UNIVERSIDADE FEDERAL DE MINAS GERAIS
Pós-Graduação em Zoologia

The effects of polymorphic characters in the phylogeny of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* (Mammalia, Carnivora, Canidae, Cerdocyonina) based on morphological characters

Leila Alessandra Martins Birkenhead

2019

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Pós-Graduação em Zoologia

The effects of polymorphic characters in the phylogeny of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* (Mammalia, Carnivora, Canidae, Cerdocyonina) based on morphological characters

Leila Alessandra Martins Birkenhead

Master's dissertation presented to Programa de Pós-graduação em Zoologia, of Universidade Federal de Minas Gerais, as part of the requirements to receive the degree of Master in Zoology

Supervisor: Fernando Araújo Perini

2019

ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO

Leila Alessandra Martins Birkenhead

Ao vigésimo sexto dia do mês de abril do ano de dois mil e dezenove, às quatorze horas, na Universidade Federal de Minas Gerais, teve lugar a defesa de Mestrado da Pós-Graduação em Zoologia, de autoria da Mestranda Leila Alessandra Martins Birkenhead intitulada: **"The effects of morphological polymorphic characters in the phylogeny of South American foxes (Mammalia, Carnivora, Canidae, Cerdocyonina)"**. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Fernando Araújo Perini, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra para a candidata para apresentação de seu trabalho.

Esteve presente a Banca Examinadora composta pelos membros: Flávio Henrique Guimarães Rodrigues, Rodrigo Parisi Dutra, e demais convidados. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata.

Após a argüição, apenas os Srs. Examinadores permaneceram na sala para avaliação e deliberação acerca do resultado final, a saber: a dissertação foi:

- [x] Aprovada sem alterações
- [ ] Aprovada com alterações (observações em anexo)
- [ ] Reprovada

Nada mais havendo a tratar, o Presidente da Comisão encerrou a reunião e lavrou a presente ata, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 26 de abril de 2019.

Comissão Examinadora	Assinatura
Prof. Dr. Fernando Araújo Perini (Orientador)	
Prof. Dr. Flávio Henrique Guimarães Rodrigues	
Prof. Dr. Rodrigo Parisi Dutra	

Programa de Pós-graduação em Zoologia
Instituto de Ciências Biológicas
Universidade Federal de Minas Gerais

The effects of morphological polymorphic characters in the phylogeny of South American foxes (Mammalia, Carnivora, Canidae, Cerdocyonina)

AUTOR: LEILA ALESSANDRA MARTINS BIRKENHEAD

Esta dissertação foi apresentada em sessão pública e submetida a avaliação em 26 de abril de 2019, pela Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Fernando Araújo Perini (Orientador / UFMG)

Prof. Dr. Flávio Henrique Guimarães Rodrigues (Membro – UFMG)

Prof. Dr. Rodrigo Parisi Dutra (Membro – UNIBH)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgments

This study is the outcome of my fortune to have encountered so many people willing to help me in the best away each one could. I thank my supervisor Fernando Perini for welcoming me in the Laboratório de Evolução de Mamíferos, for all the knowledge he has shared with me, for trusting my skills to do this study and believing on my ability even when I had moments of doubts. Thanks to Daniel Casali for helping me with the many analyses and for reviewing the manuscript. Thanks to all lab partners, classmates and professors of Programa de Pós Graduação em Zoologia with whom I had classes. Thank you to all curators and staff members of the collections I have visited: Coleção de Mamíferos do Centro de Coleções Taxonômicas da UFMG, Coleção de Mamíferos do Museu de Ciências Naturais da PUC Minas, Mammalogy Department and Paleontology Department of the American Museum of Natural History and Collection of Mammals of the Field Museum of Natural History. Special thanks to my parents that have supported me in so many ways through this two year journey, words will never be enough to express how grateful I am to have both of them as guides. This research had the financial support of CAPES and PROAP.

Abstract

The phylogeny of Canidae has been inferred by many authors using different types of data. There is a consensus among molecular and combined phylogenies that the South American canids, a clade named Cerdocyonina, are monophyletic and divided in two major clades which are *Chrysocyon* + *Speothos* and *Cerdocyon* + *Atelocynus* + *Lycalopex*. The clade *Cerdocyon* + *Atelocynus* + *Lycalopex* has 8 species and the relationships among them are contradictory in both molecular and combined phylogenies. In the morphological phylogenies the incongruence within this clade is even greater. The presence of polymorphism in datasets, molecular and morphological, has instigated authors to create different coding methods. Since the report of polymorphism within Canidae is recent, different coding methods have not been tested with this data. This study aims to evaluate how different coding methods affects the phylogeny of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex*. Our morphological dataset was coded by six coding methods and we compared the results to a molecular tree we generated based in sequences downloaded from Genbank. We also proposed a new coding method named Frequency-as-continuous. We obtained the largest sampling for the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* ever assembled. The percentage of polymorphic entries in our dataset was higher than any other matrix. We did not recover a tree that had 100% similarity to the molecular tree. The topologies of the trees obtained by the matrices coded by the Frequency-bins and Frequency-as-continuous methods were similar. The trees which were most similar to the molecular tree were obtained from the matrices coded with the Polymorphic and the Frequency-bins methods respectively. Sample size affects the coding of polymorphism. We presumed the incongruence observed between phylogenies for the study group, which used morphological characters, could be linked to the high polymorphism present in the clade. We recommend using states frequencies as a guide to code any matrix because it reduces subjectiveness when coding.

Key-words: Phylogeny Canidae, Cerdocyonina, South American canids, morphology, polymorphism, coding methods, frequency, frequency-as-continuos.

1 Introduction

The Canidae family (Mammalia, Carnivora) first appeared in North America in the Late Eocene (40 Ma). The first subfamily to diverge was Hesperocyoninae, which was extinct by the middle Miocene (15 Ma). The other two subfamilies, Borophaginae and Caninae, were first recorded at about the same time, in the early Oligocene (32 Ma) (Wang e Tedford, 2008). However, Borophaginae disappeared in late Pliocene (c.a. 2.5 Ma), with only Caninae persisting to present time, with 35 living species (Wozencraft, 2005; Tedford et al., 2009).

Canids are the most widely distributed clade within Carnivora (Wang and Tedford, 2008). Unlike the other clades of Canidae, which were confined to North America, Caninae has dispersed globally. It invaded Asia through the Beringian land bridge in the Pliocene (5 to 4 Ma) and South America when the Isthmus of Panama emerged in the late Cenozoic (3 Ma) (Wang e Tedford, 2008). The expansion of territory during the Pliocene and Pleistocene resulted in faunas partially endemic in Africa, Eurasia and South America (Tedford, 2009).

Even though the invasion of South America (SA) is relatively recent (3 Ma), this continent harbors the highest diversity of canids nowadays (Wang and Tedford, 2008). Twelve species occur in SA and only *Urocyon cineoreoargenteus* (Schreber, 1775) is not endemic (Berta, 1987). The canids endemic to SA are grouped in the subtribe Cerdocyonina (Wang e Tedford, 2008; Tedford, 2009), which today includes the following species: *Chrysocyon brachyurus* (Illiger, 1815), *Speothos venaticus* (Lund, 1842), *Cerdocyon thous* (Linnaeus, 1766), *Atelocynus microtis* (Sclater, 1883), *Lycalopex vetulus* (Lund, 1842), *L. gymnocercus* (Fischer, 1814), *L. griseus* (Gray, 1837), *L. culpaeus* (Molina, 1782), *L. fulvipes* (Martin, 1837), *L. sechurae* (Thomas, 1900) and *Dusicyon australis* (Kerr, 1792). The later considered extinct since 1880 (Berta, 1987).

Phylogenies based in molecular and combined data have strongly supported the monophyly of Cerdocyonina (Wayne et al., 1997; Bardeleben et al., 2005; Lindblad-Toh et al., 2005; Perini et al., 2009; Prevosti, 2010; Zrzavy et al., 2018). Another consensus among studies is the division of Cerdocyonina in two clades: *Chrysocyon* + *Speothos* and *Cerdocyon* + *Atelocynus* + *Lycalopex* (Wayne et al., 1997; Lindblad-Toh et al., 2005; Perini et al., 2009; Prevosti, 2010; Zrzavy et al. 2018). The clade *Cerdocyon* + *Atelocynus* + *Lycalopex* includes eight species, with *Cerdocyon* and *Atelocynus* being monotypic, the remaining species belong to the genus *Lycalopex*.

The morphological phylogeny of Berta (1987) was done at genus level and recovered *Lycalopex* as monophyletic. Tedford et al. (1995) did not recover *Lycalopex* as a monophyletic clade because *Lycalopex vetulus* is recovered within a polytomy with *Chrysocyon* + *Cerdocyon* + *Nyctereutes* + *Atelocynus* + *Speothos*. The molecular phylogenies recover *Lycalopex* monophyly (Bardeleben et al., 2005 and Lindblad-Toh et al., 2005). There are combined phylogenies that recovered the clade as monophyletic and others in which the species *Dusicyon australis* and/or *Dusicyon avus* are found within the clade (Wayne et al., 1997; Perini et al., 2009; Prevosti, 2010 and Zrzavy et al., 2018). The relationships among species of *Lycalopex* are still contentious. Zrzavy et al. (2018) phylogeny it's not fully resolved and some phylogenies do not include all species of *Lycalopex* (Wayne et al., 1997 and Bardeleben et al., 2005).

A few studies recovered the species *L. sechurae* as the first lineage to diverge within the genus (Wayne et al., 1997; Bardeleben et al., 2005; Perini et al., 2009; Prevosti, 2010). In Lindblad-Toh (2005) and Tchaicka et al. (2016) phylogenies, the *L. vetulus* lineage would have been the first to diverge. When present in the study, *L. culpaeus* is always recovered as the last lineage to diverge (Lindblad-Toh et al., 2005; Wayne et al., 1997; Perini et al., 2009; Prevosti, 2010; Tchaicka et al., 2016 and Zrzavy et al., 2018), but which lineage it is sister of varies between authors. When present, *L. gymnocercus* and *L. griseus* are recovered forming a sister clade to other species or diverging one after the other but there is no consensus about which one would have diverged first. Most studies did not sample *L. fulvipes* and there is no consensus of its position in the few phylogenies in which the species is present. In most studies, support values within *Lycalopex* are weak (Lindblad-Toh et al., 2005; Perini et al., 2009; Prevosti, 2010). This could partially explain the lack of consensus between papers. Usually, authors who obtained high support values did not sample all the species within the clade (Wayne et al., 1997; Bardeleben et al., 2005), while other studies did not present the support values for internal relations within *Lycalopex* (Zrzavy et al., 2018). Most phylogenies based on morphology also didn't estimate support values (Berta, 1987 and Tedford et al., 1995). Even though Tedford et al. (2009) estimated support values for its tree, it wasn't measured inside the clade Cerdocyonina.

The position of *Cerdocyon thous* and *Atelocynus microtis* varies. There are studies which recover *A. microtis* as the first clade to diverge (Lindblad-Toh et al., 2005; Prevosti, 2010; Zrzavy et al., 2018), while other analyses suggest that *A. microtis* and *C. thous* form a monophyletic clade sister of *Lycalopex* (Wayne et al., 1997; Perini et al.,

2009), with high and low support values presented for both hypotheses. In the morphological phylogenies *C. thous* and *A. microtis* position are much different. *Cerdocyon thous* forms a clade with the Asian species *Nyctereutes procyonoides*, and *Atelocynus microtis* forms a clade with *Speothos venaticus* (Berta, 1987 and Tedford, 1995).

Berta (1987), Tedford et al. (1995) and Prevosti (2010) used morphological characters extensively to build their phylogenetic proposals. Other phylogenies that used morphological characters usually used at least one of these authors data to make their inferences (e.g. Perini et al., 2009; Zrzavy et al., 2018). However, only in Prevosti's (2010) publication polymorphic characters were recognized and coded into the matrix.

1.1 Polymorphism

The recognition that different individuals of the same species present different degrees of variation in their characteristics is not new. In the 19th century Charles Darwin (1859) wrote that it should not be assumed that all individuals of the same species are identical, which is one of the fundamental tenets of his theory. He already noted that, at the time, most taxonomists did not admit that such variations can occur in characters considered important. Given the advent of cladistics and phylogenetic systematics it is natural to think that intraspecific variation would be an important topic of discussion. However, the literature about intraspecific variation or polymorphic characters associated with morphological phylogenies is limited.

One of the approaches into this topic is a series of papers by Dr. John J. Wiens (Wiens, 1995, 1998, 1999, 2000, Wiens and Servedio 1997, 1998). Herein we have adopted Wiens (1999) definition of polymorphism: it is the variation within species which is independent of ontogeny and sex. The variation is genetically based and heritable. In 2000, Poe and Wiens did an extensive search on the literature about how morphologists choose characters for phylogenetic analyses. The results obtained explained, in some ways, why intraspecific variation/polymorphic characters have not been a topic of discussion. Criteria for character exclusion are rarely mentioned in the literature and in the papers which inform a criterion, the most common reason for exclusion is the variation observed within terminal taxa, especially when it is intraspecific. The failure of authors in presenting an outline for the exclusion of characters hampers analyses of polymorphic characters. It also raises questions such as whether or not polymorphism is rare, if researchers are removing them from matrixes or if it is not being reported deliberately.

The use of polymorphic characters in phylogenetics is considered contentious. Nixon and Wheeler (1990) do not acknowledge attributes which are polymorphic as characters, and designate them as traits. For these authors, since traits are not constant, they do not fit the concept of phylogenetic informative characters (Nixon and Wheeler, 1990). Nixon and Wheeler (1990) statement is based in theoretical principals and not in pragmatic observation. On the other hand, Campbell and Frost (1993) assert that systematists work primarily with fixed characters, not because its always more informative, but because it is easier to deal with them than to code ambiguity, do various analytical procedures and/or having to discriminate the source of variations. As mentioned before, the lack of information of the criterion for excluding characters limits polymorphism analyses. It also constrains studies about whether or not in practice polymorphism can contribute to phylogenetic inferences.

Even though there is a stigma towards polymorphic characters, there are methods available to analyze such data (Wiens, 1995), which differ in underlying assumptions of coding (Campbell and Frost, 1993). These methods are fixed characters, Polymorphic, Any instance, Unscaled, Scaled (Campbell and Frost, 1993), Frequency (Swofford and Berloch, 1987) and Majority (Wiens et al. 2000) (explained in Material and Methods). As stated by Campbell and Frost (1993) each method has its drawbacks but these methods have been put to test and it showed that intraspecfic variable characters have evolutionary evidence.

A pragmatic approach comparing different coding methods to infer phylogenies, using morphological data, was done by Campbell and Frost (1993), Wiens and Servedio (1997,1998) and in various works of Wiens (1995, 1998, 1999, 2000). As predicted by Kluge and Farris (1969), characters which vary intraspecifically have a positive relation with homoplasy (Campbell and Frost, 1993; Wiens, 1995). In spite of homoplasy correlation, when comparing the results obtained by matrices that had only fixed characters with matrices that also included polymorphic characters, the fixed data only approach was less efficient than all polymorphic coding methods. This suggests that polymorphic data have a significant phylogenetic signal, which is supported by simulations and congruence analysis. (Campbell and Frost, 1993; Wiens and Servedio, 1997, 1998; Wiens 1995, 1998).

Canidae phylogeny has been inferred by many authors using various type of data but there are still many incongruences among published trees, specially within the South American canids. Most studies which used morphological data do not have a single polymorphic state in their matrix. In 2010, Prevosti published a matrix which presented

polymorphic morphological characters in canids but he only used the Polymorphic coding method on his data. Therefore, in this paper we explored how different coding methods can affect the phylogenetic relationships of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex*.

2 Materials and Methods

The data used in this study was obtained in the following collections: Coleção de Mamíferos do Centro de Coleções Taxonômicas da UFMG (CCT-UFMG), Coleção de Mamíferos do Museu de Ciências Naturais da PUC Minas (CMMCN-PUC), Mammalogy Department (AMNH) and Paleontology Department (AMNH-P) of the American Museum of Natural History and Collection of Mammals of the Field Museum of Natural History (FMNH). A list of specimens analyzed is presented in Appendix 1.

The species studied were *Atelocynus microtis, Cerdocyon thous, Lycalopex culpaeus, L. fulvipes, L. griseus, L. gymnocercus, L. sechurae and L. vetulus*. We chose two species as outgroup, *Hesperocyon gregarius*, which belongs to Hesperocyoninae an extinct subfamily of Canidae, and *Urocyon cinereoargentus,* since it has been recovered in multiple studies as one of the earliest divergence within extant Caninae (e.g. Lindblad-Toh et al., 2005; Perini et al., 2009).

We reviewed morphological characters from previous studies, related to discrete morphological skull variation, mandible and teeth. Some previous descriptions were left as in the original while others were modified (Berta, 1987; Tedford et al., 1995; Prevosti, 2010) (For more information see section 3.1 Characters descriptions). The nomenclature was based in the works of Evans and Lahunta (2012), and Wible and Spaulding (2013). We analysed at least one female and one male of every species, except for *Hesperocyon gregarius* since specimens did not have the gender assigned. For better sampling of species within their geographic distribution, when the number of specimens was large enough, we sampled at least one specimen of all localities present in the collections and whenever possible one female and one male for each locality. Given that species can have a wide or restricted geographical distribution, and we work with museum specimens, different localities could have been different areas, cities, states or countries (Appendix 1). All specimens observed were adults and only characters that were observed in both genders were coded (Wiens, 2000).

We used six coding methods for polymorphic characters, which were: Polymorphic, Majority, Unordered, Unscaled, Frequency-bins and Frequency-as-continuous. The frequencies observed are presented in Appendix 2. The Polymorphic method codes all character states observed as they are, keeping the polymorphisms in the matrix cells (Campbell and Frost, 1993). In the Majority method the state coded in the matrix is the most frequent, and if there is a tie the tied states are coded as a separate state (Johnson, Zink and Marten, 1988). In the Unorder method, characters are coded as absent (0), polymorphic (1) and fixed (2), all character states unordered. In this method

it is assumed that all characters must have a polymorphic state even if it was not observed in the studied sample (Wiens, 2000). In the Unscaled method, characters are coded as absent (0), polymorphic (1) and fixed (2), all character states are ordered and characters states don't have to necessarily go through a polymorphic state. In the Frequency-bins method, one of the states is chosen as reference. In this study we chose state 1. After that, the frequencies observed in state 1 are recoded into new states which are the bins (e.g: state 1 = 0-10%, state 2 = 11-20%); there is no established rule to determinate the ranges of the bins. This method it is designed for binary characters only (Wiens, 1993 and Wiens, 1995).

We also proposed a frequency coding method in which the numerical values representing the frequency of the state of a binary character (here, arbitrarily defined as state 1) are used as a continuous variable. It follows the same theoretical premises of the method of Frequency-bins, namely, that frequencies are inheritable and sufficiently stable to fulfill the role of a phylogenetics character, but without the somewhat artificial delimitation of bins. The observed frequencies values are optimized as continuous characters, which are necessarily ordered. To the best of our knowledge, this is a new approach, not implemented before in previous phylogenetics studies which included polymorphic data.

Character coding was done in MorphoBank (O'Leary and Kaufman 2011, 2012) and Mesquite (Maddison and Maddison, 2018). Maximum-parsimony analyses of matrices were done in TNT 1.5, with implicit enumeration search using equal and implied weights (k=3) (Goloboff, Farris and Nixon, 2008). Tree consensus was calculated by strict (=Nelson) method. Support was measured by Standard and Poisson Bootstrap, to equal and implied weight analyses, respectively. In all analysis, trees were rooted in *Hesperocyon gregarius*. To test the correlation between the percentage of cells containing polymorphic entries for each taxa and the number of specimens observed for each, we used the non-parametric, Spearman's rank correlation coeficient (Spearman, 1904) on the matrix coded by the Polymorphic method. Tests were conducted in RStudio (RStudioTeam, 2016), R version 3.4.2 (R Core Team, 2017), using the ggpubr package (Kassambara, 2018).

We acknowledge that the true phylogeny cannot be known, but we needed a phylogeny as a parameter to compare our results. Since there are many disagreements between published hypotheses for our study group, we decided to run a new molecular tree with data downloaded from Genbank. We elected a molecular hypothesis as our default tree, since presumable it was obtained through independent data from

morphological datasets. The molecular tree was based on 25 molecular makers (22 nuclear, 3 mitochondrial). The sequences were obtained for 8 ingroup and 21 outgroup species. We aligned each gene with MUSCLE algorithm (Edgar, 2004) in AliView 1.18.1 (Larsson, 2014) and manually edited the sequences whenever necessary. Individual alignments were concatenated into a complete dataset with 16630bp at FASconCAT-G (Kück and Longo, 2014)

The best-fit models of nucleotide substitution and best partition schemes were estimated with Partition Finder 2.1.1 (Lanfear et al., 2017), using Bayesian Information Criterion (BIC) and greedy algorithm, with linked branch lengths. For coding sequences each codon position was evaluated as a potential partition, and we consider exon and intron regions as potential different partitions.

Bayesian phylogenetics analyses were performed in MrBayes 3.2.6 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) on CIPRES (Miller et al., 2010), not using Beagle library. Two independent runs were set, with four chains each, sampling 10.000 trees per run. The convergence was checked monitoring the standard deviation of split frequencies (< 0.01) in MrBayes and ESS (>200) and trace plots in Tracer 1.7 (Rambaut et al., 2018). Support was accessed with Posterior Probabilities (PP) exhibit in a majority-rule consensus tree (Huelsenbeck et al., 2001), visualized in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

To evaluate the performance of the coding methods, we estimated the SPR-distance metric of the trees found by each in relation to the well-supported molecular tree. The matrices which produced the two most similar trees to the molecular topology were combined, separately, with Prevosti's (2010) morphological characters. Prevosti's matrix (2010) and the combined matrices were then compared to the molecular topology with SPR-distances, in order to evaluate a possible improvement in similarity among morphological and molecular topologies after adding characters of the literature to our original datasets. To compare the results of the combined matrices we also analyzed Prevosti's (2010) data set, excluding DNA nucleotides sequences, for our ingroup because his consensus tree of the osteological data included taxa that we did not observed and to compare tree topologies in SPR taxa sampling has to be the same.

We also chose to combine our results with Prevosti's (2010) morphological data because he was the first to acknowledge polymorphism in South American canids and gathered a rich amount of morphological characters, proposing one of the most

comprehensive and recent morphological proposals for the group. All polymorphisms in Prevosti (2010) were coded by the Polymorphic method. The following characters didn't vary in the ingroup and were excluded: 18, 20, 25, 27, 29, 31, 38, 43, 46, 47, 49, 65, 67, 69, 71, 74, 75, 76, 80, 84, 88, 100, 104, 126, 130, 134, 135, 138, 140, 146, 149, 150, 153, 155, 157, 158, 170, 171, 172, 173, 178, 179, 181, 183-191, 194, 195, 196, 199, 200, 206, 211, 214-218, 220, 221, 222, 223, 225, 231, 232, 233, 234, 236, 237. Characters that were the same as ours were also taken out of the matrix: 27, 26, 34, 32, 32, 30, 48, 50, 52, 57, 58, 59, 79, and 83.

3 Results

In total 158 specimens were observed (Table 1). After direct observation of specimens and revision of the literature, we were able to formalize 32 cranio-dental characters (Table 2). We observed polymorphism in 30 characters, of which 15 were characters already described in literature. All characters were coded as binary.

Table 1 - Number of specimens examined per species.

Species	Sample size
Atelocynus microtis	16
Cerdocyon thous	27
Lycalopex culpaeus	15
Lycalopex fulvipes	2
Lycalopex griseus	21
Lycalopex gymnocercus	23
Lycalopex sechurae	18
Lycalopex vetulus	18
Urocyon cinereoargenteus	28
Hesperocyon gregarius	8

Table 2 - Characters and characters states.

Characater	State 0	State 1
1. Nasal-frontal suture level related to the maxilla-frontal suture	Ends anteriorly or at the same level of maxilla-frontal suture to the maxilla-frontal suture	Extends posteriorly to the maxilla-frontal suture.
2. Eversion of jugal dorsal border	Absent	Present
3. Postorbital constriction	Immediately posterior to the post orbital process	Dislocated posteriorly approaching the frontal-parietal suture

Characater	State 0	State 1
4.Anteriormost portion of sagittal crest	Just before or after the frontral-parietal suture	Just before or after the parietal-interparietal suture
5. Profile shape of dorsal margin of neurocranium	Straight	Convex
6. Antero lateral process of nasal (lateral view)	Absent	Presen
7. Zygomatic process of maxilla	Absent	Present
8. Crest between the lacrimal foramen and the maxillary foramen	Absent	Present
9. Lateral expansion of dorsal and ventral areas of the mastoid process	Same length	Dorsal expands more
10. Infraorbital foramen shape	Round	Tear-shape or oval
11. Foramen of the postorbital process of the frontal	Slit like	Round
12. Retroarticular foramen	The foramen is near the postglenoid process	The foramen is near the distal portion of the squamosal bone
13. Distal margin of palatine	Ends anteriorly or on the same line as the distal margin of the 2nd upper molar	Ends posteriorly to the distal margin of the 2nd upper molar
14. Basioccipital bulge	Absent	Present
15. Bulla mesial posterior expansion (in relation to the jugular foramen)	Absent, doesn't cover the jugular foramen in ventral view	Present, partially covers the jugular foramen in ventral view
16. Auditory meatus tube	Absent	Present
17. Shape of the ventral margin of the mandible	Straight	Curved
18. Size of the angular process	Larger antero-posteriorlly than dorso-ventrally	Larger dorso-ventrally than antero-posteriorlly
19. Position of the tip of the angular process	Same level as the dorsal margin of the process	Higher than the dorsal margin of the process

Characater	State 0	State 1
20. Position of condylar process	On the same line as the margin that connects it to the angular process	It projects posteriorly in relation to the margin that connects it to the angular process
21. Shape of the anterior margin of coronoid process	Straight	Curved
22. Shape of the dorsal margin of coronoid process	Straight	Curved
23. Subangular lobe	Absent	Present
24. i1 distal cusplets	Absent	Present
25. i2 distal cusplets	Absent	Present
26. i3 distal cusplets	Absent	Present
27. I1 mesial cusplets	Absent	Present
28. I1 distal cusplets	Absent	Present
29. I2 mesial cusplets	Absent	Present
30. I2 distal cusplets	Absent	Present
31. Distal cusplets of p3	Absent	Present
32. m1 protostylid	Absent	Present

3.1 Characters descriptions

In this section we discuss each character and previous descriptions by other authors if it is the case. Characters illustrations are provided in Appendix 3.

1. Nasal-frontal suture level related to the maxilla-frontal suture: (0) Ends anteriorly or at the same level of maxilla-frontal suture; (1) Extends posteriorly to the maxilla-frontal suture.

The suture between the nasals and frontal bones could coalesce anteriorly/same level of the suture between the frontal and the maxilla. Alternatively, the suture between the frontal and maxilla can be posterior to the end of the suture between the frontal and the nasal. Berta (1987) and Tedford et al. (1995) described this character with two states, but Prevosti (2010) considered three states. The additional state of Prevosti (2010), "at the level of the suture", wasn't coded alone for any of the species that we observed and for the frequencies analyzes we have done characters with binary

states are mandatory. For such reason we chose to use two states. Making our description similar to Berta (1987) and Tedford (1995) which use words such as “rarely” and “usually”, these makes us believe that they also observed the character state "at the same length" but choose not to make it as a separate state. This is a modification of characters 27 of Berta (1987), no. 19 of Tedford et al. (1995) and no. 38 of Prevosti (2010).

2. Eversion of jugal dorsal border: (0) Absent; (1) Present.

The superior margin of jugal could projects outwards forming a slender dorsal margin. This is a modification of characters no. 35 of Tedford (1995), no. 7 of Wang et al. (1999) and no. 26 of Prevosti (2010).

3. Postorbital constriction: (0) immediately posterior to the post orbital process; (1) Dislocated posteriorly approaching the frontal-parietal suture

The postorbital constriction can occur immediately after the post-orbital processes or posteriorly, closer to the suture between the frontal and parietal bones. This is a modification of characters no. 34 of Prevosti (2010).

4. Anterior most portion of sagittal crest: (0) Just before or after the frontal-parietal suture; (1) Just before or after the parietal-interparietal suture.

The parasagittal crests usually coalesce into a central sagittal crest in many canid species. The beginning of the sagittal crest could be very anterior, close to the suture between the frontal and the parietal, or dislodged posteriorly, closer to the suture between the parietal and interparietal bones. This character is similar to Prevosti’s (2010) character no. 35 because he also delimits the location of the structure related to sutures. His description is more extensive than ours and that’s why we presume that Prevosti’s (2010) considers character 9 and 10 of Wang et al. (1999) to be similar to his description. We do not consider Wang’s et al. (1999) to be similar to our description, since in this work the location of the crest isn’t defined. On the other hand, we consider that Prevosti’s character is closer to ours. We observed sexual dimorphism for this character in *Lycalopex culpaeus*. For such reason the character wasn't scored on the matrix for this species. The sagittal crest of the females of *L. culpaeus* begins around

parietal-interparietal suture and males around the frontal-parietal suture. Contrary to Prevosti (2010), we didn't observe a correlation between development of the crest and sex in most species besides *Lycalopex culpaeus*. This is a modification of characters no. 35 of Prevosti (2010).

5. Profile shape of dorsal margin of neurocranium: (0) Straight; (1) Convex.

When seen in lateral view, skulls have a nearly straight dorsal profile of the neurocranium. In contrast, some skulls have a more convex profile, giving a more rounded appearance to it. This is a new character from this study.

6. Antero lateral process of nasal (lateral view): (0) Absent; (1) Present.

At the anterior lateral end of the nasal there is a process that projects anteriorly, which can be viewed clearly in lateral view. This is a new character from this study.

7. Zygomatic process of maxilla: (0) Absent; (1) Present.

In the first state the zygomatic process of maxilla is absent; the maxilla-jugal suture is straight, with no curvature. In the second state the maxilla-jugal suture makes a curve forming the zygomatic process of maxilla. Curvatures can be more obtuse or acute due to the curve on the anterior portion of the jugal bone. This will divide the anterior portion of the bone into a dorsal and a ventral area. This is a new character from this study.

8. Crest between the lacrimal foramen and the maxillary foramen: (0) Absent; (1) Present.

Inside the orbital cavity, just above the maxillary foramen, a clearly discernible crest runs horizontally in some specimens. This character is monomorphic and new character from this study.

9. Lateral expansion of dorsal and ventral areas of the mastoid process (Posterior view): (0) Same length; (1) Dorsal expands more.

The dorsal portion of the mastoid process can be of the same length as the ventral portion or, alternatively, it could projects further outwards. This is a new character from this study.

10. Infraorbital foramen shape: (0) Round; (1) Tear-shaped or oval.

The infraorbital foramen could be perfectly rounded, but could also have the dorsal portion narrower, giving it a tear shape, or laterally compressed giving it an oval shape. This is a modification of the character no. 32 of Prevosti (2010).

11. Foramen of the postorbital process of the frontal: (0) Slit like; (1) Round.

The foramen present in the postorbital process could have the opening narrow, slit-like, or round. This is a new character from this study.

12. Retroarticular foramen position: (0) The foramen is near the postglenoid process; (1) The foramen is near the distal portion of the squamosal bone.

The retroarticular foramen can be positioned immediately after the postglenoid process, or dislocated dorsally near the distalmost suture of the squamosal bone. This is a new character from this study.

13. Distal margin of palatine: (0) Ends anteriorly or on the same line as the distal margin of the 2nd upper molar; (1) Ends posteriorly to the distal margin of the 2nd upper molar.

The distal most margin of the palatine can ends anteriorly or with the end of the dental series. In contrast, the palatine can project posteriorly beyond the end of the dental series. This is a modification of characters no. 12 of Berta (1987), no. 26 of Tedford et al. (1995) and no. 30 of Prevosti (2010).

14. Basioccipital bulge: (0) Absent ;(1) Present.

A clear swallowing of the basioccipital is present near the internal margin of the bulla in some specimens. This basioccipital bulge can be shaped as a line or rounded. This is a new character from this study.

15. Bulla mesial posterior expansion (in relation to the jugular foramen): (0) Absent, doesn't cover the jugular foramen in ventral view (1) Present, partially covers the jugular foramen in ventral view.

The posterolateral portion of the auditory bulla can partially cover the jugular foramen, giving the impression that the foramen is narrower and smaller in ventral view. When the posterolateral portion of the bulla doesn't cover the jugular foramen, it appears that the foramen is larger. This is a new character from this study.

16. Auditory meatus tube: (0) Absent; (1) Present.

Berta (1987) described the diameter and the presence of the auditory meatus tube as a single character. We didn't observe variation in the diameter of the auditory canal. Wang et al. (1999) described two character states related to the length of the tube. Prevosti (2010) also described the length of the tube, but considered different states than Wang et al. (1999). We didn't observe difference in the length of the auditory meatus tube in the specimens examined, so we follow Prevosti's (2010) description. In the first state, the ventral portion of the external meatus does not extend laterally; making the opening of the external auditory meatus looks closer to the bulla. In the second state, the ventral portion of the external meatus expands laterally, forming a tube, separating the external auditory meatus opening from the bulla. This is a modification of characters no. 20 Berta (1987), no. 18 of Wang et al. (1999) and no. 48 of Prevosti (2010).

17. Shape of the ventral margin of the mandible: (0) Straight; (1) Curved.

In the first state, the ventral margin of the mandibular ramus has a more or less straight profile throughout its extension, while in the second state it forms a slight curvature, giving a concave appearance to the ventral margin of the mandible. This character is better observed in fused mandibles and on top of a straight surface in lateral view. This is a new character from this study.

18. Size of the angular process: (0) Larger antero-posteriorly than dorso-ventrally; (1) Larger dorso-ventrally than antero-posteriorly.

Tedford et al. (1995) and Prevosti (2010) described a single character each for the angular process, which is here considered two different characters, characters no. 18 and 19. For their characters related to the angular process, Tedford et al. (1995) have two states, while Prevosti (2010) has three states. We choose to separate them in two characters because we couldn't observe a correlation between the states described. In this character we considered the proportions of the angular process. The anteroposterior length of the process is larger than its height; in the other the height is greater than the anteroposterior length. This is a modification of characters no. 38 of Tedford et al. (1995) and no. 52 of Prevosti (2010).

19. Position of the tip of the angular process: (0) Same level as the dorsal margin of the process; (1) Higher than the dorsal margin of the process.

Tedford et al. (1995) and Prevosti (2010) descriptions of their characters for the angular process correspond to characters no. 18 and 19 of this study (see discussion in character 18). This character is related to the posterior tip of the angular process, which could be aligned with the dorsal margin of the process, or higher, curving upwards. This is a modification of characters no. 38 of Tedford et al. (1995) and no. 52 of Prevosti (2010).

20. Position of condylar process: (0) On the same line as the margin that connects it to the angular process; (1) Projects posteriorly in relation to the margin that connects it to the angular process.

The condylar process could be positioned immediately above the proximal base of the angular process, aligned with the posterior margin of the mandibular ramus, or it could project posteriorly. This is a new character from this study.

21. Shape of the anterior margin of coronoid process: (0) Straight; (1) Curved.

The anterior margin of the coronoid process could be relatively straight or curves gently outwards, giving a convex profile. Character no. 13 of Berta (1987) and no. 37 of Tedford et al. (1995) are related to the coronoid process, but describing aspects that we weren't able to observe in this study. This is a new character from this study.

22. Shape of the dorsal margin of coronoid process: (0) Straight; (1) Convex.

The upper margin of the coronoid process can assume different shapes, being straight, giving a square like appearance to the process, or curved, giving it a convex appearance. This is a new character from this study.

23. Subangular lobe: (0) Absent; (1) Present.

The subangular lobe is a convexity identifiable in the distoventral portion of the mandible in some carnivora taxa, particularly in some canids as *Urocyon*, *Cerdocyon*, *Nyctereutes* and *Otocyon*, and was described as a derived condition by Berta (1987). Our description is the same as Tedford et al. (1995), with two states. Prevosti (2010) described three states for the subangular lobe: (0) smooth or absent (1) sharp and wide and (2) well developed but laterally compressed. The last state is present in a species that wasn't included in our study group. In taxa examined here, this character is present only in *Urocyon cinereoargenteus* and *Cerdocyon thous*, with both species having similar subangular lobes, in which its height it shorter than that of the angular process. Therefore, we considered only two states. This is a modification of characters no. 26 of Berta (1987), no. 24 of Tedford et al. (1995), no. 27 of Wang et al. (1999) and no. 50 of Prevosti (2010).

24. i1 distal cusplets: (0) Absent; (1) Present.

Some specimens present small cusplets on the distal face of the first lower incisors. This is a new character from this study.

25. i2 distal cusplets: (0) Absent; (1) Present.

Some specimens present small cusplets on the distal face of the second lower incisors. This is a new character from this study.

26. i3 distal cusplets: (0) Absent; (1) Present.

Some specimens present small cusplets on the distal face of the third lower incisors. This is a new character from this study.

27. I1 mesial cusplets: (0) Absent; (1) Present.

Berta (1987) described the absence and presence of the cusplets of I1-2 on the same character and didn't differentiate between mesial and distal positions. Tedford et al. (1995) also described as a single character the mesial cusplets of I1-3. We follow Prevosti (2010) in which each incisor and cusplets are considered a separate character. This is a modification of characters no. 5 of Berta (1987), no. 17 of Tedford et al. (1995) and no. 57 of Prevosti (2010).

28. I1 distal cusplets: (0) Absent; (1) Present.

Some specimens present small cusplets on the distal face of the first upper incisors. This is a new character from this study.

29. I2 mesial cusplets: (0) Absent; (1) Present.

Berta (1987) and Tedford et al. (1995) described both the presence and absence of upper incisors cusplets as a single character (see discussion above). This is a modification of characters no. 5 of Berta (1987), no. 17 of Tedford et al. (1995) and no. 58 of Prevosti (2010).

30. I2 distal cusplets: (0) Absent; (1) Present.

Some specimens present small cusplets on the distal face of the second lower incisors. This is a new character from this study.

31. Distal cusplets of p3: (0) Absent; (1) Present.

Tedford et al. (1995) described the presence and absence of the posterior cusplets of P3 and p2-3 as a single character. We follow Prevosti (2010) in considering the presence of cusplets in each premolar as a separated character. In our observation only

p3 showed variation. This character was modified from no. 7 of Tedford et al. (1995) and no. 79 of Prevosti (2010).

32. m1 protostylid: (0) Absent; (1) Present.

The protostylid is a small cusp located on the mesiobuccal face of the lower molars of some specimens. This is a modification of characters Character no. 14 of Berta (1987), no. 29 of Tedford et al. (1995) and no. 83 of Prevosti (2010).

3.2 Phylogenetic analyses

The molecular analysis is summarized in a Bayesian majority-rule consensus (Fig. 1). Here we present only the result for our ingroup and without branch lenghts (See the full tree in Appendix 4). The clade *Cerdocyon* + *Atelocynus* + *Lycalopex* was recovered as monophyletic with maximum support (PP=1). The lineage leading to *Atelocynus microtis* would have been the first to diverge, followed by the *Cerdocyon thous* + *Lycalopex* lineages which had a weak support (PP=0.55). The monophyly of the genus *Lycalopex* had high support value (PP=1). The phylogenetic relationships within *Lycalopex* were moderately supported, with posterior probabilities ranging from 0.79 to 0.91. The first lineage to diverge in *Lycalopex* was *L. fulvipes* (PP=1), followed by *L. sechurae* (PP=0.87) *L. gymnocercus* (PP=0.79), *L. griseus* (PP=0,79), and *L. culpaeus* and *L. vetulus* (PP= 0.9).

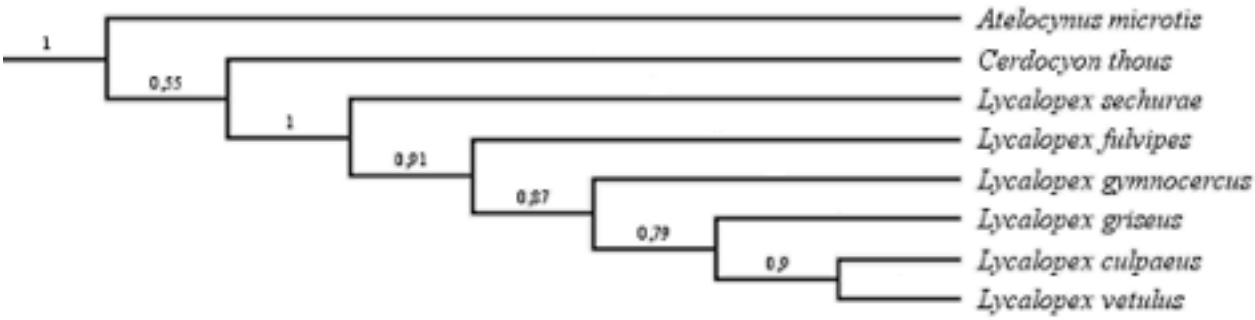

Figure 1. Bayesian Molecular phylogeny, topology only, of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* based on 25 molecular makers. Numbers on nodes refer to Bayesian Posterior Probabilities.

The molecular tree we obtained was not 100% similar to any other tree of prior studies (e.g. Berta 1987, Tedford et al. 1995, Lindblad-toh et al., 2005). It was most similar to the tree in Perini et al. (2009) study, the disagreement being the relationship

of *Atelocynus microtis* and *Cerdocyon thous*. In Perini et al. (2009) *Atelocynus microtis* and *Cerdocyon thous* form a clade which is sister of *Lycalopex*. In our tree the *Atelocynus microtis* lineage would have diverged first, followed by *Cerdocyon thous* and *Lycalopex* respectively. Both hypotheses have been presented by previous authors (Wayne et al., 1997; Lindblad-toh et al., 2005). The relationships for the ingroup had strong support values.

The following results were compared with the molecular tree we obtained. The analyses of the matrix coded using the Polymorphic method resulted in a single most parsimonious tree with 20 steps under equal weights (Fig. 2). With implied weight (k=3) it was obtained a single most parsimonious tree with fit= 0.25 (Fig. 3). The topologies found with equal weights and implied weight were the same, with most of the tree pectinated. The clade *Cerdocyon* + *Atelocynus* + *Lycalopex*, and the genus *Lycalopex* were not recovered as monophyletic groups, unlike the molecular tree. The lineage leading to *Atelocynus microtis* was recovered as the first divergence, even before *Urocyon cinereoargenteus*, which belongs to the outgroup. *Urocyon cinereoargenteus* was grouped with *Cerdocyon thous* in a polytomy with all species of *Lycalopex*. Overall, the tree is poorly resolved and with low support.

For the analyses of the matrix coded by the Majority method, a single most parsimonious tree was found with 57 steps under equal weights (Fig. 4). The tree topology was resolved but with low support values, all under 40. In this tree the monophyletic clades *Cerdocyon* + *Atelocynus* + *Lycalopex* and the genus *Lycalopex,* obtained in the molecular tree, were not recovered. *Cerdocyon thous* was the first lineage to diverge, even before *Urocyon cinereoargenteus,* which is an outgroup species, this last one having a sister group relationship with *L. vetulus. Atelocynus microtis* was sister clade of *Lycalopex culpaeus*. With implied weights (k=3) it was obtained a single most parsimonious tree with fit= 5.95 (Fig. 5). The monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex,* obtained in the molecular analysis, was recovered, but not the monophyly of the genus *Lycalopex*. The first lineage to diverge would have been *L. vetulus*, followed by *Cerdocyon thous* and *Atelocynus microtis*, which were placed in a polytomy among other *Lycalopex* species.

The analyses for the matrix using the Unordered coding method found a single most parsimonious tree, with 65 steps under equal weights (Fig. 6). The highest support value was 81. With implied weights (k=3) it was obtained a single most parsimonious tree with fit= 4.15 (Fig. 7). Support values were under 80. The topologies found with equal

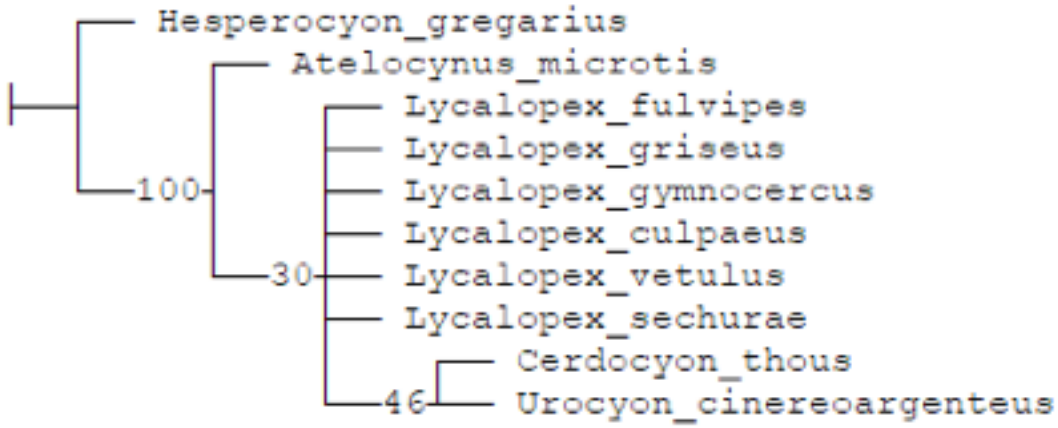

Figure 2. Most parsimonious tree obtained by the matrix coded by the Polymorphic coding method equal weights. Support measured with bootstrap.

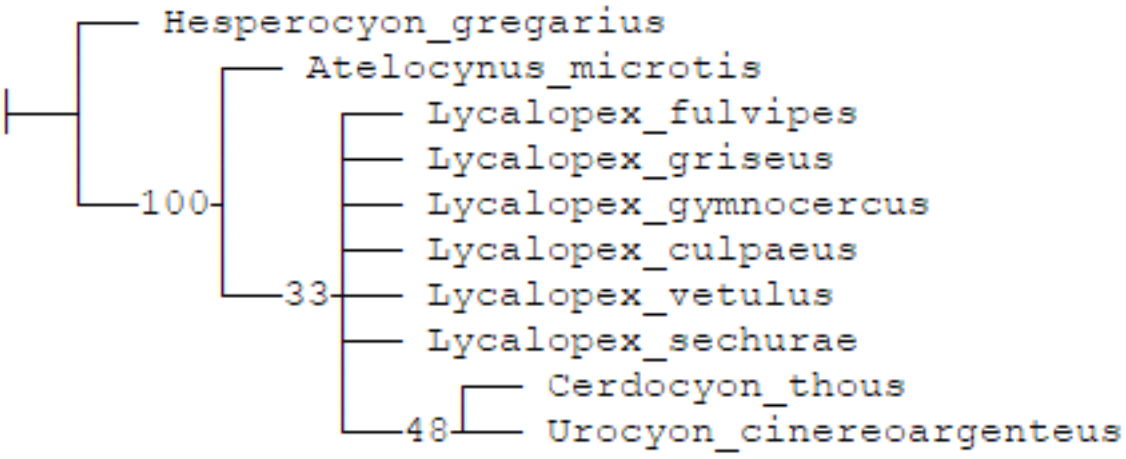

Figure 3. Most parsimonious tree obtained by the matrix coded by the Polymorphic coding method under implied weights (k=3). Support measured with bootstrap.

weights and implied weights were the same. The monophyletic clades *Cerdocyon* + *Atelocynus* + *Lycalopex* and the genus *Lycalopex*, obtained in the molecular tree, were not recovered. The lineage *Urocyon cinereoargenteus*, an outgroup species, was clustered with *Cerdocyon thous* among species of *Lycalopex*. *Lycalopex fulvipes* would have been the first lineage to diverge followed by *Atelocynus microtis*.

The analyses based on the matrix using the Unscaled method found a single most parsimonious tree with 75 steps under equal weights (Fig. 8). With implied weights (k=3) it also was obtained a single most parsimonious tree with fit= 5.8 (Fig. 9). The topologies found with equal weights and implied weights were the same. The highest support value

was 81 on the tree under equal weights for the clade *C. thous* and *U. cinereoargenteus*. The monophyletic clades *Cerdocyon* + *Atelocynus* + *Lycalopex* and the genus *Lycalopex*, obtained in the molecular tree, weren't recovered. The outgroup lineage *Urocyon cinereoargenteus* was clustered with *Cerdocyon thous* within the *Lycalopex* clade. In these hypotheses the *Atelocynus microtis* lineage would have diverged first.

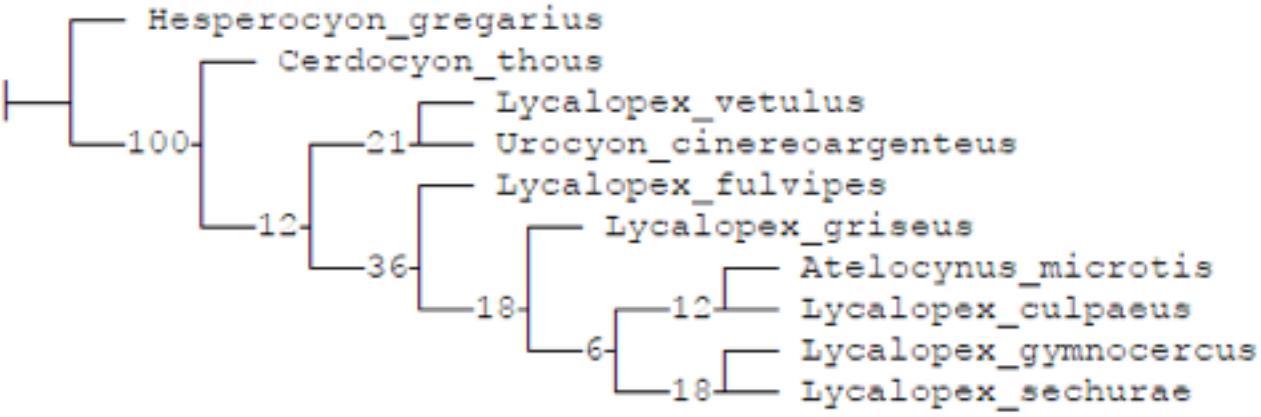

Figure 4. Most parsimonious tree obtained by the matrix coded by the Majority coding method under equal weights. Support measured with bootstrap.

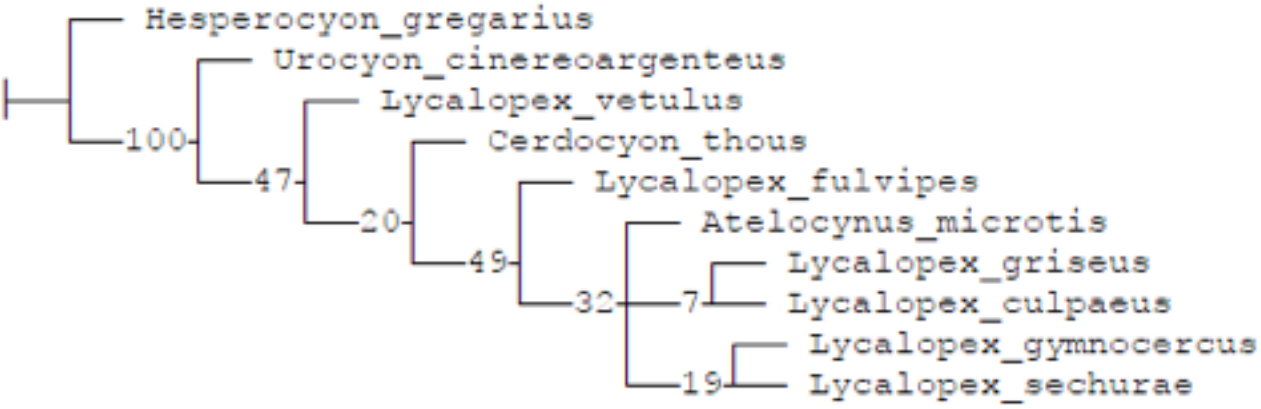

Figure 5. Most parsimonious tree obtained by the matrix coded by the Majority coding method under implied weights. Support measured with bootstrap.

The analyses of the matrix coded by the Frequency-bins method found a single most parsimonious tree with 503 steps under equal weights (Fig. 10). The implied weights (k=3) analysis also obtained a single most parsimonious tree with fit= 49.2 (Fig. 11). Tree topology was the same for the analyses done under equal and implied weights. The highest support value was 64 on the implied weight tree for a clade including species of *Lycalopex* (excluding *L. vetulus*) and *A. microtis*. On these phylogenies the monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* was recovered, but not of

the genus *Lycalopex*. The lineage *Lycalopex vetulus* would have diverged first followed by *Cerdocyon thous*, and *Atelocynus microtis* forms a sister group with *L. fulvipes*.

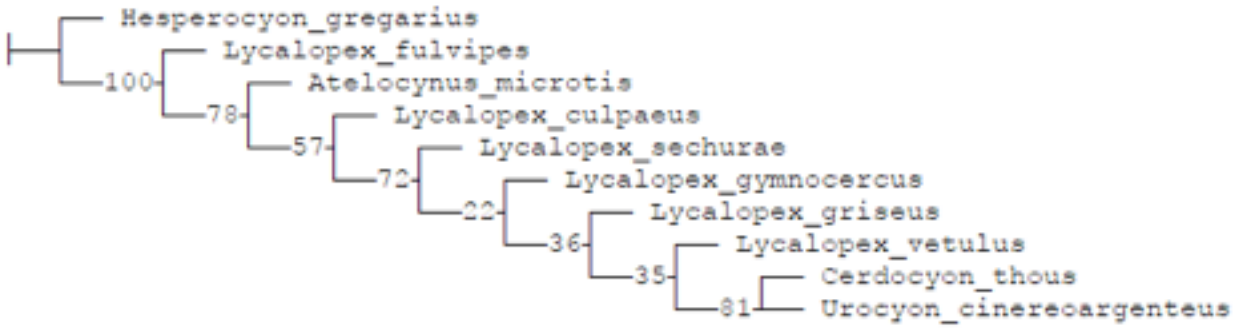

Figure 6. Most parsimonious tree obtained by the matrix coded by the Unordered coding method under equal weights. Support measured with bootstrap.

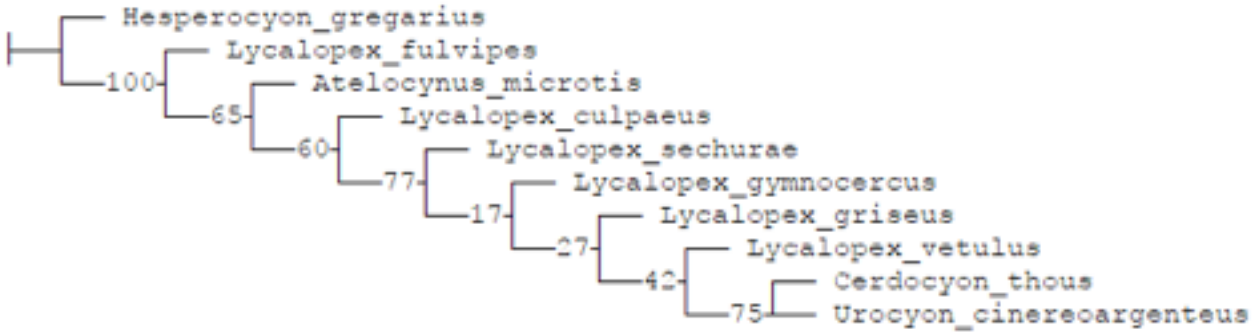

Figure 7. Most parsimonious tree obtained by the matrix coded by the Unordered coding method under implied weights. Support measured with bootstrap.

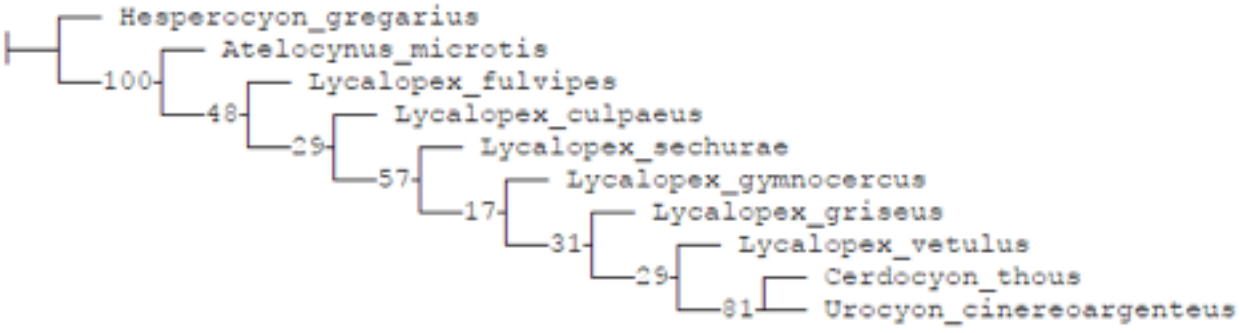

Figure 8. Most parsimonious tree obtained by the matrix coded by the Unscaled method under equal weights. Support measured with bootstrap.

For the matrix using the Frequency-as-continuous method a single most parsimonious tree was found with 53.98 steps under equal weights (Fig. 12). Support

values were under 60. The monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex spp*. and the genus *Lycalopex*, which are recovered in the molecular tree, weren't recovered. The lineage *Cerdocyon thous* would have diverged prior to *Urocyon cinereoargenteus*, which belongs to the outgroup and in this hypothesis forms a sister clade with *Lycalopex vetulus*. The species *Atelocynus microtis* forms a clade with *L. fulvipes*. With implied weights (k=3) it was obtained a single most parsimonious tree with fit= 6.02 (Fig. 13). Support values were under 60. In this tree the monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex spp*. was recovered but not for *Lycalopex*, unlike the molecular tree. *Lycalopex vetulus* would have been the first to diverge followed by *Cerdocyon thous*, while *Atelocynus microtis* was sister group of *L. fulvipes*.

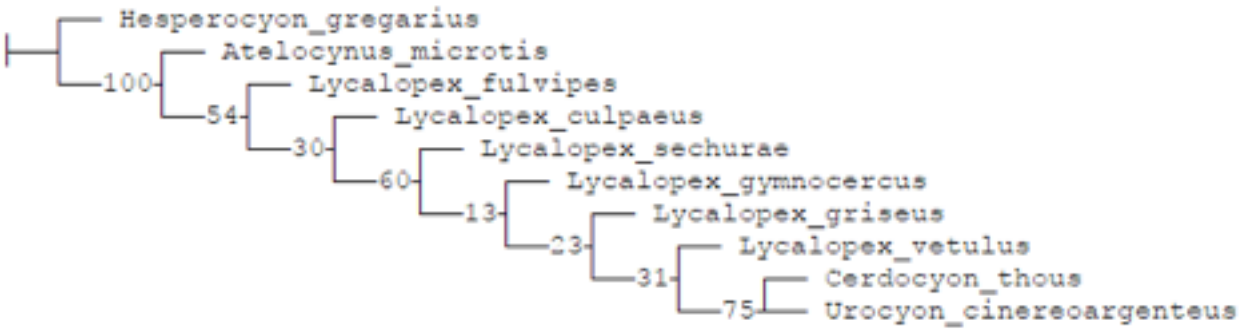

Figure 9. Most parsimonious tree obtained by the matrix coded by the Unscaled method under implied weights. Support measured with bootstrap.

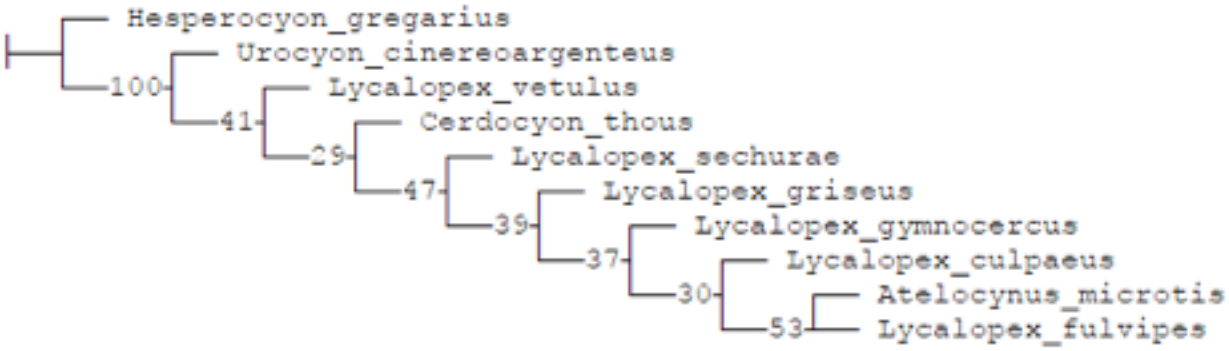

Figure 10. Most parsimonious tree obtained by the matrix coded by the Frequency-bins method under equal weights. Support measured with bootstrap.

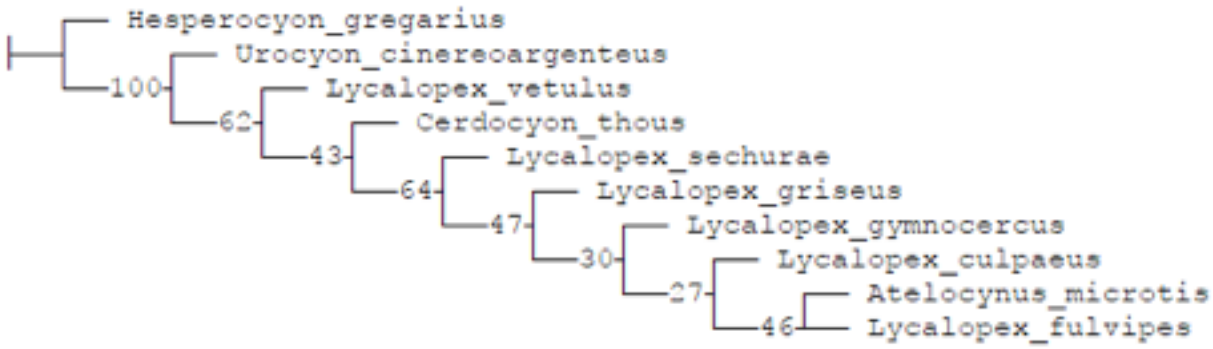

Figure 11. Most parsimonious tree obtained by the matrix coded by the Frequency-bins method under implied weights. Support measured with bootstrap.

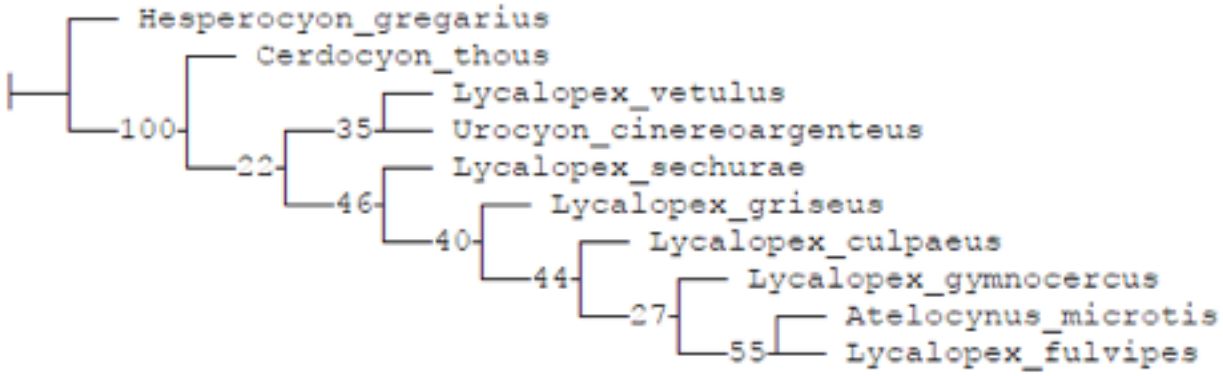

Figure 12. Most parsimonious tree obtained by the matrix coded by the Frequency-as-continuous method under equal weights. Support measured with bootstrap.

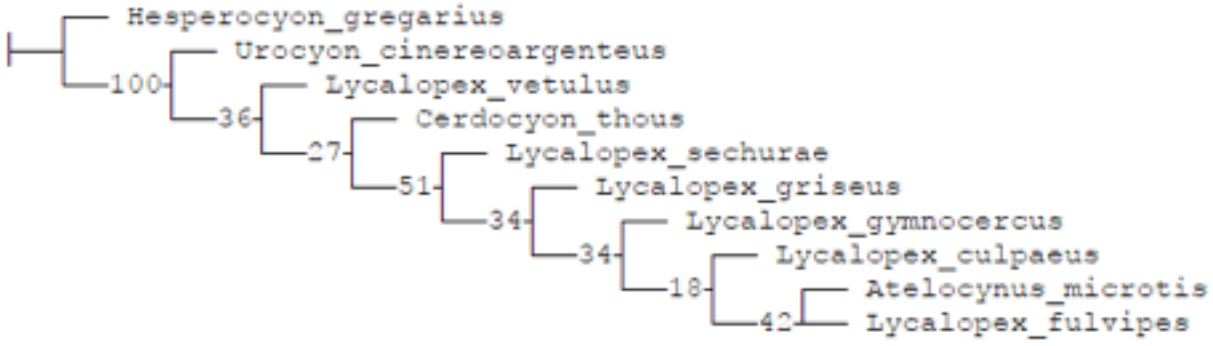

Figure 13. Most parsimonious tree obtained by the matrix coded by the Frequency-as-continuous method under implied weights. Support measured with bootstrap.

According to the SPR-distances test, the matrix that recovered the most similar

topology to the molecular tree we obtained was coded by the Polymorphic method with 86% of similarity, which would require 1 SPR move. The second most similar tree was coded by the Frequency-bins method with 43% of similarity which would require 4 SPR moves. The results were the same in equal and implied weight analyses using both coding methods. The Majority and Frequency-as-continuous methods also presented 43% of similarity related to the molecular topology, but only when analyzed with implied weights. The other approaches had lower similarity scores (Tab. 3).

Table 3. Similarity in percentage and SPR moves between trees recovered by matrices coded according to the Polymorphic, Majority, Unordered, Unscaled, Frequency-bins and Frequency-as-continuous methods. Prevosti (2010) matrix, only morphological characters, and combined matrices.

	EW		IW	
Matrix	**Similarity**	**SPR moves**	**Similarity**	**Similarity**
Polymorphic	86%	1	86%	1
Majority	14%	6	43%	4
Unordered	0%	7	0%	7
Unscaled	0%	7	0%	7
Frequency-bins	43%	4	43%	4
Frequency-as-continuous	28%	5	43%	4
Prevosti (2010) morphology	43%	4	43%	4
Prevosti + Polymorphic	57%	3	57%	3
Prevosti + Frequency-bins	43%	4	43%	4

The analysis including the morphological data of Prevosti (2010) resulted in a single most parsimonious tree with 131.770 steps under equal weights (Fig. 14). With implied weights (k=3) it was obtained also a single most parsimonious tree with fit= 2.37 (Fig. 15). There was no difference in tree topologies found by searches done with equal and implied weights. In both topologies the monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* was recovered, but not of the genus *Lycalopex*. The first lineage to diverge would have been *Lycalopex vetulus*, followed by two divergencies where *Cerdocyon thous* and *Atelocynus microtis* formed a sister clade grouped with L. fulvipes. This tree had 43% of similarity with our molecular tree which required 4 SPR moves (Tab. 3).

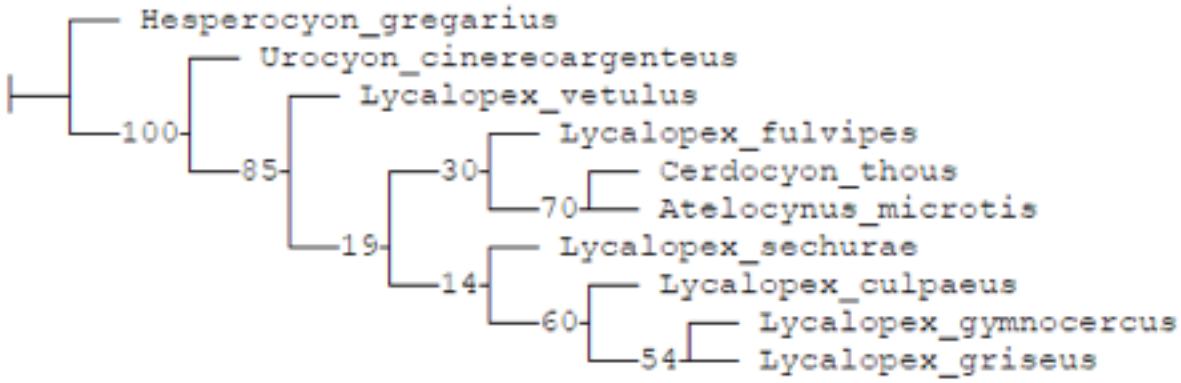

Figure 14. Most parsimonious tree obtained by the morphological characters of Prevosti (2010) under equal weights. Support measured with bootstrap.

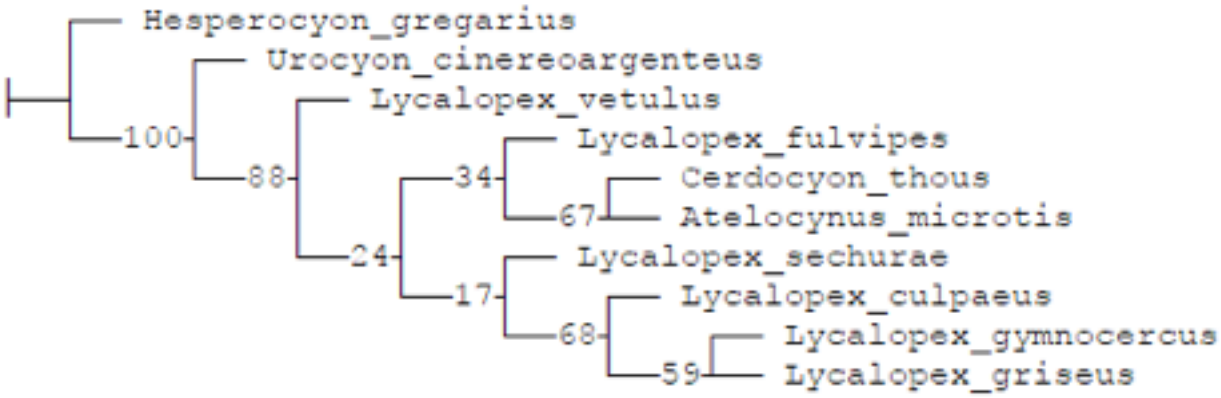

Figure 15. Most parsimonious tree obtained by the morphological characters of Prevosti (2010) implied weights. Support measured with bootstrap.

The combined analyses of our matrix coded by the Polymorphic method with Prevosti's (2010) morphological characters resulted a single most parsimonious tree with 152.786 steps under equal weights (Fig. 16). With implied weights (k=3) it was obtained a single most parsimonious tree with fit= 2.87 (Fig. 17). Highest support values was 80 on the tree under equal weights for the clade clades *Cerdocyon* + *Atelocynus* + *Lycalopex*. Tree topologies were the same for equal and implied weights. On these phylogenies both clades *Cerdocyon* + *Atelocynus* + *Lycalopex* and the genus *Lycalopex* were monophyletic, as in the molecular tree. *Cerdocyon thous* and *Atelocynus microtis* formed the sister group of *Lycalopex*. The lineage *Lycalopex* first diverged into two clades composed by *L. vetulus* + *L. sechurae* and *L. fulvipes* + *L. culpaeus* + *L.*

gymnocercus + *L. griseus*. In last clade the first lineage to diverge was *L. fulvipes* followed by *L. culpaeus*, which is sister clade of the group formed by *Lycalopex gymnocercus* and *Lycalopex griseus*.

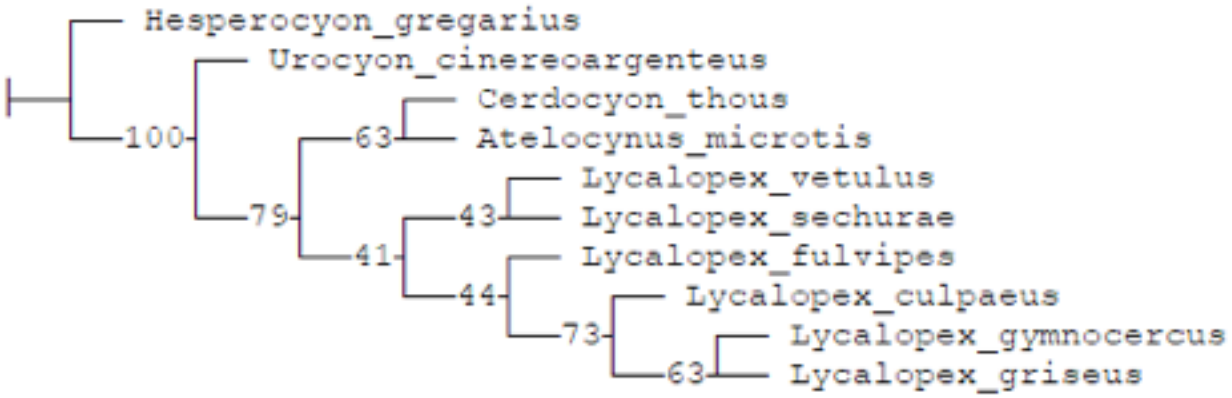

Figure 16. Most parsimonious tree obtained by the combined matrix, Polymorphic matrix + morphological characters of Prevosti (2010), under equal weights. Support measured with bootstrap.

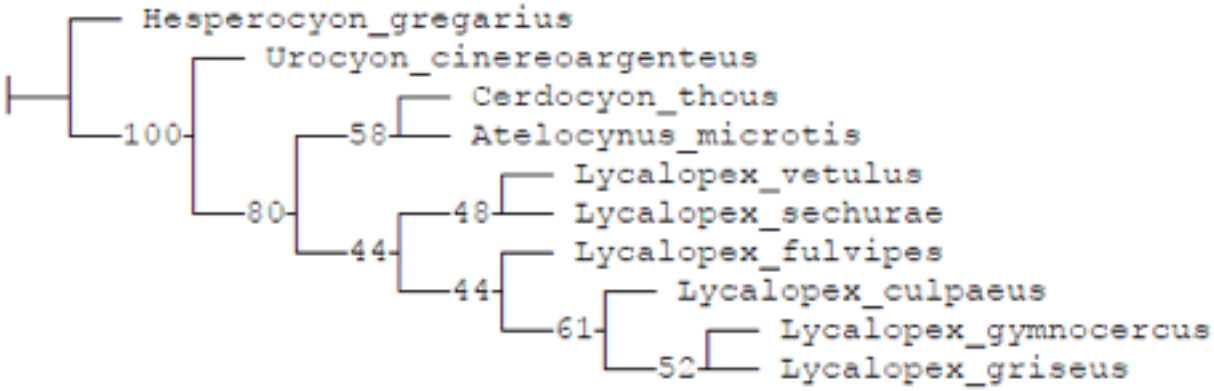

Figure 17. Most parsimonious tree obtained by the combined matrix, Polymorphic matrix + morphological characters of Prevosti (2010), under implied weights. Support measured with bootstrap.

The combined analyses of our matrix coded by the Frequency-bins method with Prevosti's (2010) morphological characters resulted in a single most parsimonious tree with 646.795 steps under equal weights (Fig. 18). With implied weights (k=3) it was obtained a single most parsimonious tree with fit= 52.77 (Fig. 19). The highest support value was 68 on the tree under implied weights. Tree topologies were the same for equal and implied weights. In these trees the monophyly of the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* was recovered, but not for the genus *Lycalopex*. The lineage of *Cerdocyon*

thous would have been the first to diverge, followed by the genus *Lycalopex*, while *Atelocynus microtis* formed a clade with *Lycalopex fulvipes*.

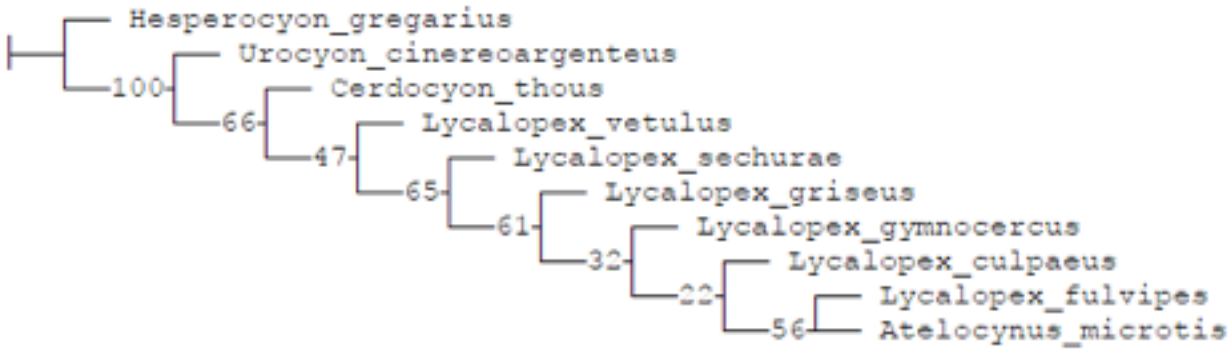

Figure 18. Most parsimonious tree obtained by the combined matrix, Frequency-bins matrix + morphological characters of Prevosti (2010), under equal weights. Support measured with bootstrap.

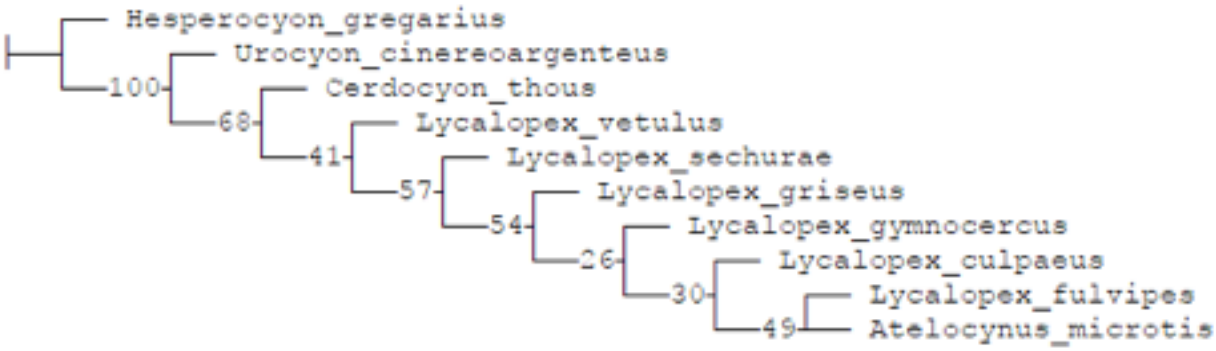

Figure 19. Most parsimonious tree obtained by the combined matrix, Frequency-bins matrix + morphological characters of Prevosti (2010), under implied weights. Support measured with bootstrap.

Overall the polymorphic matrix had 50% of polymorphic entries. It varied between taxa from 3,1% to 68,8% (Table 4). The percentage of missing entries in total was 0,03% and varied from 3,1% to 34,4% among species (Tab. 4). The relationship between cells containing polymorphic entries in the matrix coded by the Polymorphic method and sample size were strongly and positively correlated (R = 0.8, p=0.005) across taxa in our dataset (Fig. 20).

Table 4 - Percentage of polymorphic and missing entries for the Polymorphic matrix

Species	Polymorphic entries	Missing entries
Atelocynus microtis	40,6%	0,0%
Cerdocyon thous	68,8%	0,0%
Lycalopex culpaeus	43,8%	3,1%

Species	Polymorphic entries	Missing entries
Lycalopex fulvipes	3,1%	0,0%
Lycalopex griseus	68,8%	0,0%
Lycalopex gymnocercus	68,8%	0,0%
Lycalopex sechurae	62,5%	0,0%
Lycalopex vetulus	75,0%	0,0%
Urocyon cinereoargenteus	68,8%	0,0%
Hesperocyon gregarius	3,1%	34,4%

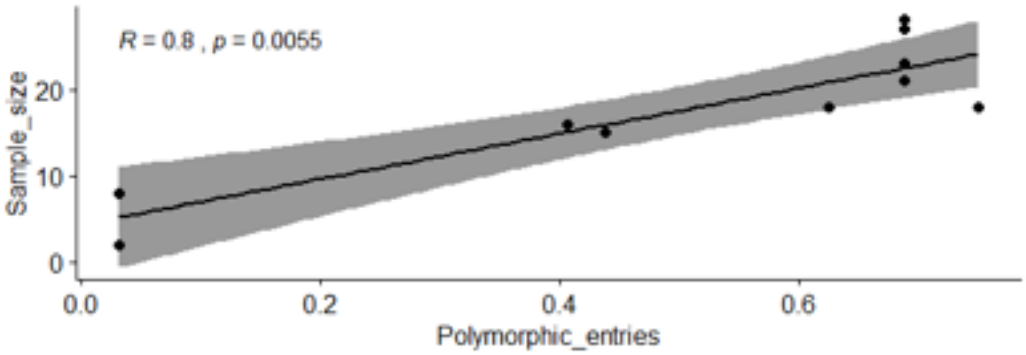

Figure 20. Spearman rank test results for the correlation between the percentage of cells containing polymorphic entries and sample sizes.

4 Discussion

In this study we observed 140 specimens of the in ingroup, being the largest sampling effort done with the intention of gathering morphological characters for the clade *Cerdocyon* + *Atelocynus* + *Lycalopex*. Tedford et al. (1995) and Prevosti (2010) collected morphological data for more inclusive clades of canids, which included our study group. Our sample size was four times larger than Tedford et al. (1995), which was of 34 specimens and it did not included specimens of *Lycalopex fulvipes*. Compared to Prevosti (2010), our sample size was 30% larger, since he observed 107 specimens of our study group. Our observations revealed a larger percentage of polymorphic characters than in previous studies (e.g. Berta, 1987; Tedford et al., 1995; Perini et al, 2009; Prevosti, 2010), most of which do not have a single polymorphic character in their matrix. This could be an outcome of our sample size or exclusion of polymorphic characters (Poe and Wiens, 2000) by previous authors.

It can be expected that, when working with qualitative characters, there will be some coding disagreements between authors. However, it was a surprise when we noted that even in binary characters coded for absence and presence, there was disagreement. We presume that this could be the result of sample size, exclusion of polymorphic characters (Poe and Wiens, 2000) and we add that even character states could have been excluded in previous studies. This reinforces the need for authors to specify their criteria to exclude data from their matrix (Poe and Wiens, 2000). Even though disagreement might be related to sample size, it is a matter of concern to observe it in characters which are supposed to have a straightforward coding, such as in characters that have absent and present as states.

In most of the trees we obtained from the different matrices, the topologies were not affected by the use of equal and implied weights. This suggest that our matrix could have a greater percentage of homoplasy. This is expected because it has been shown that polymorphic characters are more homoplastic (Campbell and Frost, 1993). Overall support values were low for the trees we found. The strong and positive correlation obtained in the Spearman rank test, which analyzed the relationship between cells containing polymorphic entries and sample size in the Polymorphic matrix, confirmed that within our matrix polymorphism coding was affected by sample size, similar to Prevosti's (2010) results.

None of the trees obtained were 100% similar to the molecular tree, but since there are similarities between our data and the molecular tree, we believe this suggests that polymorphic characters may have phylogenetic signal (Campbell and Frost, 1993;

Wiens, 1998). According to the SPR-distance test, the "Polymorphic" tree had the most similar topology to the molecular tree, with 86% of similarity, which is much higher than all other trees. However, the "Polymorphic" tree has a big portion of its topology pectinated, and we suspect that this could be the reason for such high similarity, since a SPR move in a tree that has low resolution can have a big effect on its topology. For this reason we chose to make a combined matrix with the second most similar tree, which is the “Frequency-bins” tree. The “Frequency-bins" tree topology was resolved and it had 43% of similarity with the molecular tree. Prevosti (2010) tree also had 43% of similarity with the molecular tree even though his matrix had many more morphological characters. The similarity and the topology of coding method we proposed here, Frequency-as-continuous, was the same as in the Frequency-bins method, therefore better than most other methods, and we believe it is a valid method of coding data.

The difference on the similarity of the trees found by the analyses of the combined matrices was less contrasting when compared with the single matrices. The Polymorphic matrix + Prevosti obtained a tree with 57% of similarity, it was higher when compared with Prevosti’s tree, but less when compared with the “Polymorphic” tree.

According to Wiens (1998), the most accurate results in parsimony analyses it is obtained when using frequencies in the coding method. The most similar tree was obtained from the matrix coded by Polymorphic method, but, as we mentioned before, the “Polymorphic" tree has a few peculiarities which we believe could have affected the similarity results. Therefore we cannot discard the Frequency-bins method, which obtained the second most similar tree to the molecular tree, as the best method to code our matrix. We certainly disagree with Campbell and Frost (1993) in that the unscaled method is the best, since the tree we obtained by coding with this method had no similarity whatsoever with our molecular tree.

We suggest that authors should make it clear their criteria to excluded characters and/or character states, because we cannot be sure if polymorphism is uncommon in data in general, or it has not been reported. Sample size affects the coding of polymorphism, and sampling must be throughout whenever possible. Polymorphic characters should not be excluded from matrices because they have phylogenetic signals, despite being more homoplastic. The inconsistency between phylogenies for the clade *Cerdocyon* + *Atelocynus* + *Lycalopex* could be due to the high polymorphism we observed for this clade. We feel that the rapid diversification patterns observed within Canidae and recent divergency within the South American canids (Perini et al., 2009) could be correlated to the high polymorphism we observed. Previous authors could have

coded only one of the states when there are actually two states. This results in 50% chance that, in binary characters, coding could be different between authors for each character in common. We recommend researchers to take notes of states frequencies and publish with its results because frequencies can also be used as a guide to minimize subjectiveness when coding by other methods. Publishing the observed frequencies also provides more transparency to the data obtained by authors, it can facilitate the replication of results by other researches and it can provide information for future studies such as how polymorphic characters affect phylogenies.

References

Bardeleben, C., Moore, R. L., and Wayne, R. K. (2005). A molecular phylogeny of the Canidae based on six nuclear loci. *Molecular Phylogenetics and Evolution, 37*, 815–831. https://doi. org/10.1016/j.ympev.2005.07.019

Berta, A.. 1987. Origin, diversification, and zoogeography of the South American Canidae. Fieldiana Zoology, 39, 455-471.

Campbell, J. A., Frost, D. R.. 1993. Anguid lizards of the genus *Abronia*: Revisionary notes, descriptions of four new species, a phylogenetic analysis, and key. Bulletin of the American Museum of Natural History. 216, 121 pp.. ISSN 0003-0090.

Darwin, C.. 1859. The origin of species by means of natural selection, or preservation of favored races in the strugge for life. London: John Murray. Print.

Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5): 1792-97.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. Science, 294(5550), 2310-2314.

Kluge, A. G., Farris J. S.. 1969. Quantitative phylogenetic and the evolution of anurans. Syst. Zool. 18: 1-32.

Kück, P., and Longo, G. C. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Frontiers in Zoology, 11(1), 81.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30(22): 3276-3278.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution, 34(3), 772-773.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., ... Lander, E. S. (2005). Genome sequence, comparative analysis, and haplotype structure of the domestic dog. *Nature*, *438*, 803–819. https://doi.org/10.1038/nature04338

Maddison, W. P. and D.R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51 http://www.mesquiteproject.org

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.

Nixon, K.C., Wheeler, Q.D., 1990. An amplification of the phylogenetic species concept. Cladistics 6, 211–223.

Perini, F. A., Russo, C. A. M., Schrago, C. G.. 2009. The evolution of South American endemic canids: a history of rapid diversification and morphological parallelism. Journal of evolutionary biology, 23, 311–322. doi10.1111/j. 1420-9101.2009.01901.x.

Prevosti, F. J. 2010. Phylogeny of the large extinct South American Canids (Mammalia, Carnivora, Canidae) using a 'total evidence' approach. *Cladistics*, *25*, 1–26. https://doi. org/10.1111/j.1096-0031.2009.00298.x

Poe, S. and Wiens, J.J. 2000. Chapter 2 - Character selection and methodology of morphological phylogenetics in Wiens J. J. (ed.) Phylogenetic analysis of morphological data. Smithsonian Institute Press, Washington: 21-35.

Rambaut, A., Drummond AJ, Xie D, Baele G and Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biology. syy032. doi:10.1093/sysbio/syy032

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

Swofford DL, Berlocher SH. 1987. Inferring evolutionary trees from gene frequency data under the principle of max- imum parsimony. *Systematic Biology* **36:** 293–325.

Tchaicka et al. 2016. Molecular assessment of the phylogeny and biogeography of a recently diversified endemic group of South American canids (Mammalia: Carnivora: Canidae). Genetics and Molecular Biology, 39, 3, 442-451. DOI: http://dx.doi.org/ 10.1590/1678-4685-GMB-2015-0189.

Tedford, R. H., Taylor, B. E., and Wang, X. (1995). Phylogeny of the Canidae (Carnivora, Canidae): The living taxa. *American Museum Novitates*, *3126*, 1–37.

Tedford, R. H., Wang, X., and Taylor, B. E. (2009). Phylogenetic sys- tematics of the North American fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History*, *325*, 1–218. https://doi.org/10.1206/574.1

Wang, X., Tedford, R. H., and Antón, M. (2008). *Dogs: Their fossil relatives and evolutionary history*. New York, NY: Columbia University Press. https://doi.org/10.7312/ wang13528

Wayne, R. K., Geffen, E., Girman, D. J., Koepfli, K. P., Lau, L. M., and Marshall, C. R. (1997). Molecular systematics of the Canidae. *Systematic Biology*, *46*, 622–653. https://doi.org/10.1093/ sysbio/46.4.622

Wiens, J.J., 1995. Polymorphic characters in phylogenetic systematics. Syst. Biol. 44, 482–500.

Wiens, J.J., 1998. Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. Syst. Biol. 47, 427–444.

Wiens, J.J., 1999. Polymorphism in systematics and comparative biology. Annu. Rev. Ecol. Syst. 30, 327–362.

Wiens, J.J., Servedio, M.R., 1997. Accuracy of phylogenetic analysis including and excludin polymorphic characters. Syst. Biol. 46, 332–345.

Wiens, J.J., Servedio, M.R., 1998. Phylogenetic analysis and intraspecific variation: performance of parsimony, likelihood, and distance methods. Syst. Biol. 42, 228–253.

Wozencraft, W.C. 2005. Order Carnivora. In: Mammal Species of the World – A Taxonomic and Geographic Reference (D.E. Wilson and D.M. Reeder, eds), pp. 532–628. The Johns Hopkins University Press, Baltimore, MD.

Zrzavy, J., Duda, P., Robovsky, J., Okrinová, I., Rincanková, V. P.. 2018. Phylogeny of the Caninae (Carnivora): Combining morphology, behaviour, genes and fossils. Zoologica Scripta, 1-17. DOI 10.1111/zsc.12293. 2018