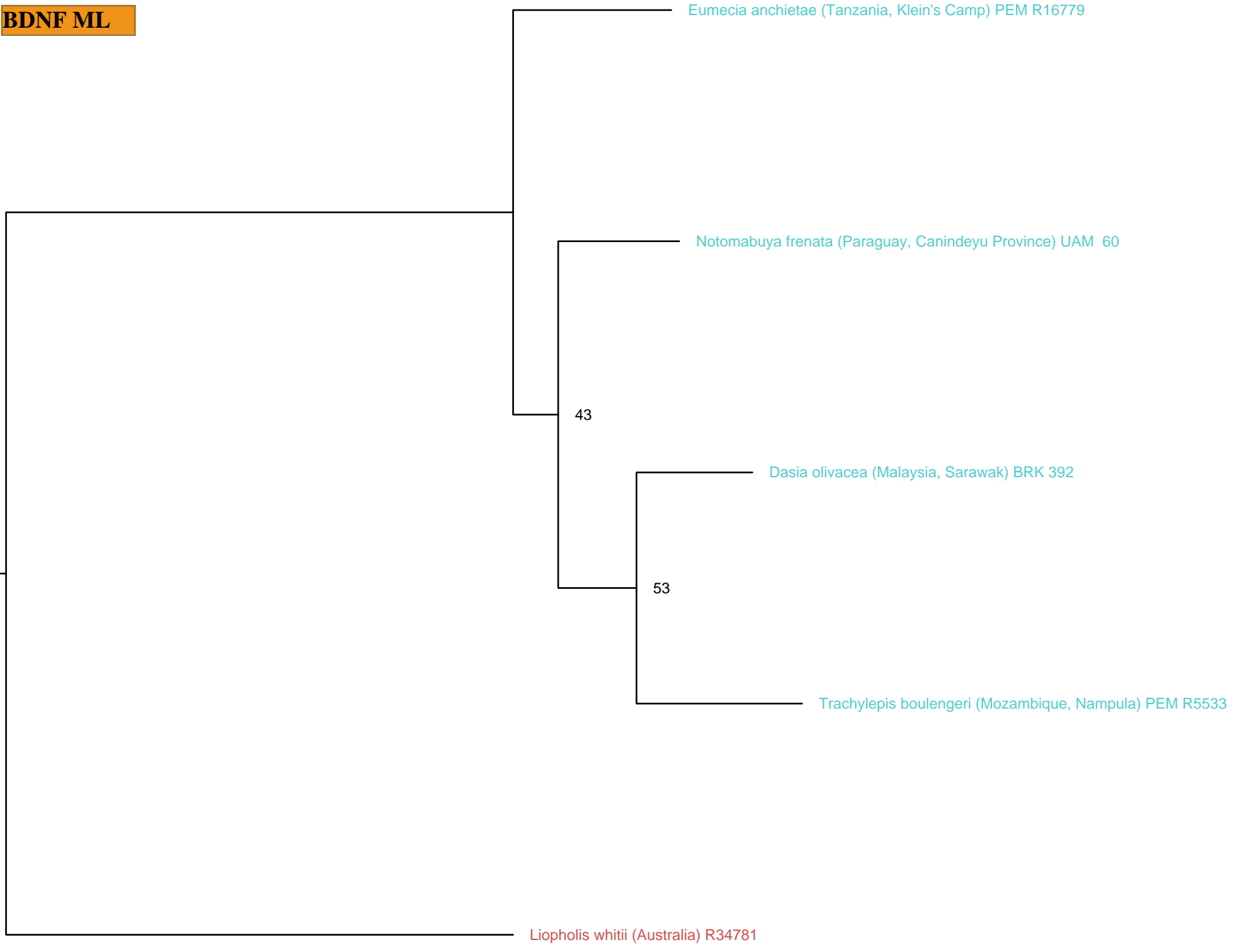
0.2



0.3

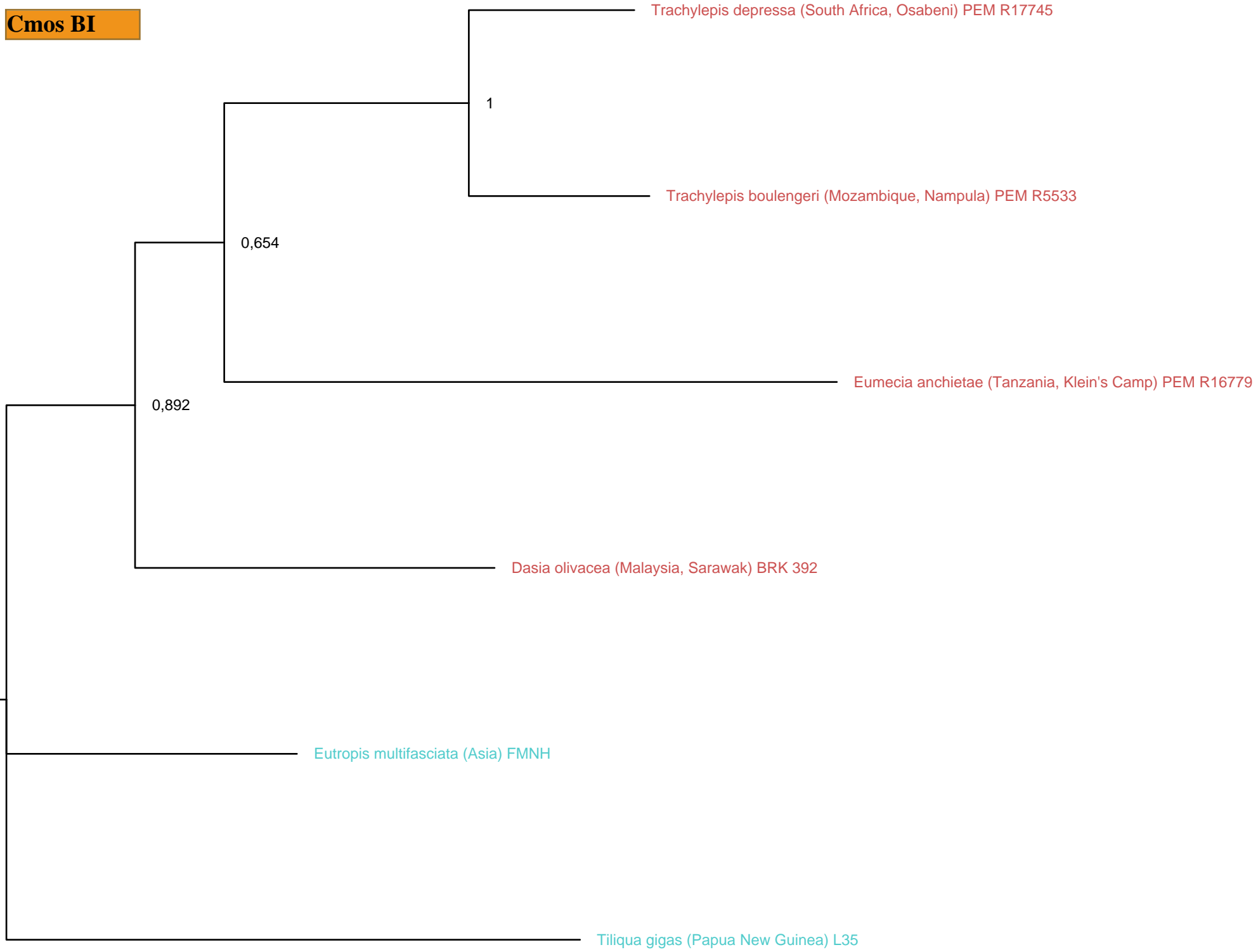


BDNF ML

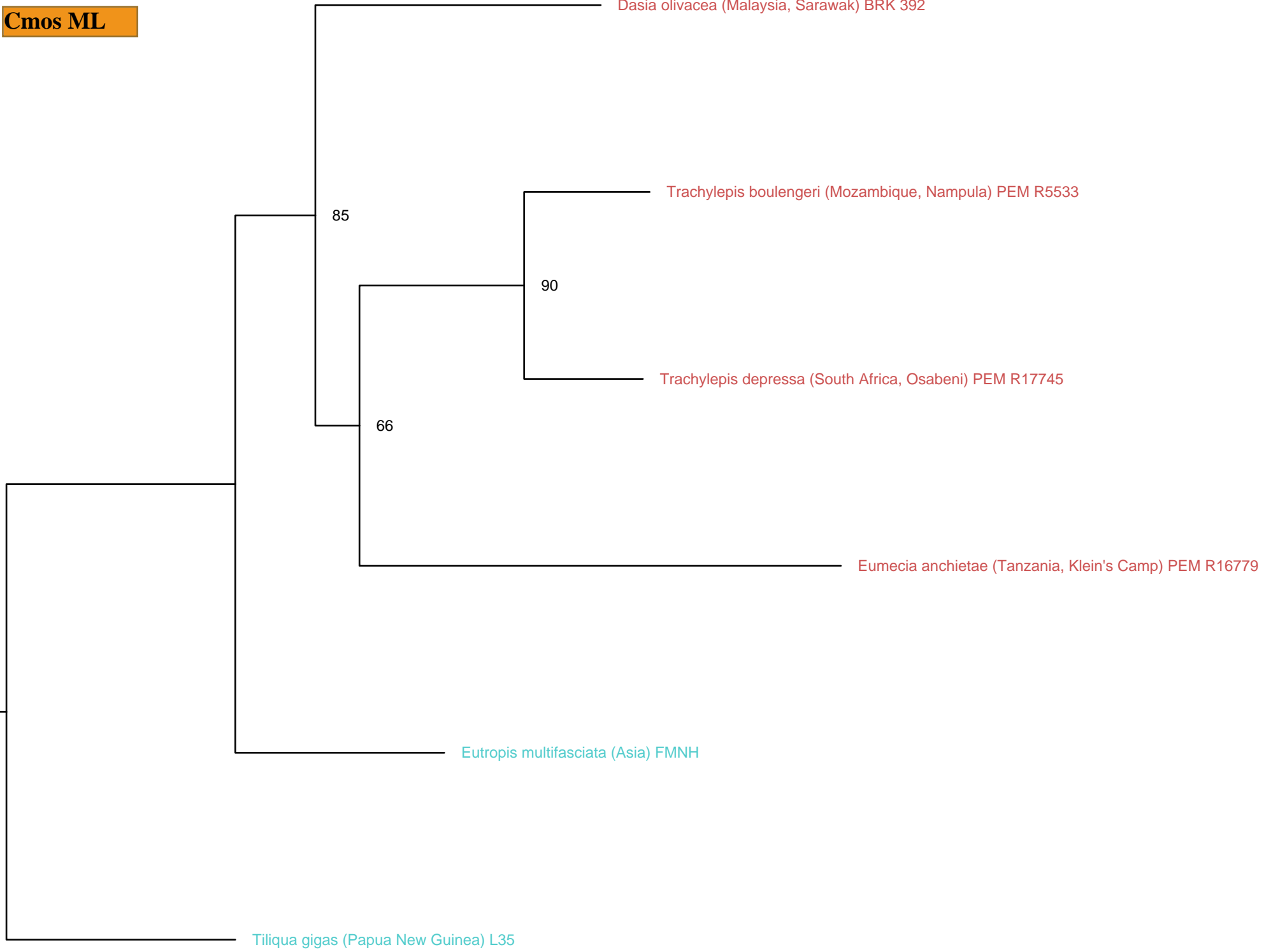


0.02

Cmos BI

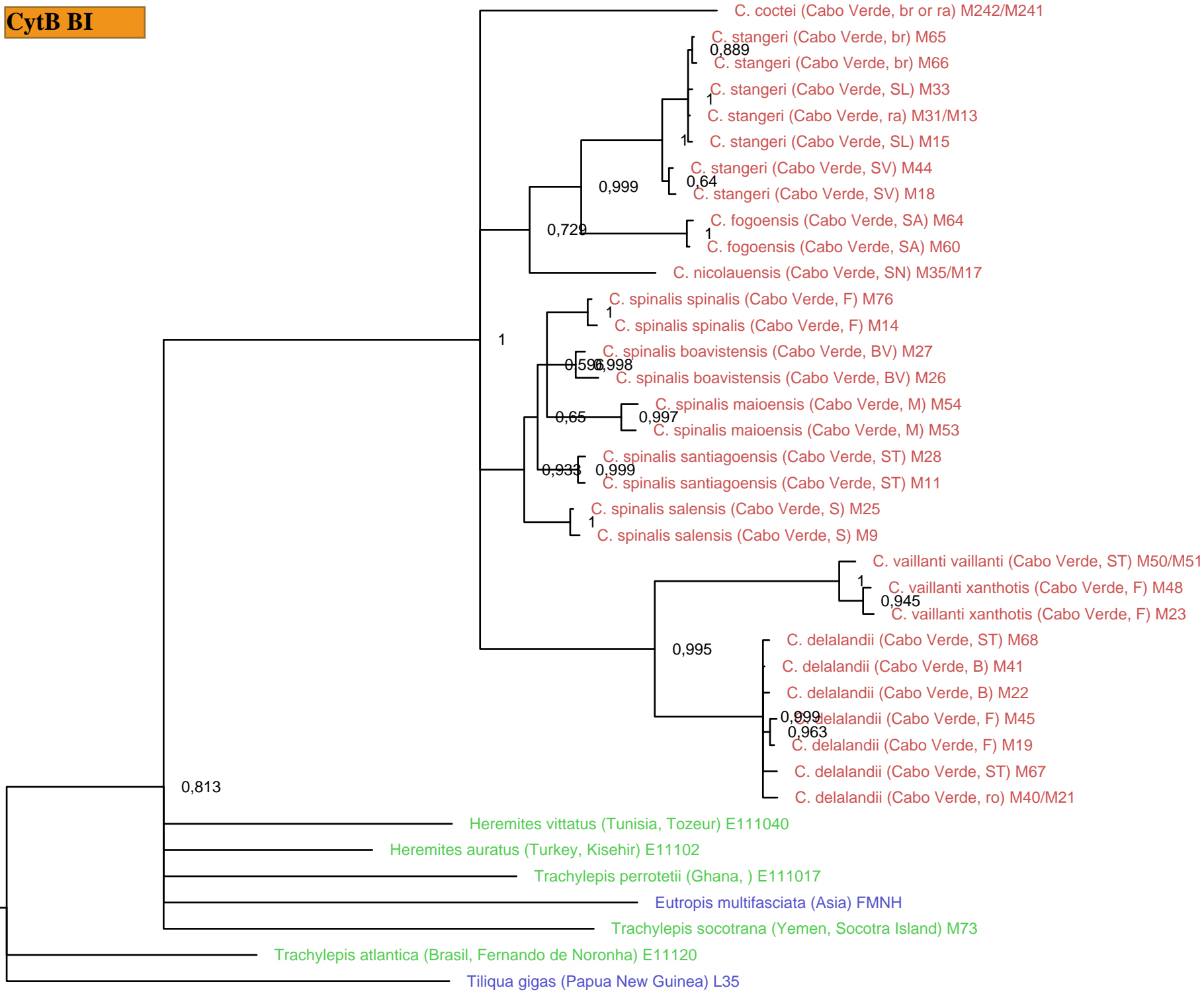


Cmos ML

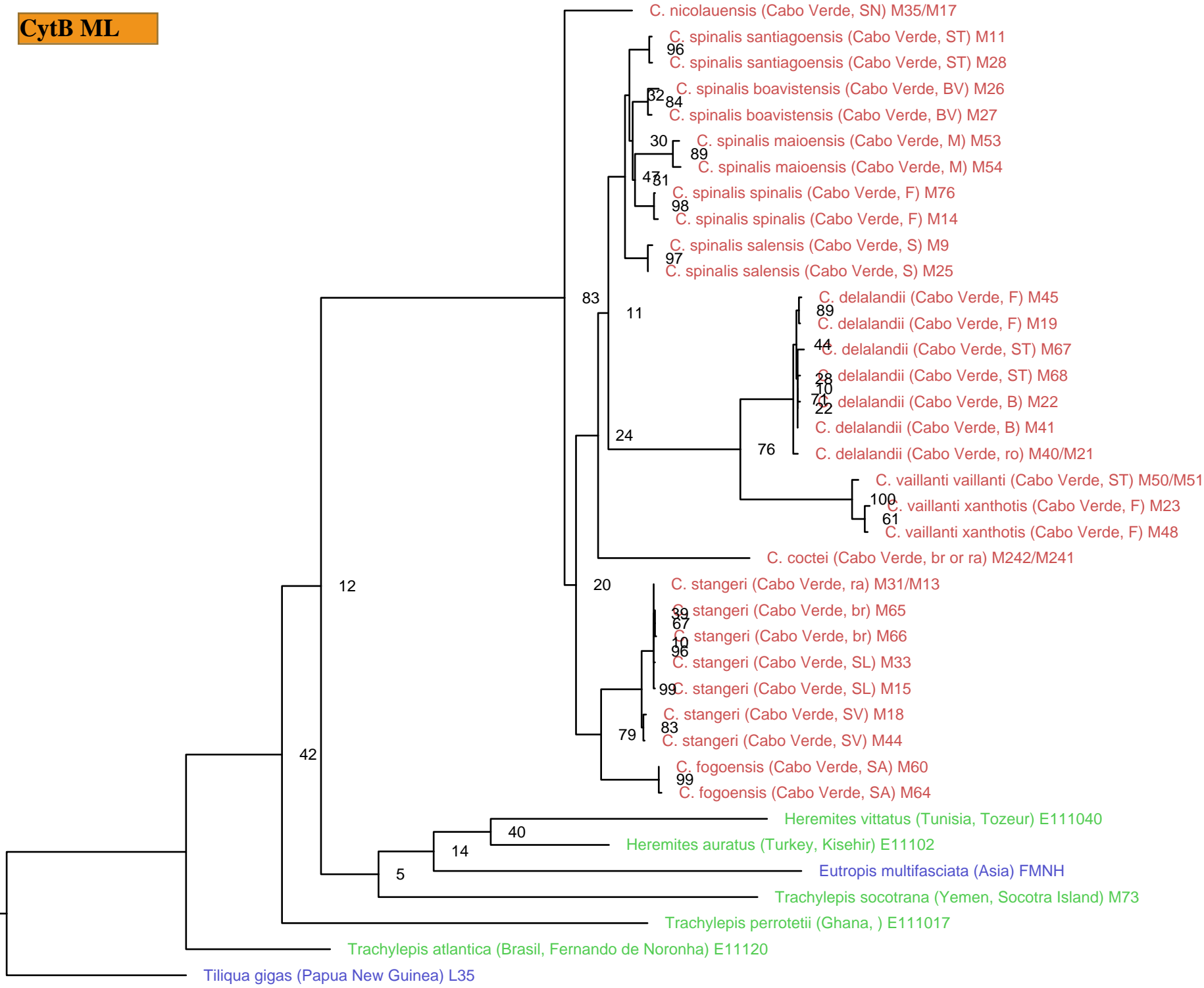


0.006

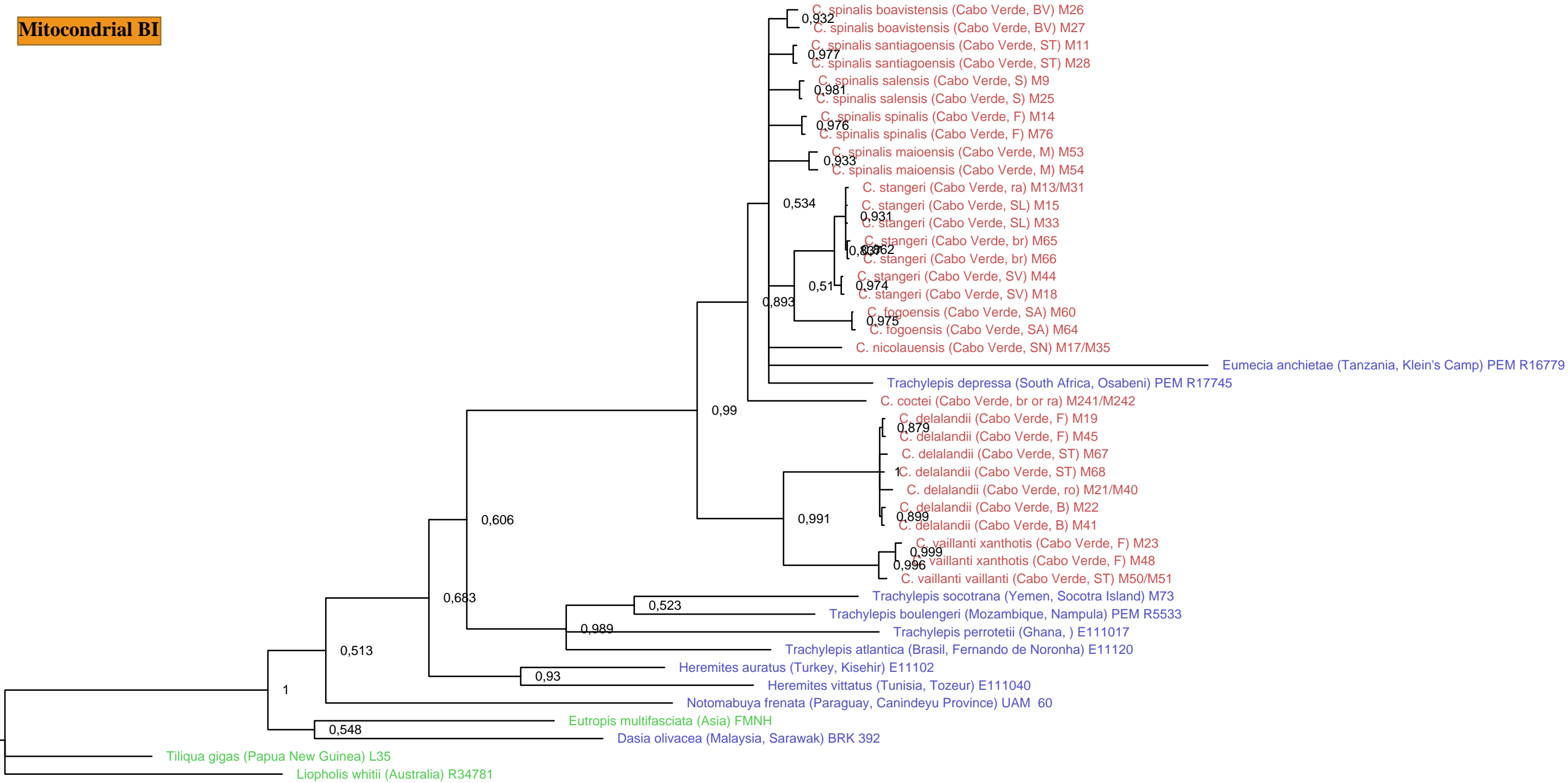
CytB BI



0.05

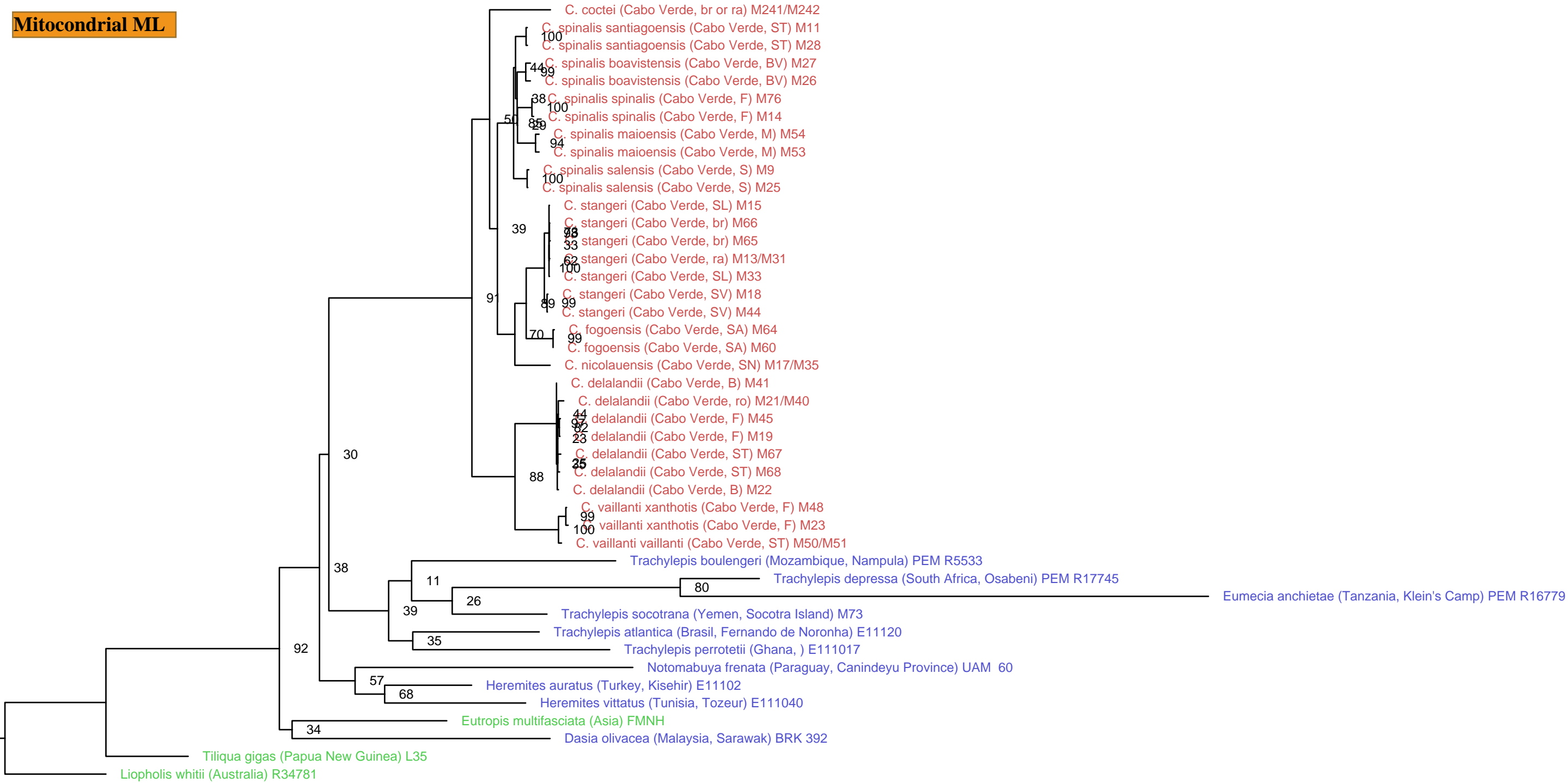


Mitochondrial BI



0.08

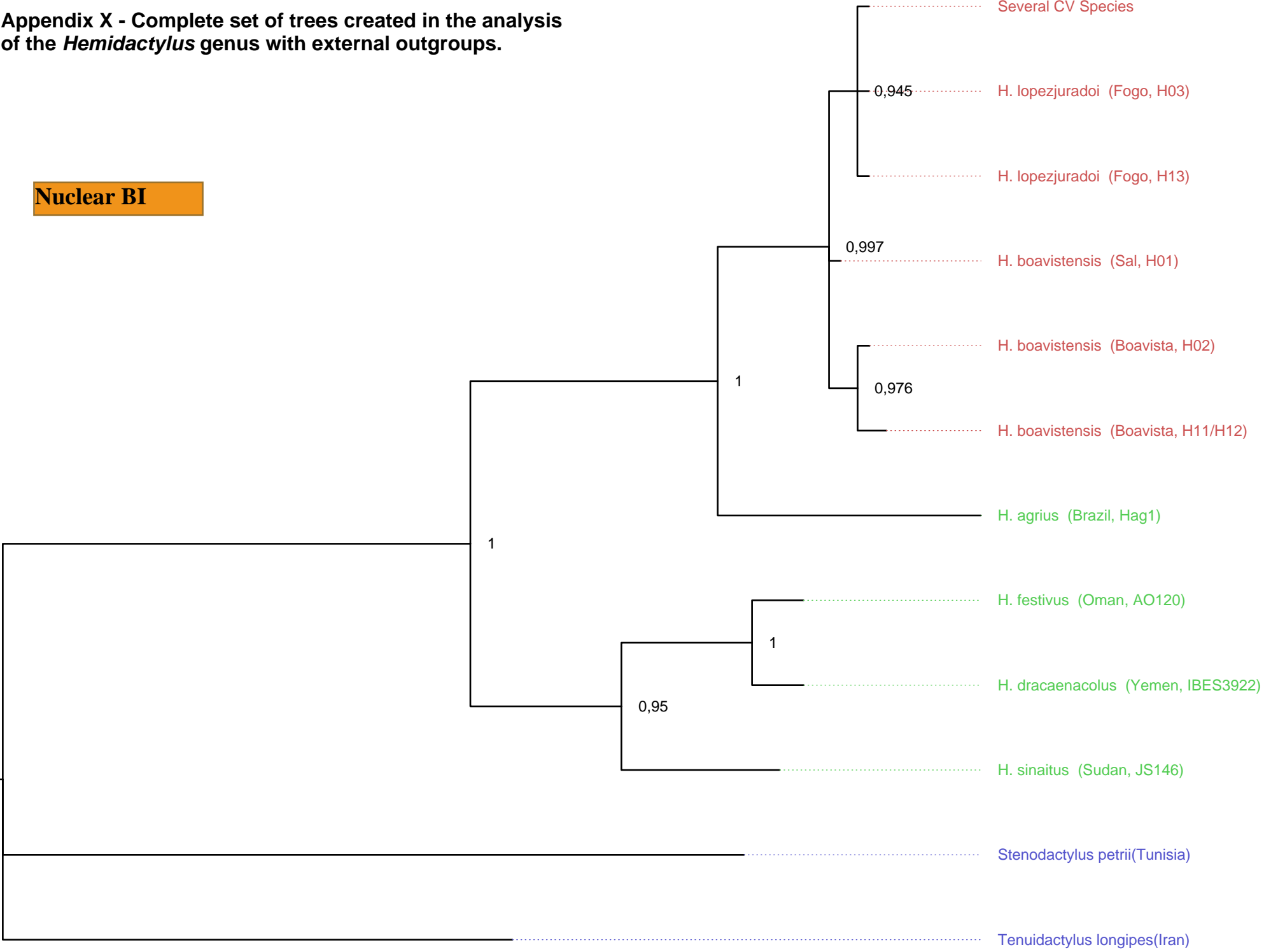
Mitochondrial ML



0.2

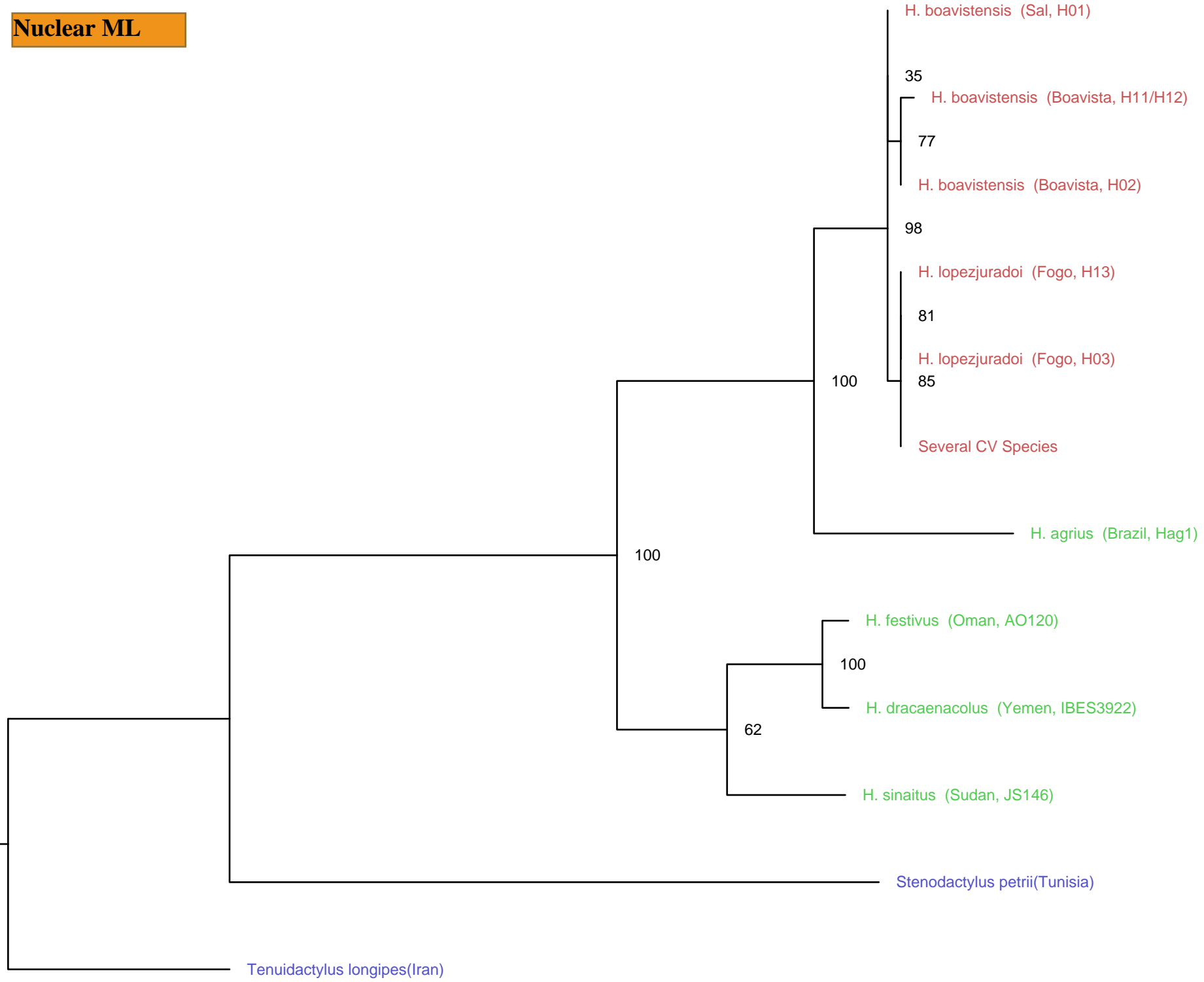
Appendix X - Complete set of trees created in the analysis of the *Hemidactylus* genus with external outgroups.

Nuclear BI



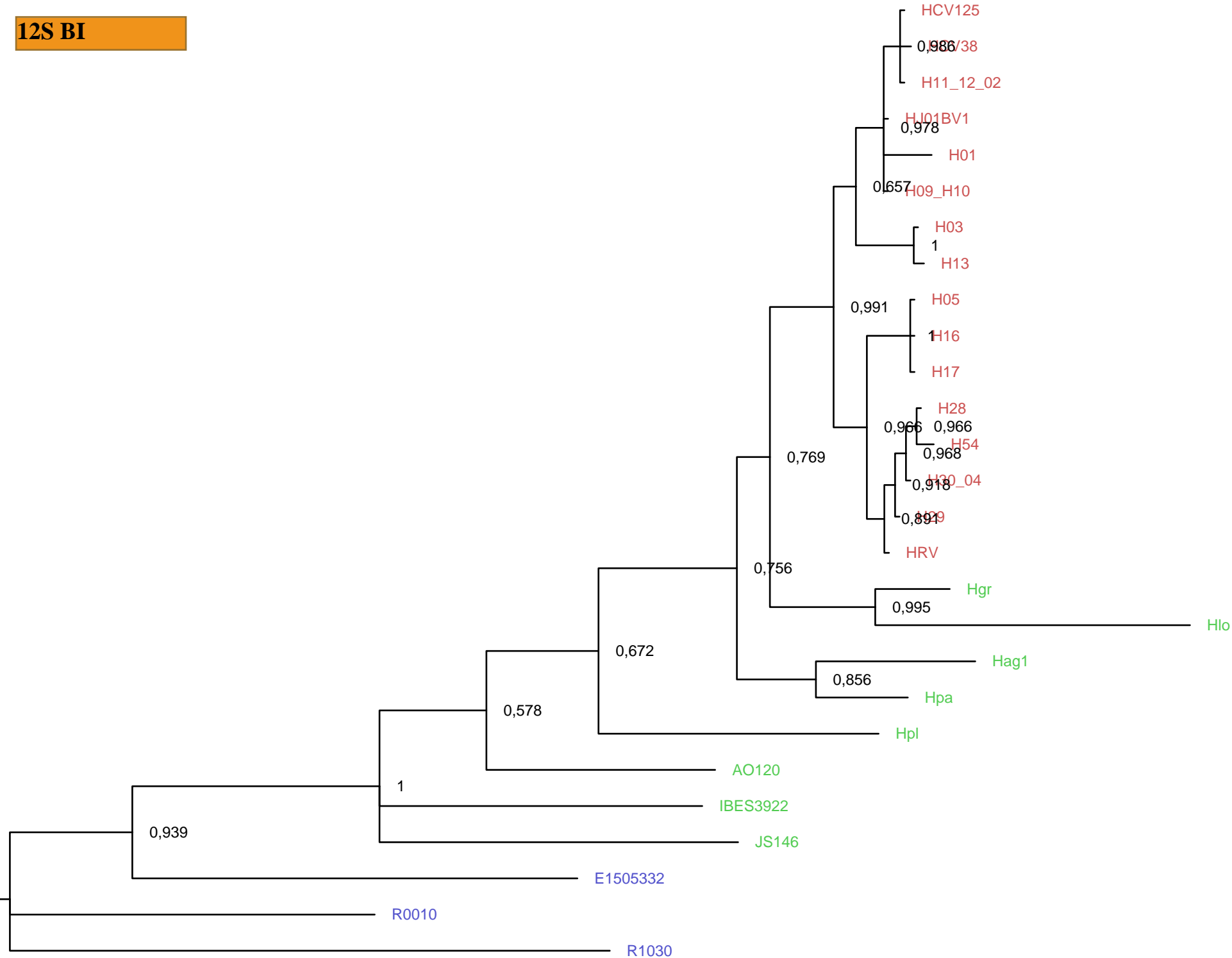
0.007

Nuclear ML

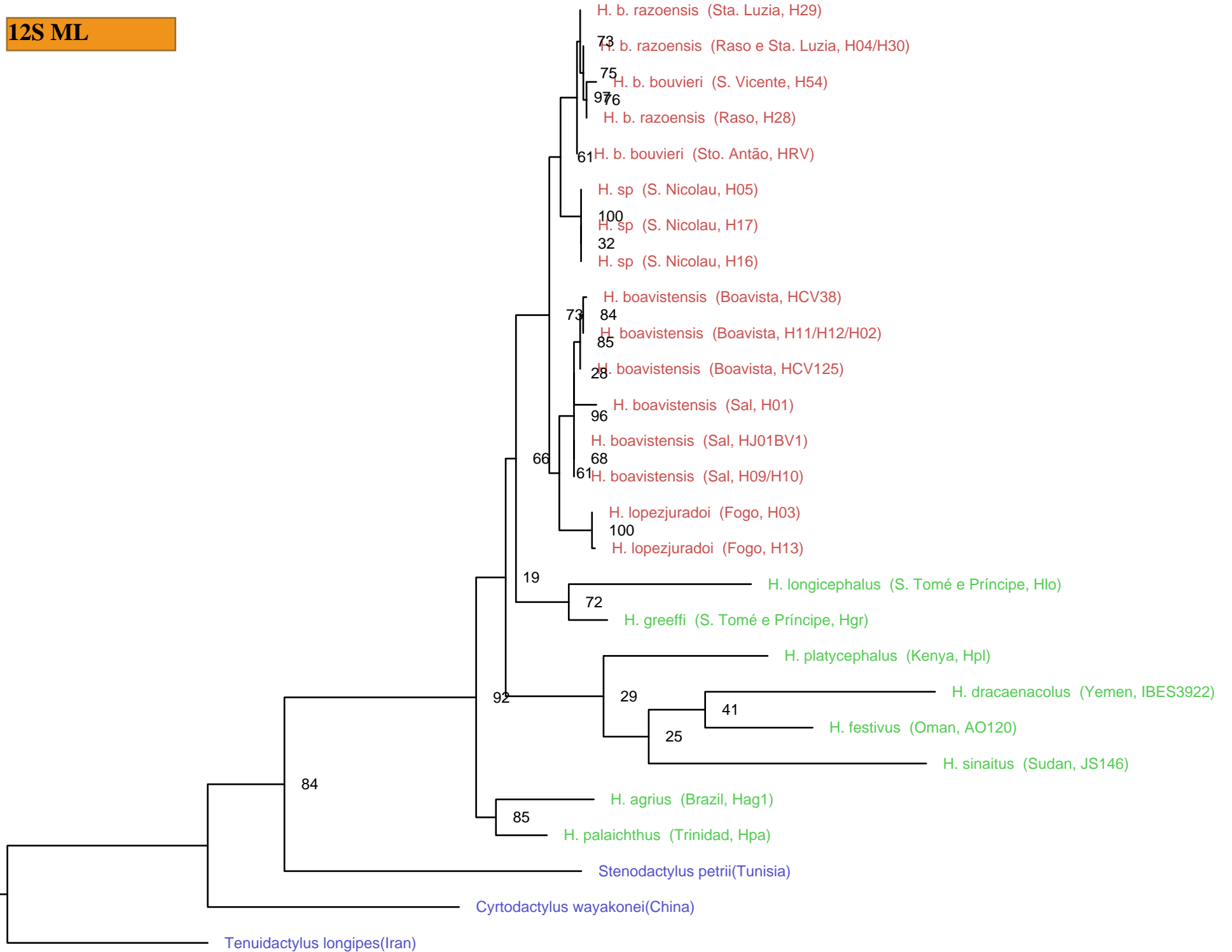


0.01

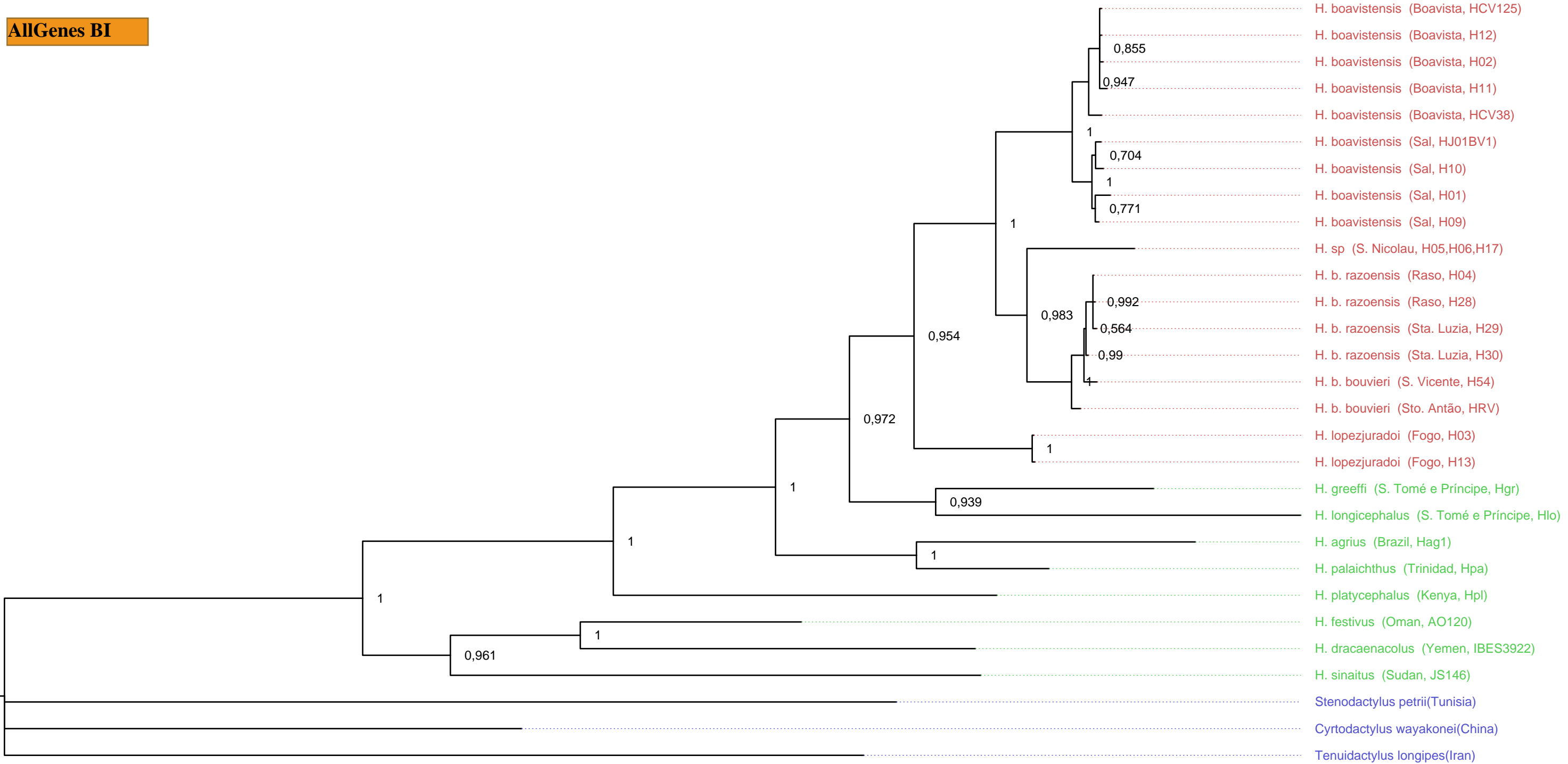
12S BI

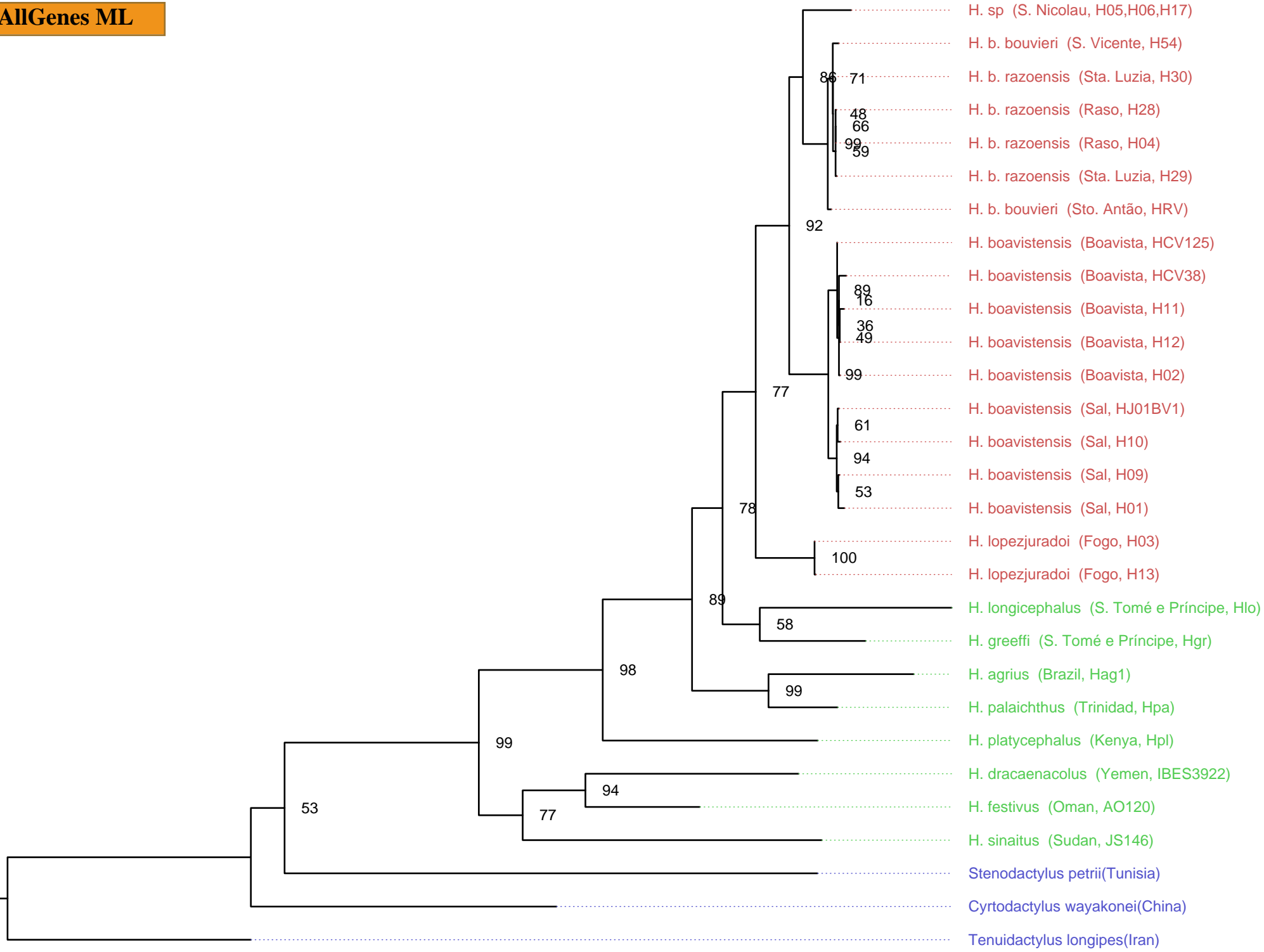


0.09



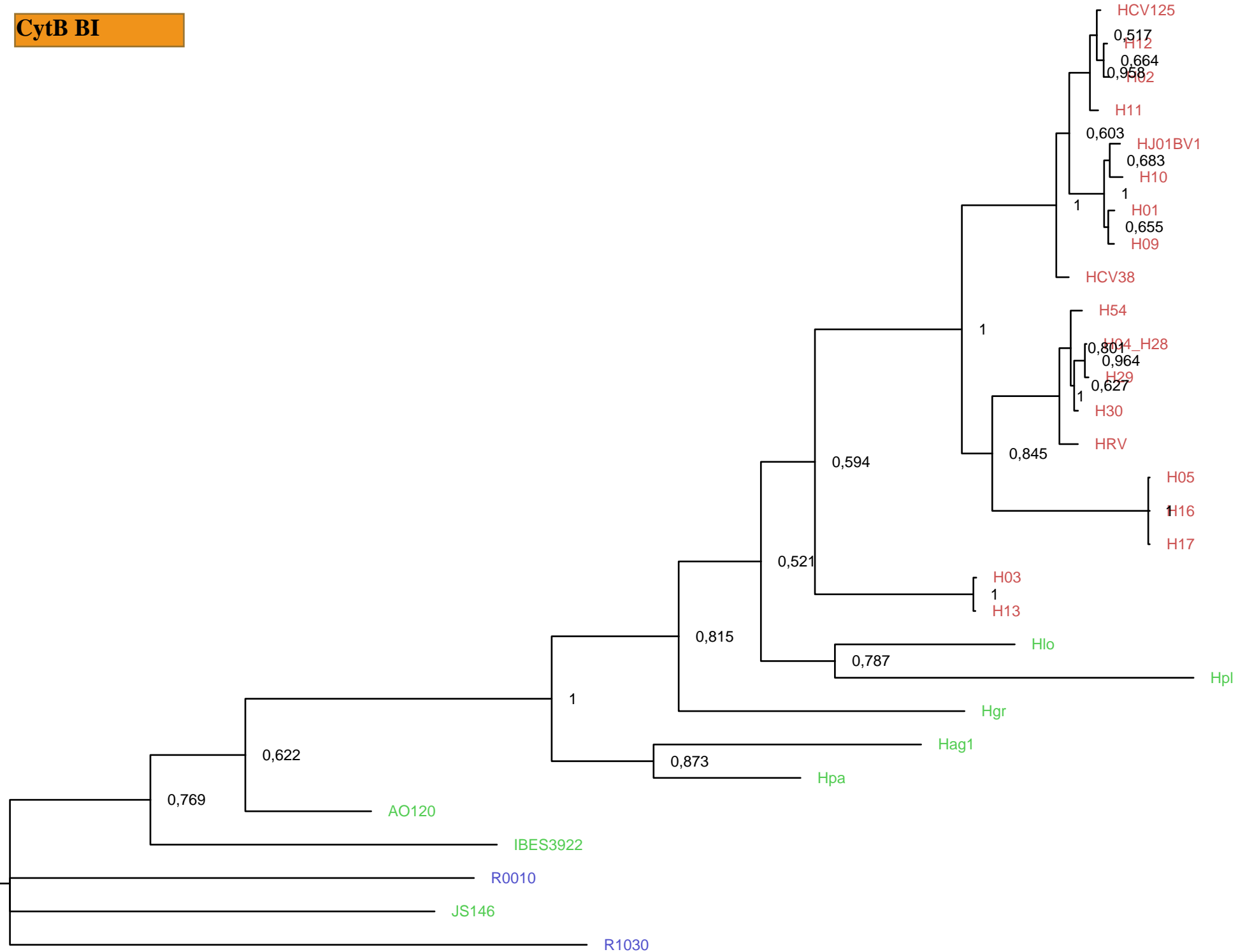
AllGenes BI





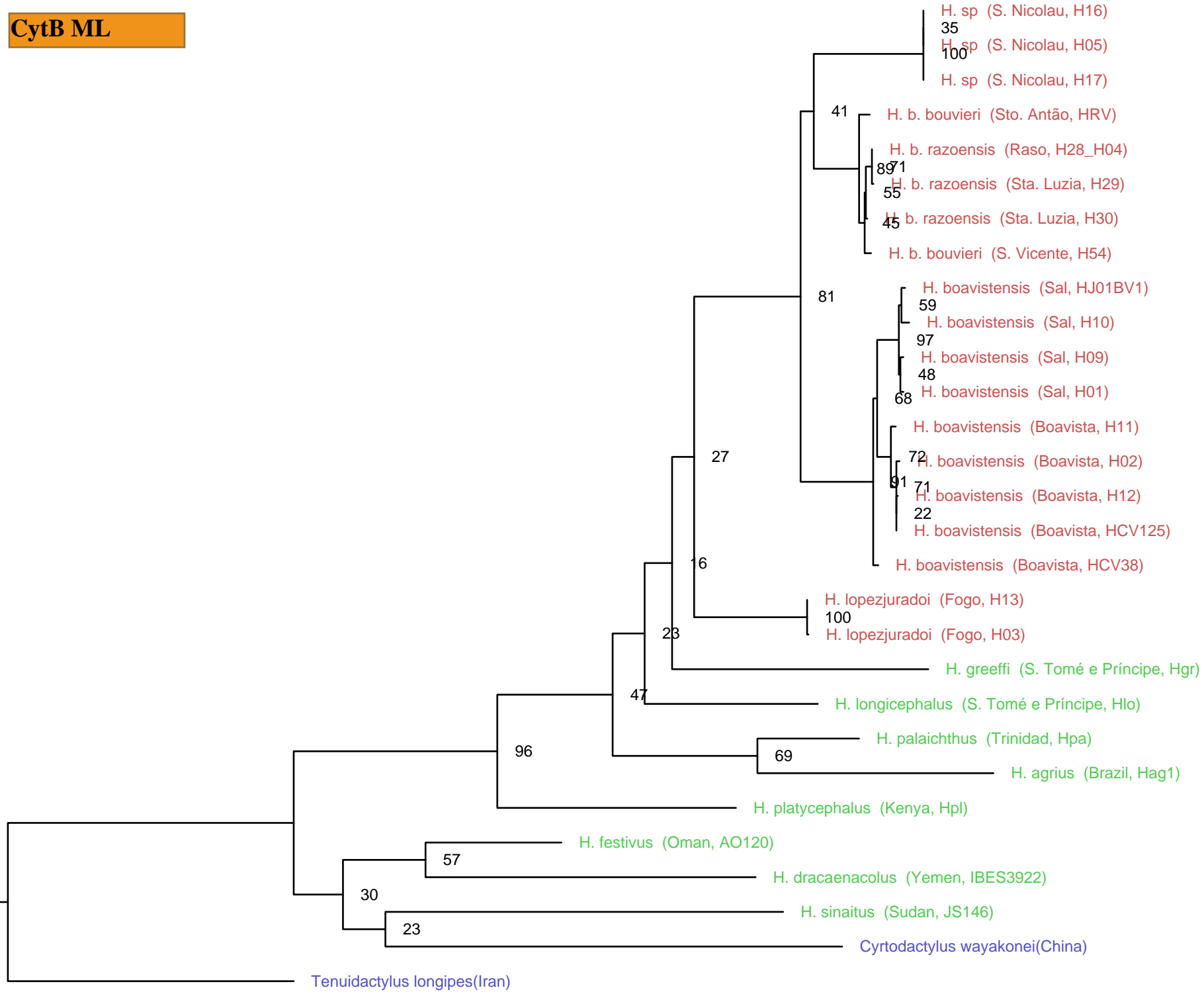
0.2

CytB BI



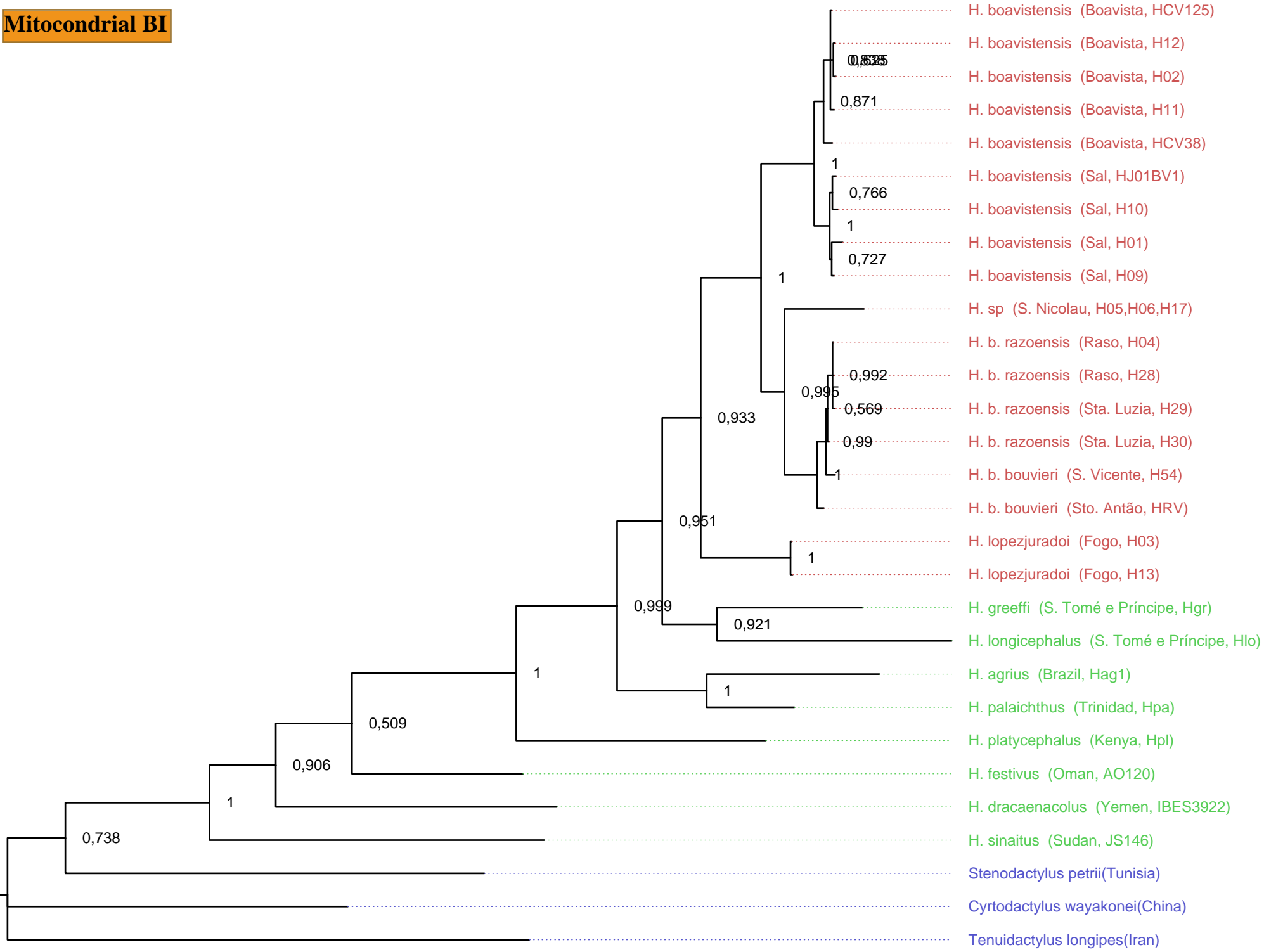
0.09

CytB ML

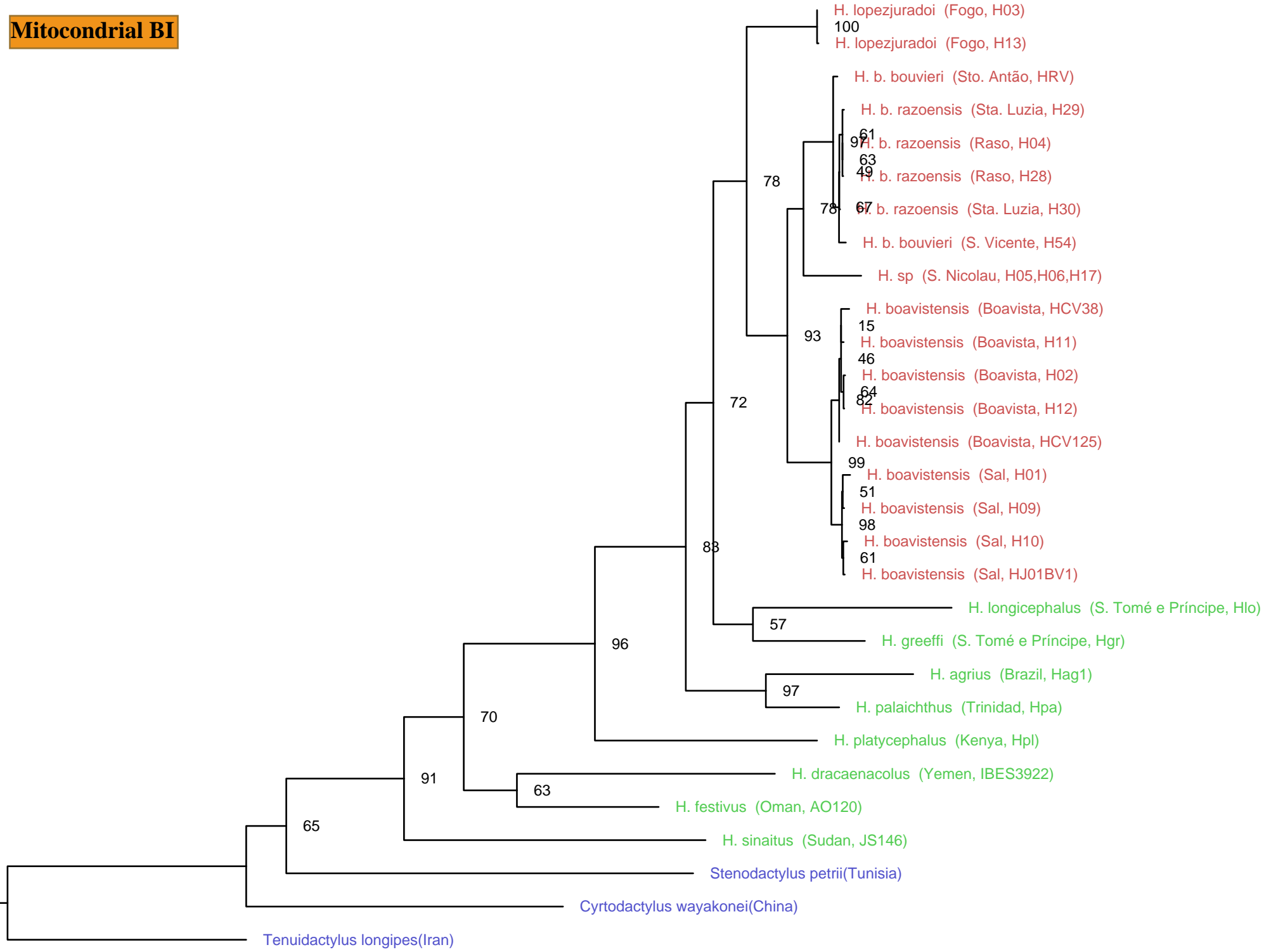


0.2

Mitochondrial BI



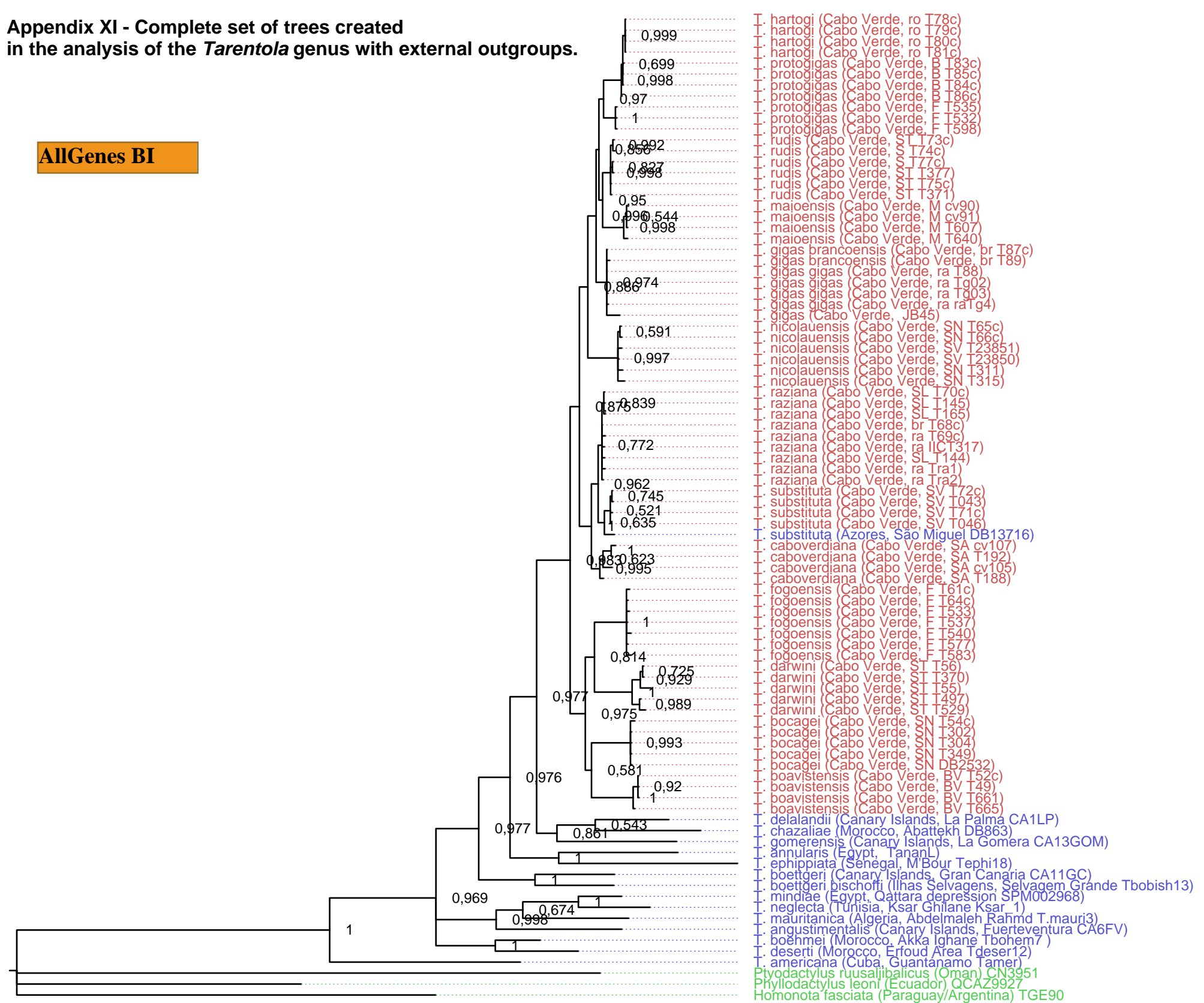
Mitochondrial BI



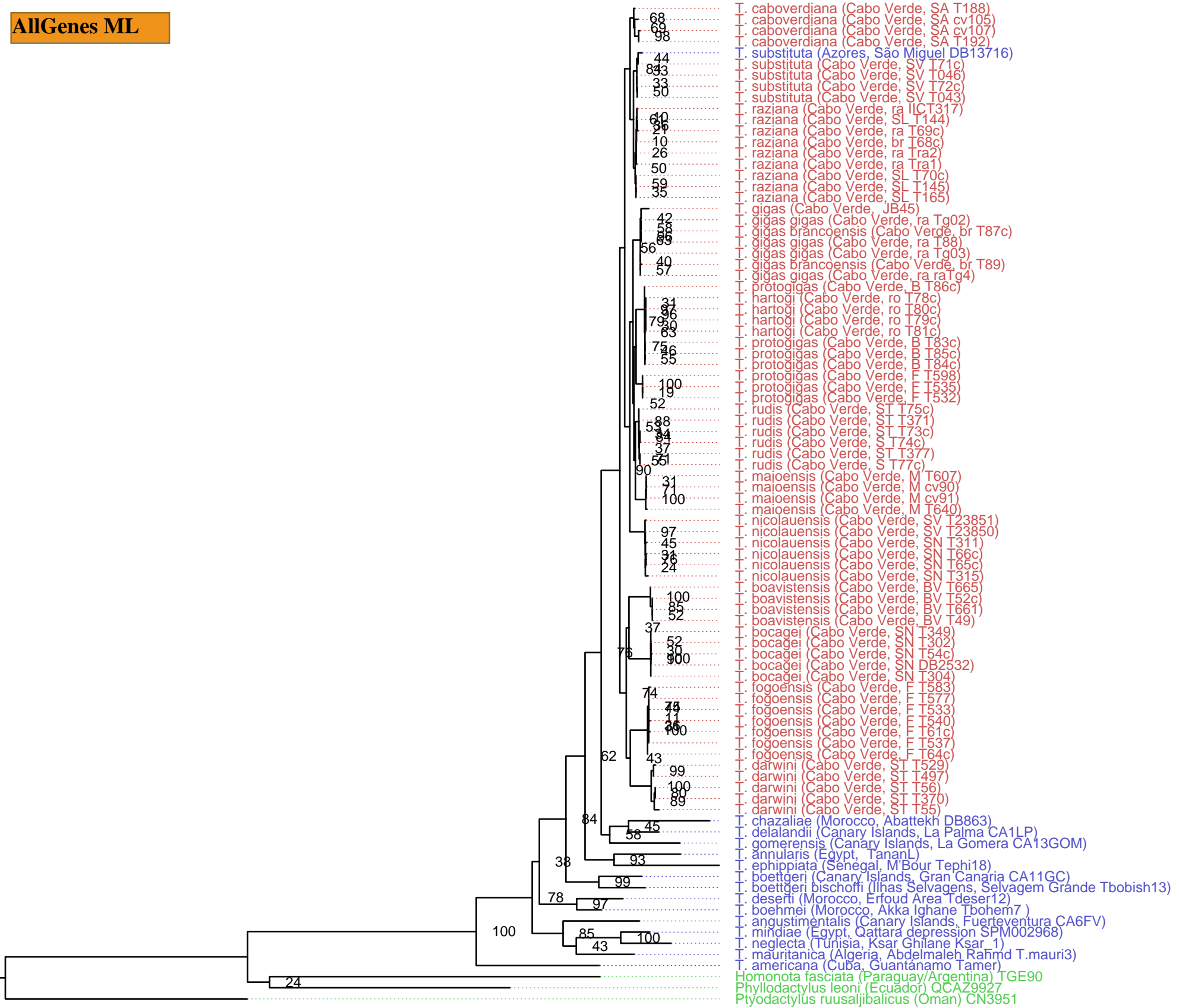
0.2

Appendix XI - Complete set of trees created
in the analysis of the *Tarentola* genus with external outgroups.

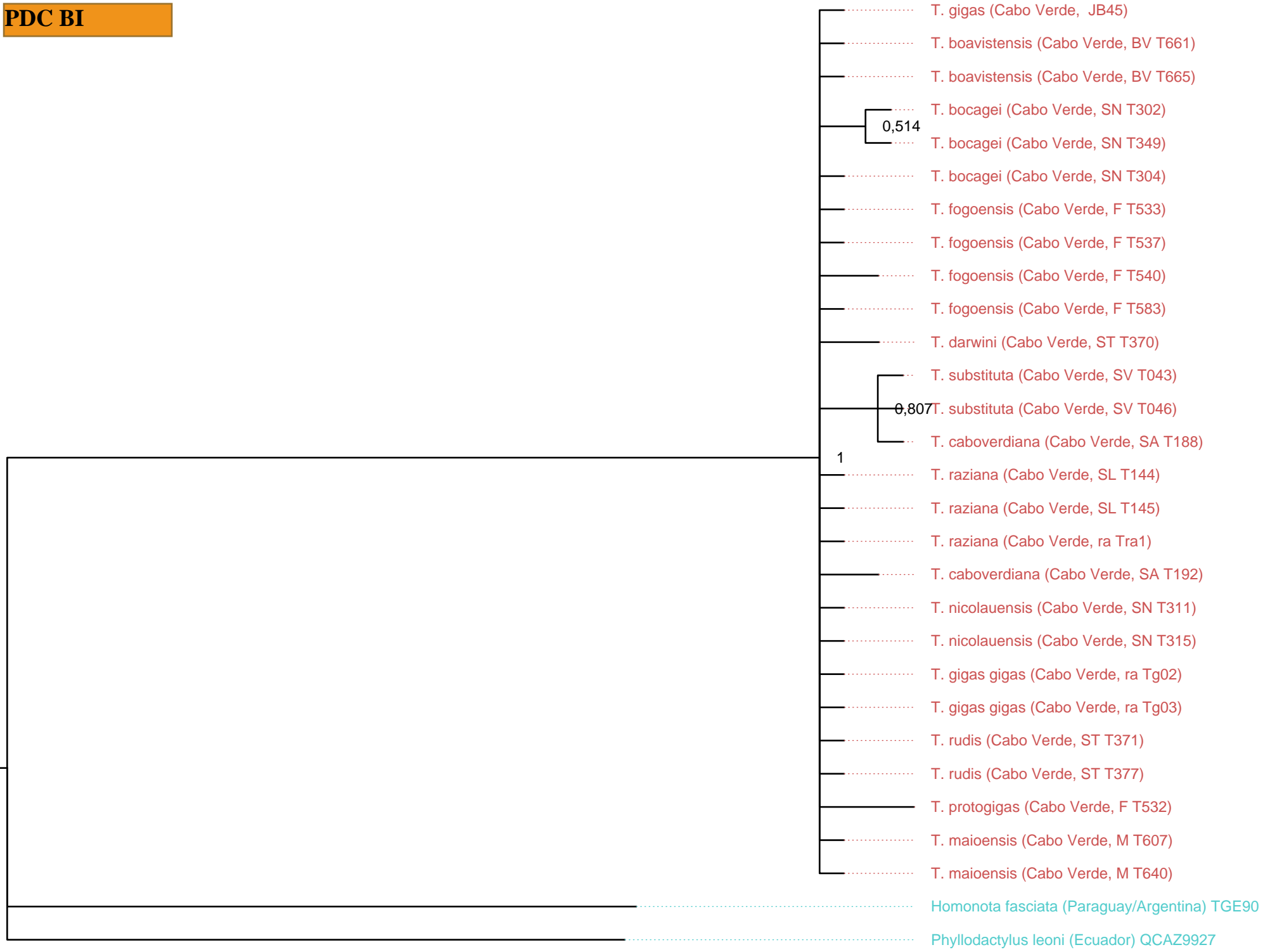
AllGenes BI



0.04



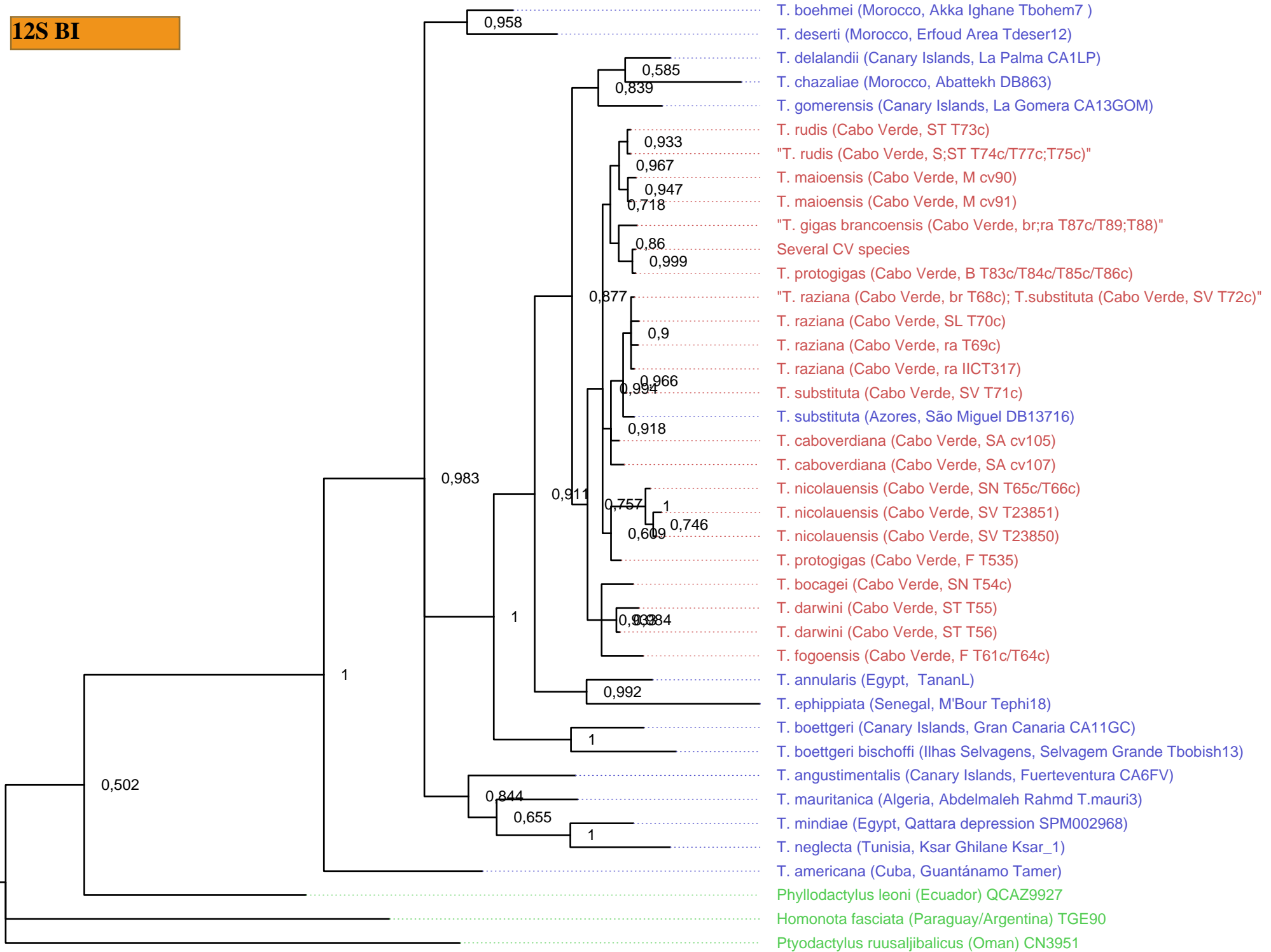
0.07



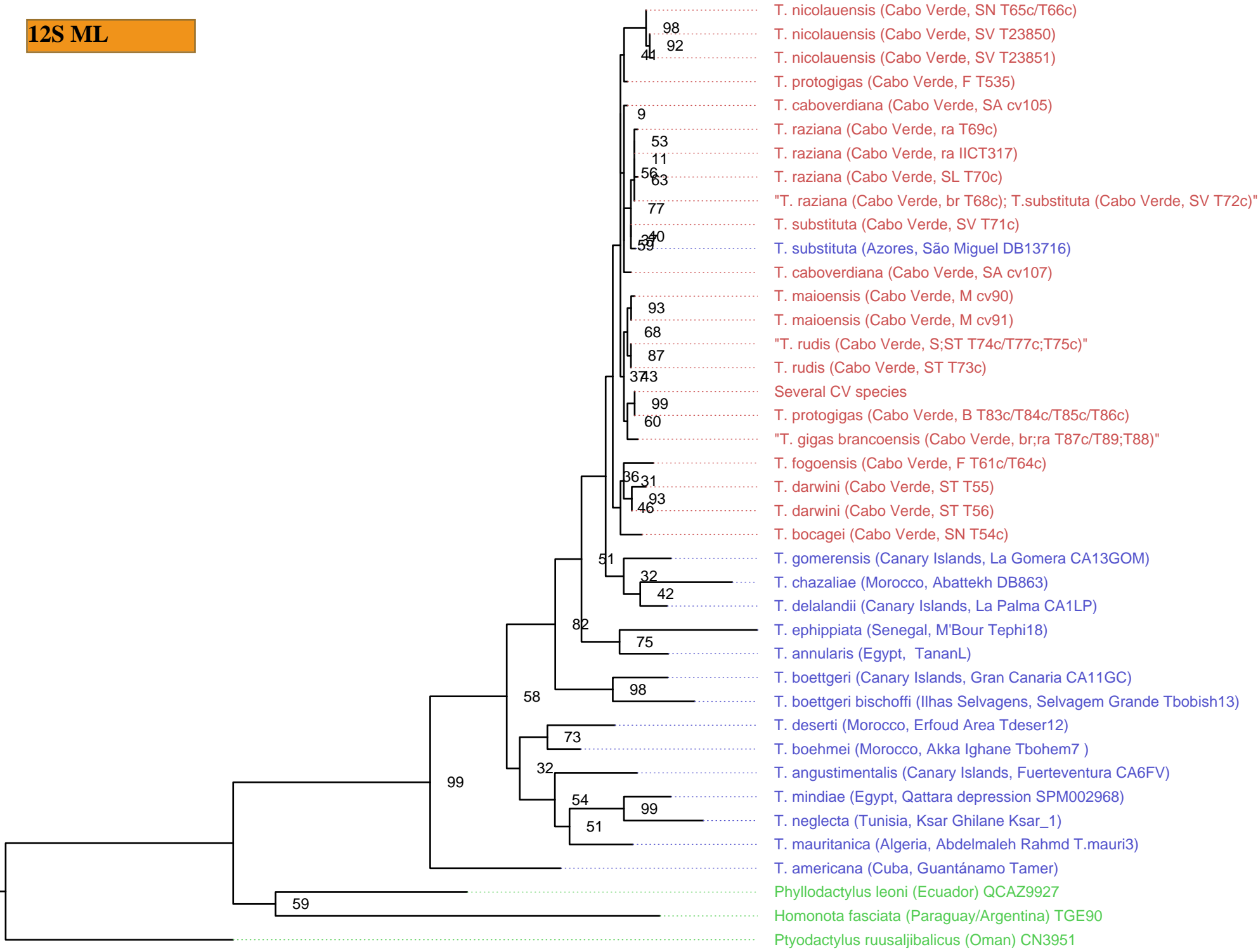
0.004



0.02

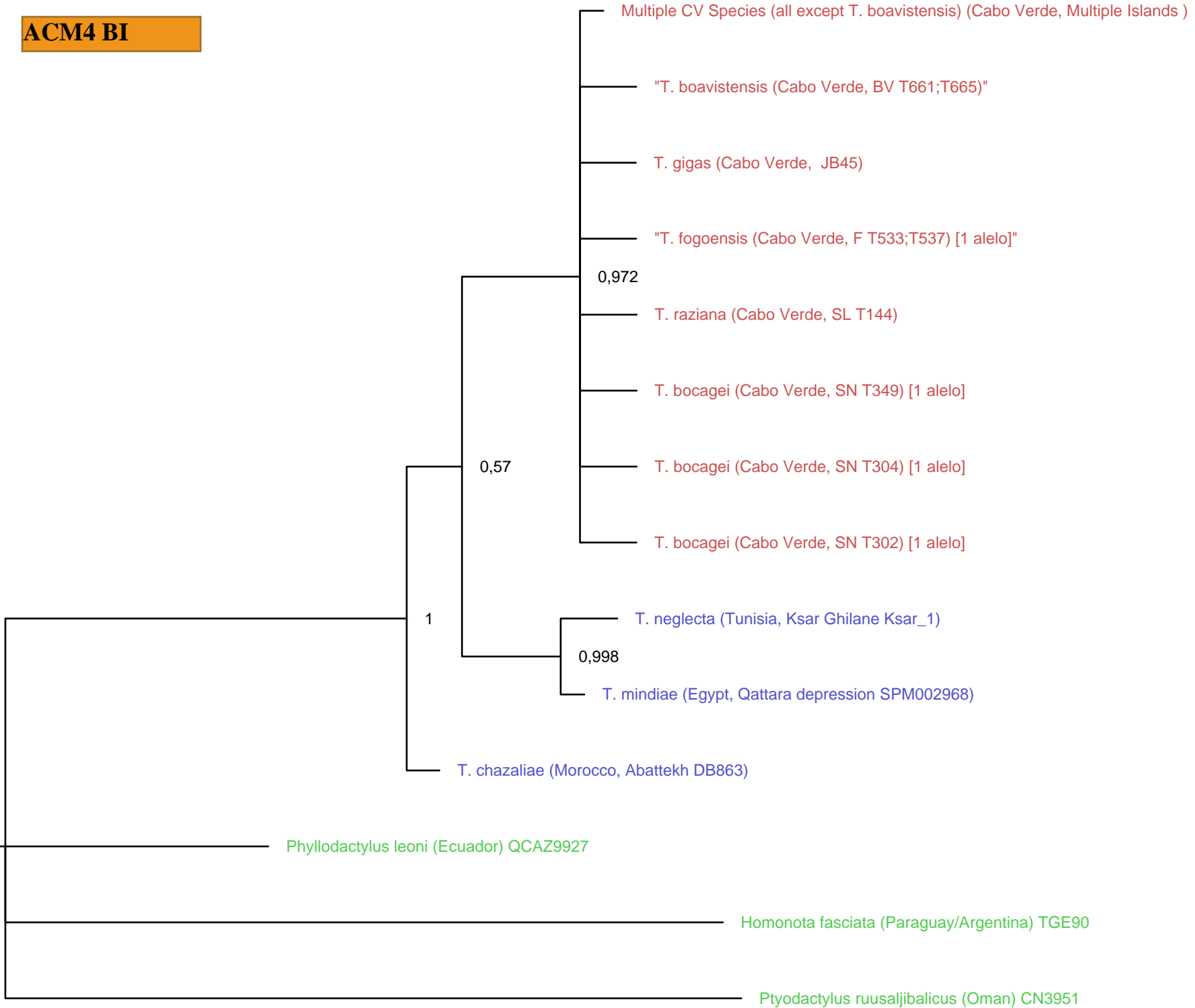


12S ML

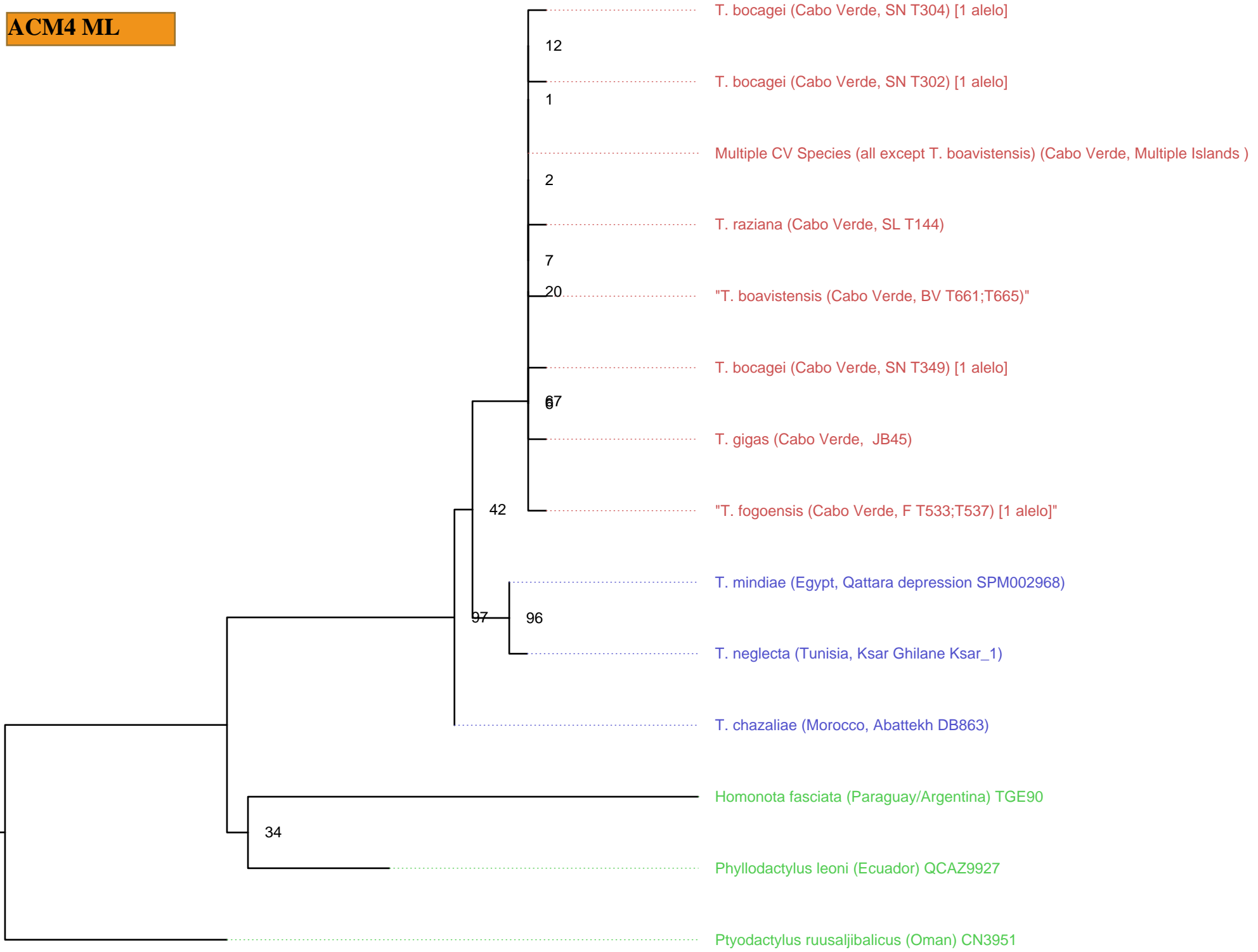


0.09

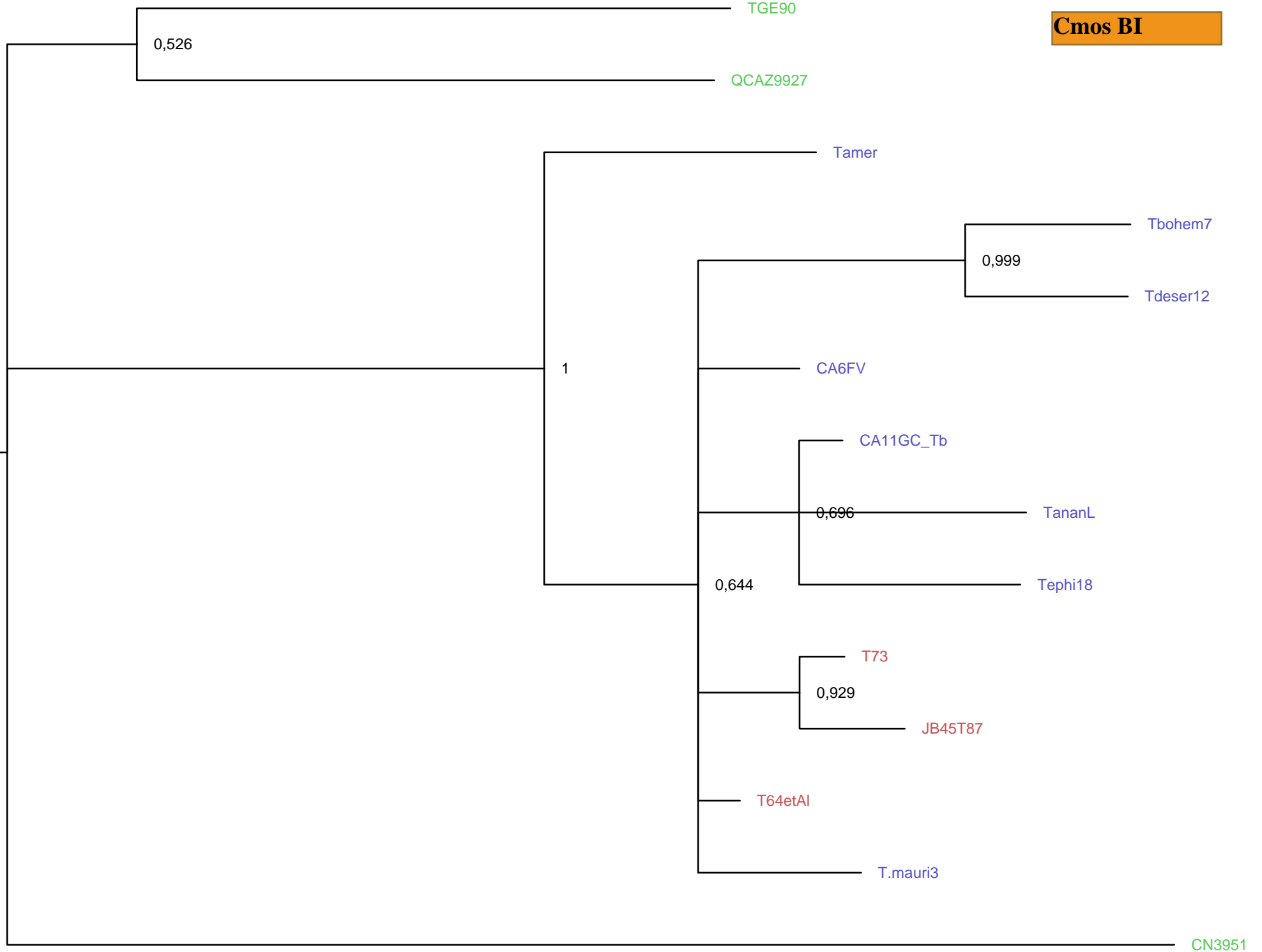
ACM4 BI



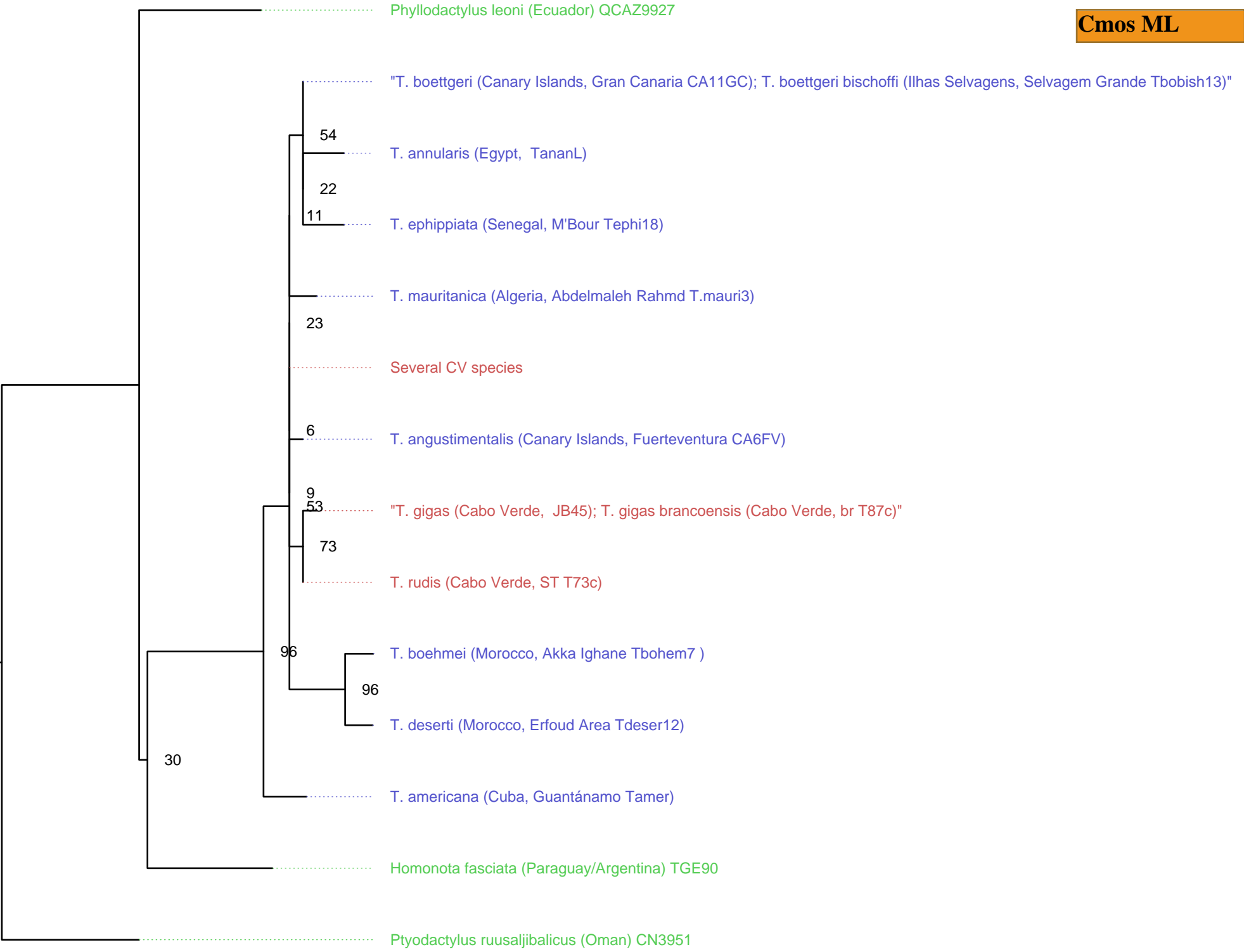
ACM4 ML



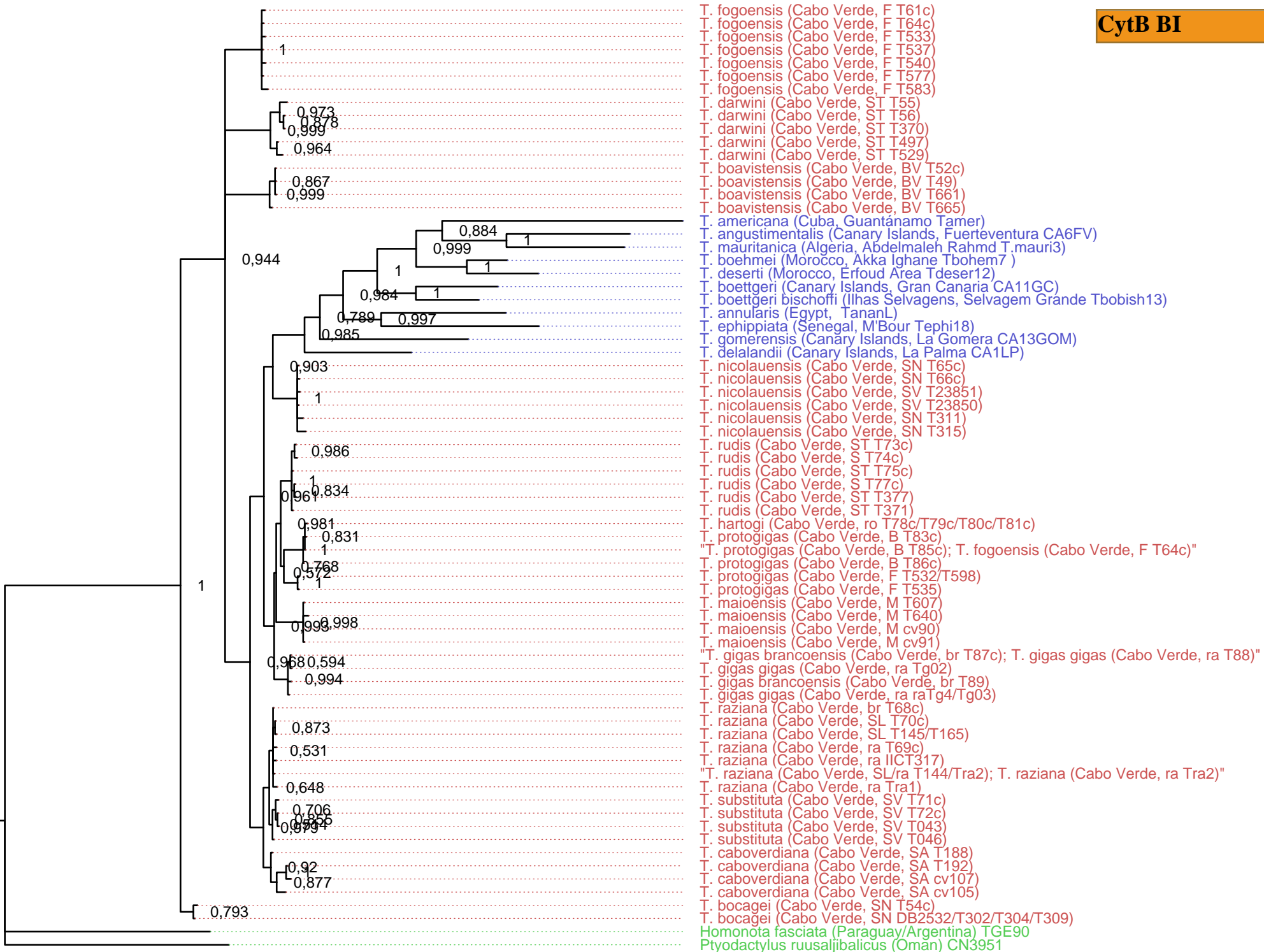
0.009



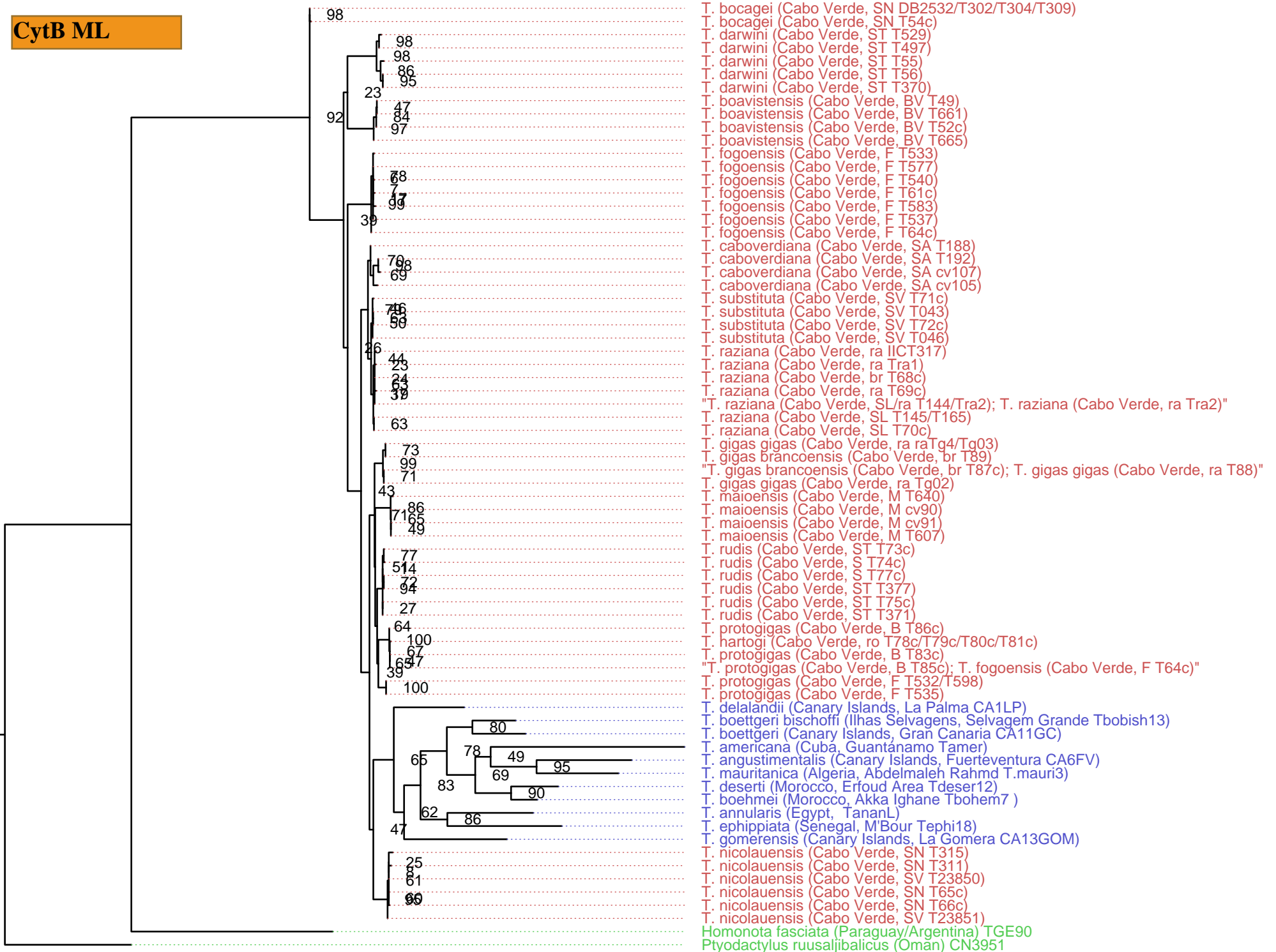
0.004



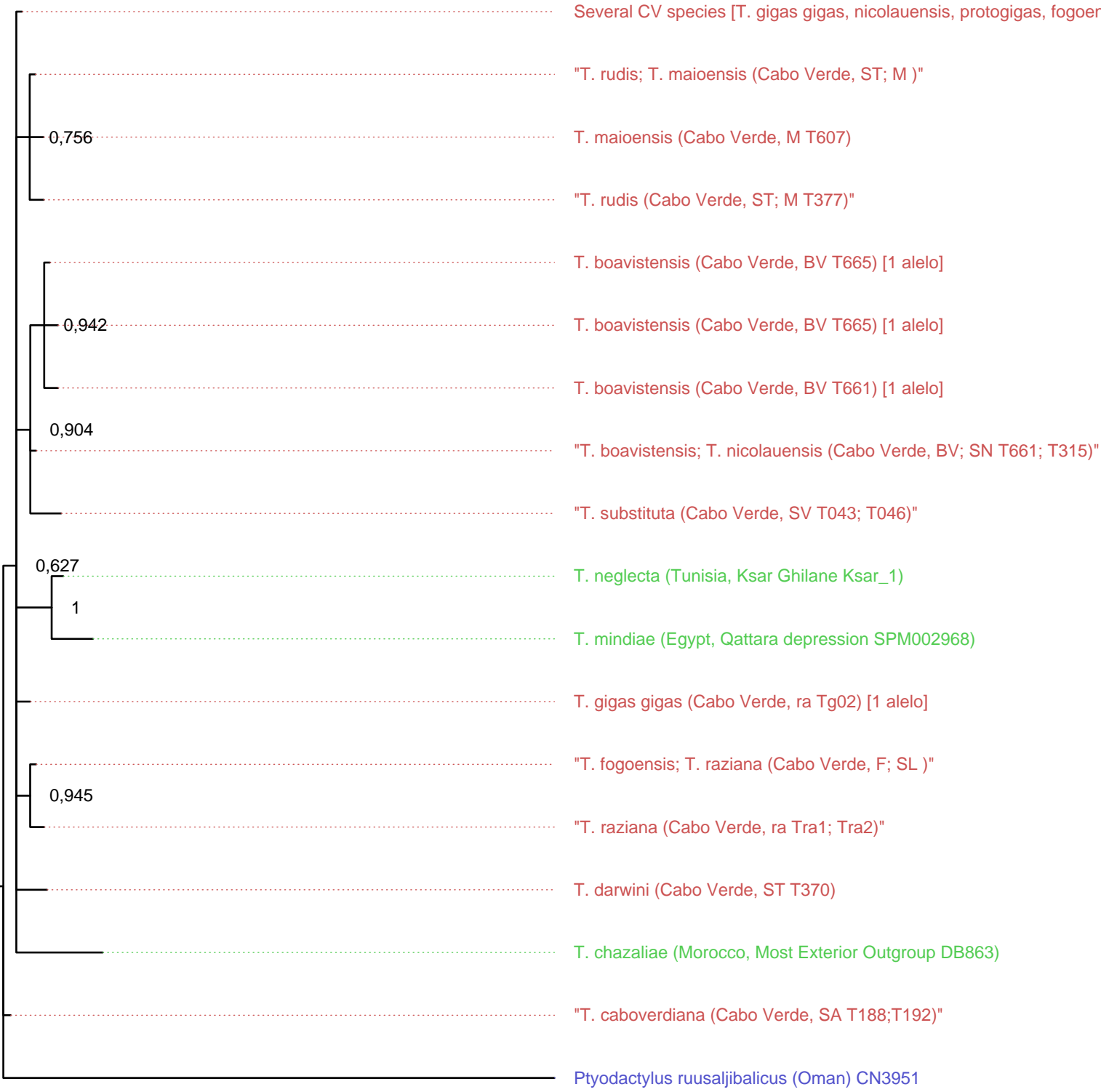
0.008



CytB ML



0.3



0.008

"*T. caboverdiana* (Cabo Verde, SA T188;T192)"

T. mindiae (Egypt, Qattara depression SPM002968)

91

T. neglecta (Tunisia, Ksar Ghilane Ksar_1)

31 *T. rudis* (Cabo Verde, ST; M T377)"

10

T. maioensis (Cabo Verde, M T607)

54

"*T. rudis*; *T. maioensis* (Cabo Verde, ST; M)"

T. gigas gigas (Cabo Verde, ra Tg02) [1 alelo]

T. boavistensis (Cabo Verde, BV T665) [1 alelo]

42

49 *T. boavistensis* (Cabo Verde, BV T661) [1 alelo]

71

24 *T. boavistensis* (Cabo Verde, BV T665) [1 alelo]

65 "*T. substituta* (Cabo Verde, SV T043; T046)"

2

"*T. boavistensis*; *T. nicolauensis* (Cabo Verde, BV; SN T661; T315)"

30 "*T. fogoensis*; *T. raziana* (Cabo Verde, F; SL)"

72

9 "*T. raziana* (Cabo Verde, ra Tra1; Tra2)"

15

T. darwini (Cabo Verde, ST T370)

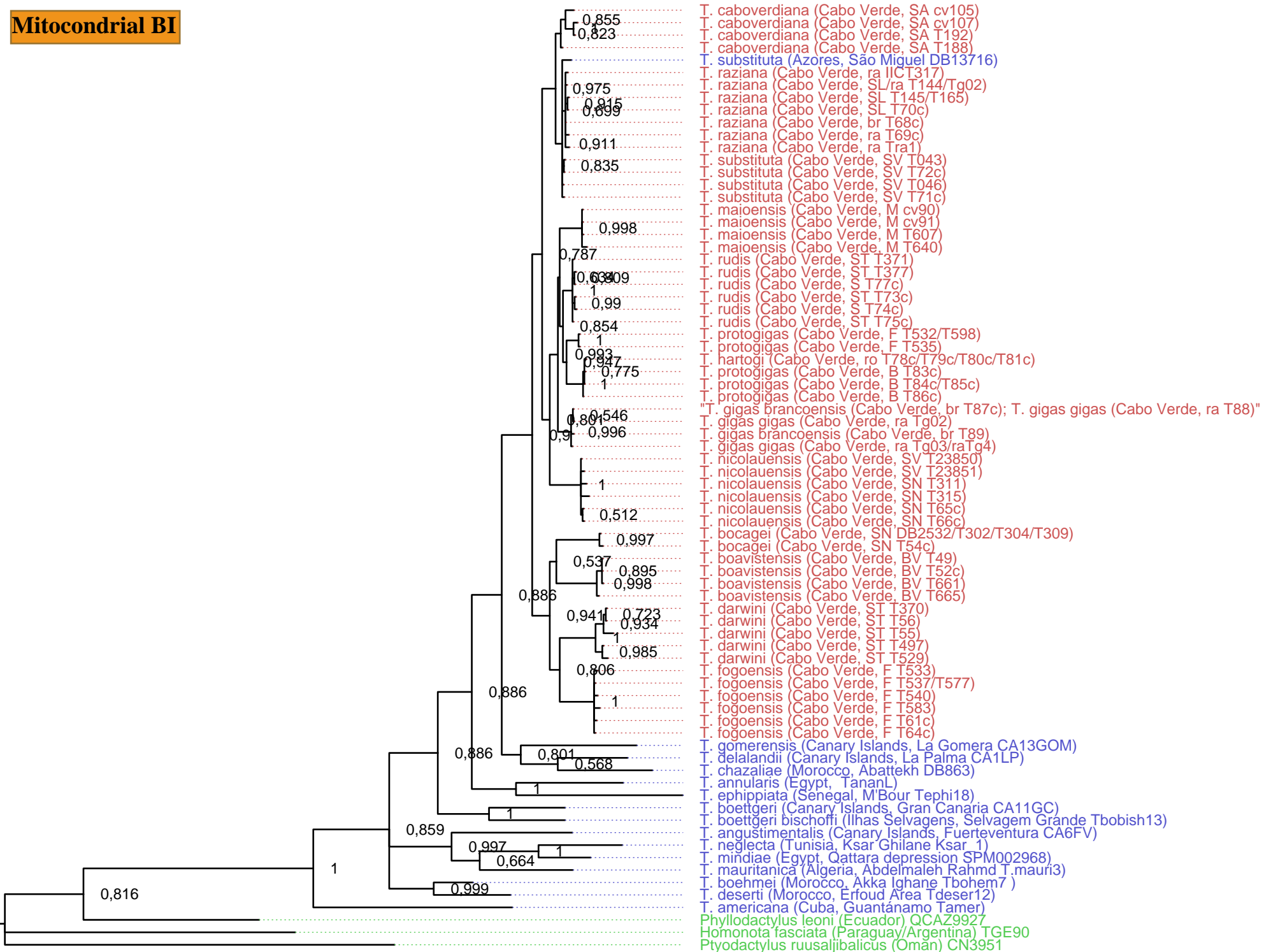
Several CV species [*T. gigas gigas*, *nicolauensis*, *protogigas*, *fogoensis* e *bocagei*] (Cabo Verde, SN, ra e F)

T. chazaliae (Morocco, Most Exterior Outgroup DB863)

Ptyodactylus ruusaljibalicus (Oman) CN3951

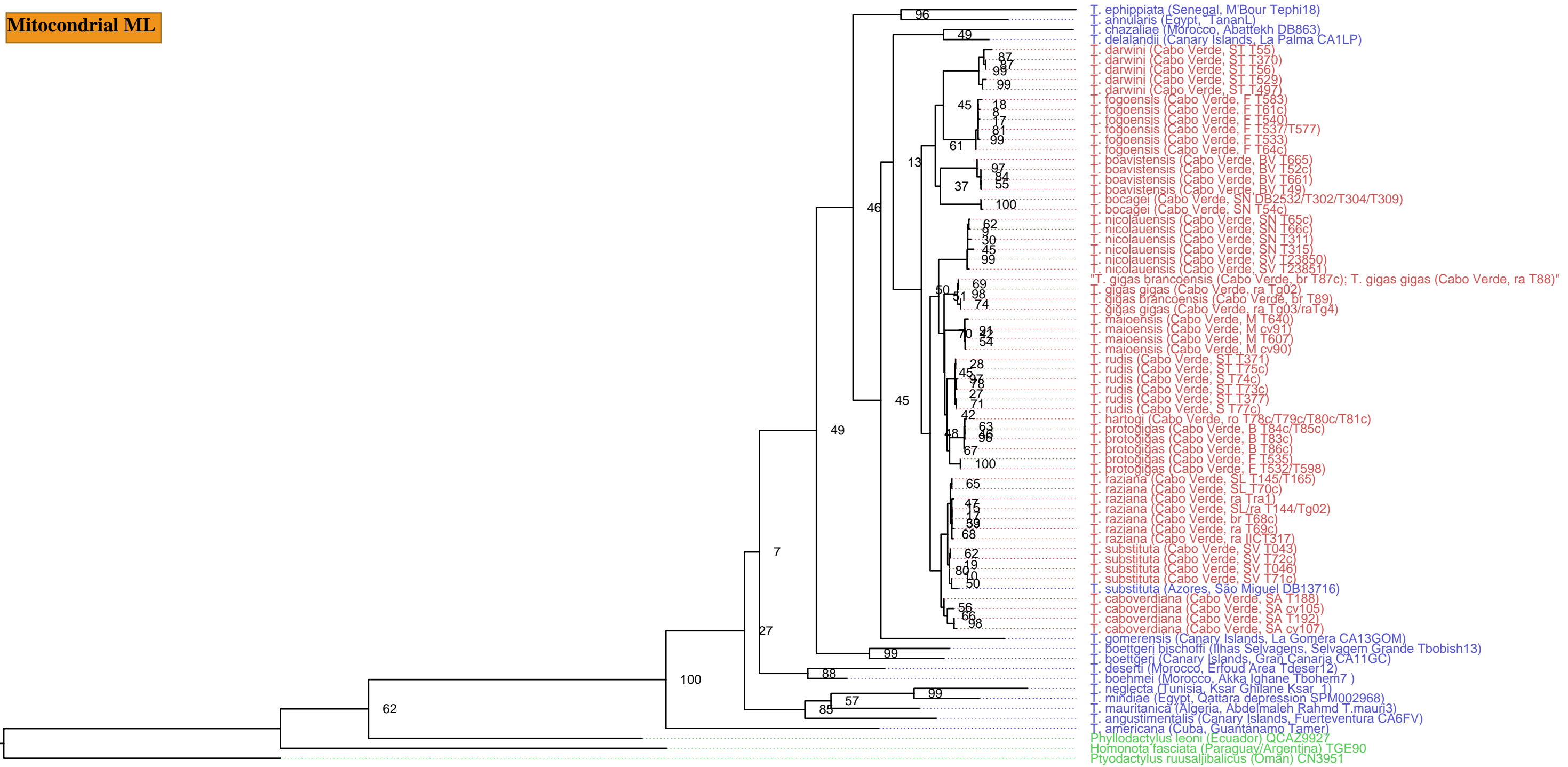
0.008

Mitochondrial BI



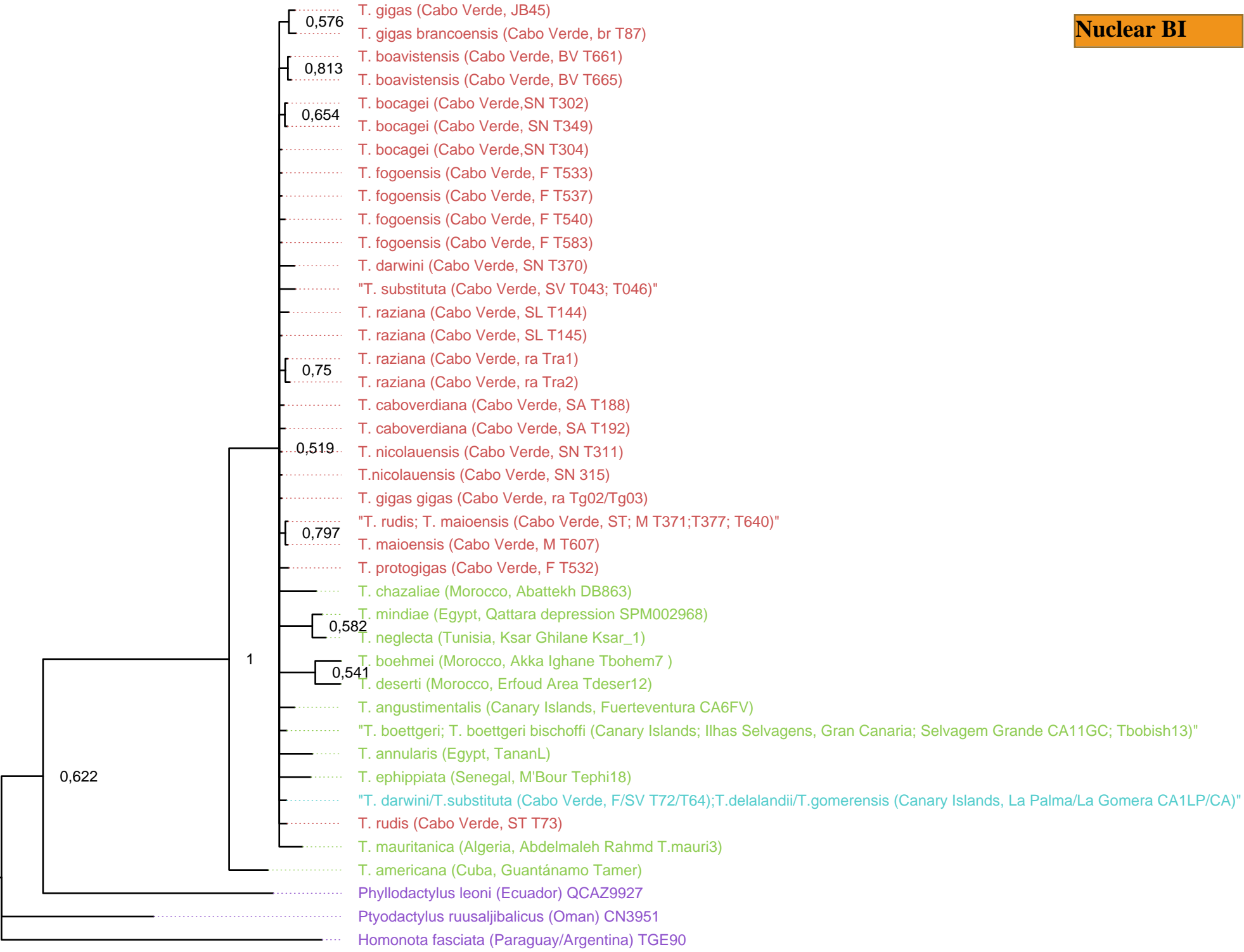
0.08

Mitochondrial ML



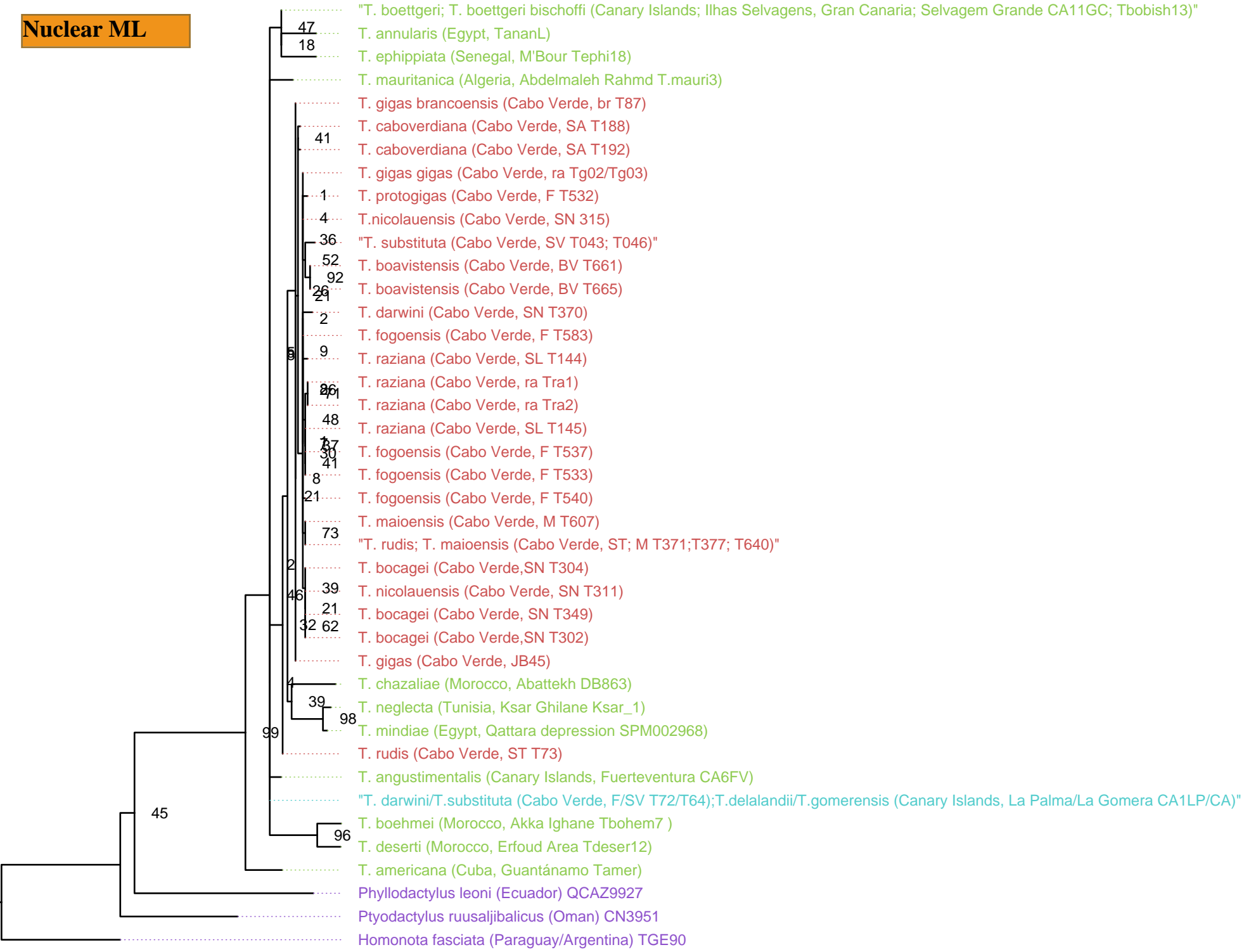
- T. ephippiata (Senegal, M'Bour Tephi18)
- T. annularis (Egypt, TananL)
- T. chazaliae (Morocco, Abattekh DB863)
- T. delalandii (Canary Islands, La Palma CA1LP)
- T. darwini (Cabo Verde, ST T55)
- T. darwini (Cabo Verde, ST T370)
- T. darwini (Cabo Verde, ST T36)
- T. darwini (Cabo Verde, ST T56)
- T. darwini (Cabo Verde, ST T529)
- T. darwini (Cabo Verde, ST T497)
- T. fogoensis (Cabo Verde, F T583)
- T. fogoensis (Cabo Verde, F T61c)
- T. fogoensis (Cabo Verde, F T540)
- T. fogoensis (Cabo Verde, F T537/T577)
- T. fogoensis (Cabo Verde, F T533)
- T. fogoensis (Cabo Verde, F T64c)
- T. boavistensis (Cabo Verde, BV T665)
- T. boavistensis (Cabo Verde, BV T52c)
- T. boavistensis (Cabo Verde, BV T661)
- T. boavistensis (Cabo Verde, BV T49)
- T. bocagei (Cabo Verde, SN DB2532/T302/T304/T309)
- T. bocagei (Cabo Verde, SN T54c)
- T. nicolauensis (Cabo Verde, SN T65c)
- T. nicolauensis (Cabo Verde, SN T66c)
- T. nicolauensis (Cabo Verde, SN T311)
- T. nicolauensis (Cabo Verde, SN T315)
- T. nicolauensis (Cabo Verde, SV T23850)
- T. nicolauensis (Cabo Verde, SV T23851)
- T. gigas brancoensis (Cabo Verde, br T87c); T. gigas gigas (Cabo Verde, ra T88)"
- T. gigas gigas (Cabo Verde, ra Tg02)
- T. gigas brancoensis (Cabo Verde, br T89)
- T. gigas gigas (Cabo Verde, ra Tg03/raTg4)
- T. maioensis (Cabo Verde, M T640)
- T. maioensis (Cabo Verde, M cv91)
- T. maioensis (Cabo Verde, M T607)
- T. maioensis (Cabo Verde, M cv90)
- T. rudis (Cabo Verde, ST T371)
- T. rudis (Cabo Verde, ST T75c)
- T. rudis (Cabo Verde, S T74c)
- T. rudis (Cabo Verde, ST T73c)
- T. rudis (Cabo Verde, ST T377)
- T. rudis (Cabo Verde, S T77c)
- T. hartogi (Cabo Verde, ro T78c/T79c/T80c/T81c)
- T. protogigas (Cabo Verde, B T84c/T85c)
- T. protogigas (Cabo Verde, B T83c)
- T. protogigas (Cabo Verde, B T86c)
- T. protogigas (Cabo Verde, F T535)
- T. protogigas (Cabo Verde, F T532/T598)
- T. raziana (Cabo Verde, SL T145/T165)
- T. raziana (Cabo Verde, SL T70c)
- T. raziana (Cabo Verde, ra Tra1)
- T. raziana (Cabo Verde, SL/ra T144/Tg02)
- T. raziana (Cabo Verde, br T68c)
- T. raziana (Cabo Verde, ra T69c)
- T. raziana (Cabo Verde, ra IICT317)
- T. substituta (Cabo Verde, SV T043)
- T. substituta (Cabo Verde, SV T72c)
- T. substituta (Cabo Verde, SV T046)
- T. substituta (Cabo Verde, SV T71c)
- T. substituta (Azores, São Miguel DB13716)
- T. caboverdiana (Cabo Verde, SA T188)
- T. caboverdiana (Cabo Verde, SA cv105)
- T. caboverdiana (Cabo Verde, SA T192)
- T. caboverdiana (Cabo Verde, SA cv107)
- T. gomerensis (Canary Islands, La Gomera CA13GOM)
- T. boettgeri bischoffi (Ilhas Selvagens, Selvagem Grande Tbobish13)
- T. boettgeri (Canary Islands, Gran Canaria CA11GC)
- T. deserti (Morocco, Erfoud Area Tdeser12)
- T. boehmei (Morocco, Akka Ighane Tbohem7)
- T. neglecta (Tunisia, Ksar Ghilane Ksar_1)
- T. mindiae (Egypt, Qattara depression SPM002968)
- T. mauritanica (Algeria, Abdelmaleh Rahmd T.mauri3)
- T. angustimentalis (Canary Islands, Fuerteventura CA6FV)
- T. americana (Cuba, Guantánamo Tamer)
- Phyllodactylus leoni (Ecuador) QCAZ9927
- Homonota fasciata (Paraguay/Argentina) TGE90
- Ptyodactylus ruusaljibalicus (Oman) CN3951

0.2



0.005

Nuclear ML



0.01