http://www.icn.unal.edu.co/ Caldasia 33(2):637-658. 2011

DATA SET INCONGRUENCE, MISLEADING CHARACTERS, AND INSIGHTS FROM THE FOSSIL RECORD: THE CANID PHYLOGENY

Incongruencia, datos conflictivos e indicios del registro fósil: la filogenia de los cánidos

JOAO MUÑOZ-DURÁN
Departamento de Biología, Universidad Nacional de Colombia, Bogotá D.C., Colombia. jvmunozd@unal.edu.co

ABSTRACT

Identifying and accounting for sources of significant, explicit phylogenetic conflicts among data sets is an issue that requires further study. In this paper I explore the usefulness of the known fossil record for assessing the accuracy of conflicting sister taxa hypotheses, and in identifying and accounting for misleading characters (MCs). The alternative I present begins with a parsimony analysis of each available data set. Bootstrap proportions >95% supporting conflicting clades among most parsimonious trees (MPTs) identify instances of "strong" data set incongruence. The accuracy of conflicting sister taxa hypotheses is assessed through a comparison of their temporal gaps (*T*). Conflicting clades, with a significantly longer than average *T,* are called into question. As exceptionally long *T*s can result from incompleteness and/or biases in the fossil record, it is necessary to differentiate the effect of MCs from the effect of a fragmentary fossil record. For this, the effect that characters supporting questioned conflicting clades have on data set homoplasy is assessed. If resetting these characters to missing values reduces data set homoplasy in a manner that is significantly different from random, then conflicting clades with exceptionally long *T* must arise from the effect of MCs. If so, new MPTs are calculated for modified data sets and the testing process is repeated until no more well-supported, conflicting clades are found. Finally, data sets are combined and the MPT is calculated. I applied this approach to the phylogeny of the Caninae using morphological and mtDNA data sets. Among the MCs characters identified were some that cannot be accounted for by commonly used a priori weighting schemes. The phylogeny of canids is also briefly discussed. The resulting MPT suggests the colonization of South America by three canid lineages and that the trenchant heel, a trait associated with hypercarnivory and sociality, evolved only once within the Caninae.

Key words. Stratigraphic data, Phylogenetic inference, temporal gap, misleading characters, Caninae.

RESUMEN

Identificar y controlar las fuentes de conflictos filogenéticos explícitos entre particiones de datos es un tema que requiere mayor esfuerzo de investigación. En este trabajo exploro la utilidad del registro fósil en la evaluación de hipótesis conflictivas de clados hermanos y en la identificación y control de caracteres artificiosos (MCs). La alternativa que presento inicia con un análisis de máxima parsimonia independiente para cada partición de datos disponible. Clados conflictivos y soportados por porcentajes de Bootstrap >95% entre árboles mas parsimoniosos (AMPs) identifican instancias de incongruencia "fuerte" entre particiones de datos. La precisión de las hipótesis conflictivas de clados hermanos se prueba mediante la comparación de saltos temporales (*T*). Clados conflictivos con *T*s significativamente más largos que el promedio son cuestionados y objeto de análisis adicionales. Puesto que los *T*s excepcionalmente largos pueden ser el resultado de un registro fósil incompleto o sesgado, se hace necesario diferenciar el efecto de un registro fósil inadecuado del efecto que sobre el nivel de homplasia tienen los caracteres que soportan los clados cuestionados. Si la recodificación de estos caracteres como datos faltantes reduce el nivel de homoplasia de una manera que es significativamente diferente a la de un efecto aleatorio, entonces los clados conflictivos con *T*s excepcionalmente largos deben ser el resultado de MCs. De ser así, el siguiente paso es calcular AMPs para las particiones modificadas y el proceso se repite hasta que no se obtienen clados conflictivos fuertemente soportados. Finalmente, las particiones se combinan y se calcula el AMP. Esta aproximación fue aplicada en la estimación de la filogenia de los cánidos utilizando dos particiones: ADN mitocondrial y morfología. Entre los MCs identificados hay algunos que no pueden ser controlados mediante métodos de pesaje a priori. También se discute brevemente la filogenia de los cánidos. El AMP resultante sugiere la colonización de Sur América por tres linajes de cánidos. La topología de este cladograma también indica que el talón cortante de los géneros hipercarnívoros y sociales evolucionó una única vez entre los Caninae.

Palabras clave. Datos estratigráficos; inferencia filogenética; saltos temporales, caracteres artificiosos, Caninae.

INTRODUCTION

The growing amount and diversity of phylogenetic data has motivated many efforts to generate more accurate phylogenetic inferences and several approaches for managing diverse sources of phylogenetic information have been proposed (Mickevich 1978, Fisher 1988, Kluge 1989, Fisher 1992, Bull *et al.* 1993, de Queiroz 1993, Rodrigo *et al.* 1993, Farris *et al.* 1995, Miyamoto & Fitch 1995, Huelsenbeck & Bull 1996, Clyde & Fisher 1997, Ballard *et al.* 1998, Lapointe 1998, Wiens 1998, O'Keefe & Sander 1999, Maddison & Knowles 2006, Ané *et al.* 2007, Edwards *et al.* 2007, Liang & Pearl 2007). Similarly, a number of tests have been implemented to determine whether there is significant incongruence among different data sets (Templeton 1983, Felsenstein 1985, Kishino & Hasegawa 1989, de Queiroz 1993, Rodrigo *et al.* 1993, Farris *et al.* 1995, Huelsenbeck & Bull 1996). Weighting schemes have been suggested to reduce the effect of misleading characters and improve the detection of the phylogenetic signal from molecular data (Martin 1995, Hillis *et al.* 1996, Naylor & Brown 1997, Hassanin *et al.* 1998). In addition, many attempts have been made to assess the quality of the results of phylogenetic inference by testing for congruence between cladistic patterns and stratigraphic data (Fisher 1992, Norell & Novacek 1992, Benton & Storrs 1994, Huelsenbeck 1994, Wagner 1995, Huelsenbeck & Rannala 1997, Siddall 1998, Wills 1999, Angielczyk 2002, Pol & Norell 2006, Wills *et al.* 2008, Pyron 2010). Nevertheless, identifying and accounting for sources of significant, explicit phylogenetic conflicts among data sets is an issue that requires further study. The objective of this paper is to explore the use of the known fossil record as an independent source of information in assessing the accuracy of conflicting, well-supported sister taxa hypotheses derived from different data sets, and in identifying and accounting for misleading characters. For this, I propose an approach based on the comparison of temporal gaps among conflicting sister taxa hypotheses. The phylogeny of the sub-family Caninae will be used as an example.

For the purposes of this paper, misleading characters (MCs) are defined as the fraction of all homoplastic characters and character states whose inclusion in a phylogenetic analysis may obscure the pattern of shared ancestry present in a data set and may lead to “strong” incongruence among data sets. MCs are not necessarily evenly distributed across all taxa in an analysis, and thus they only confuse the phylogenetic signal in few regions of a phylogeny. MCs differ from noisy characters (random data *sensu* Wenzel & Siddall 1999) in that they include not only homoplasies that result from random processes, but also those caused by processes of evolutionary convergence (adaptive and nonadaptive), by among lineages differences in life history traits, or by structural, functional and developmental constraints. MCs are not uninformative; on the contrary, they provide valuable information about evolutionary patterns and mechanisms, especially those related to structural and functional constraints. However, controlling for MCs is desirable in trying to improve the accuracy of phylogenetic relationships derived from a data set.

Assessing Conflicting Phylogenetic Hypotheses with Paleontological Data

Paleontological data and phylogenies are independent sources of information on the evolutionary history of taxa. Thus, the fossil record might be used in testing the accuracy of conflicting phylogenetic hypotheses and in detecting misleading characters. It has been shown that paleontological data may provide valuable information about characters or combinations of characters that elucidate aspects of character polarity and distribution of synapomorphies, thus improving the accuracy of phylogenetic estimation (Patterson 1981, Donoghue *et al.* 1989, Novacek 1992, Eernisse & Kluge 1993, Wheeler *et al.* 1993, Benton 1995, Miyamoto & Fitch 1995, Novacek 1996, Brochu 1997, Smith 1998, Naylor & Adams 2001, Asher *et al.* 2003, Santini & Tyler 2003, Hermsen *et al.* 2006, Cobbett *et al.* 2007, Magallón 2010). Fossil data may also subdivide long branches leading to erroneous sister taxa hypotheses resulting from parallel character changes; therefore this type of data may help in correcting for long branch attraction effects (Wiens 2005). Furthermore, the inclusion of fossil data in phylogenetic analyses has the potential of overturning proposed evolutionary relationships based on data from extant taxa alone (Gauthier *et al.* 1988, Donoghue *et al.* 1989, Novacek 1992, Smith 1998, Rothwell & Nixon 2006, Cobbett *et al.* 2007).

Paleontology also provides information on the approximate age of origination of clades which may be useful in testing the reliability of conflicting hypotheses of sister taxa relationships derived from the separate analyses of data sets. A variety of metrics have been proposed for assessing the congruence between stratigraphic and cladistic data (Fisher 1992, Norell & Novacek 1992, Benton & Storrs 1994, Huelsenbeck 1994, Smith & Littlewood 1994, Wagner 1995, Siddall 1998, Wills 1999, Angielczyk 2002, Pol & Norell 2006, Wills *et al.* 2008, Pyron 2010). These approaches are used to find a cladogram, from a number of competing cladograms, that best matches the known stratigraphic data for a group of taxa. Competing cladograms may represent alternative topologies from a single data set or from multiple data sets. Assuming parsimony, the cladogram that shows the best fit to the stratigraphic data is considered to reflect most accurately the evolutionary history of the group. The drawback that still remains in the above approaches is that they estimate the degree to which an entire tree fits the available stratigraphic data. Since phylogenetic trees most likely present a combination of clades with different degrees of fit to stratigraphic data, the identification of specific clades in the tree that show a poor fit is obscured when a good, global fit to stratigraphic sequences is found. Distinguishing between clades with relatively low and high levels of fit to stratigraphic sequences may help in testing the accuracy of conflicting sister taxa hypotheses from different MPTs. The analysis of characters supporting poor-fitting clades can help also to identify and account for characters misleading the phylogenetic signal within data sets.

An Alternative: Comparing Temporal Gaps of Conflicting Sister Taxa Hypotheses

With multiple data sets and a good fossil record for a given group, it seems appropriate to begin a phylogenetic analysis by finding the most parsimonious tree (MPT) or trees(s) for each data set. Then, data set incongruence can be evaluated using bootstrap proportions (Wiens 1998). Specific instances of "strong" incongruence among data sets are identified when conflicting clades are supported by >95% bootstrap proportions (Efron *et al.* 1996). Contrary to conflicting clades supported by <95% bootstrap proportions, "strong" instances of data set incongruence are not the result of undersampling. Instead, different character sets may record discordant evolutionary histories.

"Strong" incongruence can arise from different sources (for a review see Wendel & Doyle, 1998), including violation of the assumptions of maximum parsimony or the use of inaccurate models of character evolution. For instance, it is common to assume that all characters included in an analysis are independent estimators of phylogeny and that there are homogeneous processes through time in all lineages included in a tree (Felsenstein 1985, Hillis & Bull 1993, de Queiroz *et al.* 1995, Efron *et al.* 1996, Lopez *et al.* 2002, Ruano-Rubio & Fares 2007, Kolaczkowski & Thornton 2008, Shavit Grievink *et al.* 2010). Simple factors, such as among lineage differences on life history traits, including generation times (For an example see Jackson *et al.* 2009), age of sexual maturity, average litter size, reproductive strategies, matting systems, among others may affect substitution rates. These differences, as far as I am aware of, are not included in any model of sequence evolution. Assuming that the taxa included in the analysis are comparable on the above factors, subsets of characters may show higher than expected levels of covariation which may deviate from the historical signal in the data set and produce a pattern of "coarse-grained" homoplasy (Naylor & Brown 1998, Shavit Grievink *et al.* 2010). It is also possible that this misleading signal is supported by high bootstrap values (See also Hillis & Bull 1993, Naylor & Brown 1998, Takezaki & Takashi 1999, Ruano-Rubio & Fares 2007). Thus, when separate analyses of data sets show conflicting, well-supported clades, such clades may correspond to either instances in which different character sets record discordant evolutionary histories or instances in which historical signal within character sets is perturbed by a sub-set of linked, or unlinked, misleading characters.

Assessing the accuracy of conflicting sister taxa hypotheses. After specific instances of "strong" incongruence among data sets are identified, the next step is to assess the phylogenetic accuracy of the conflicting hypotheses. For this, temporal gaps are calculated for all well-supported clades in all cladograms under analysis. If fossil information is adequate, confidence intervals on stratigraphic ranges must be calculated using some of the available quantitative methods to correct for the incompleteness of the fossil record (Strauss & Sadler 1989, Springer 1990, Marshall 1991, Solow 1996, Marshall 1997).

After temporal gaps are calculated, the null hypothesis is tested that the temporal gap of each conflicting clade is not significantly longer than the mean temporal gap of all other well-supported clades. Conflicting clades for which the null hypothesis is rejected are called into question and further explored. Rejection of the null hypothesis may result from incompleteness and/or undersampling of the fossil record, or it may suggest that hypothesized sister taxa did not diverge from a common ancestor and that the clade results from the effect of MCs. To determine which of the above alternatives is more likely, the next step is to assess the effect that the characters supporting questioned conflicting clades have on data set homoplasy.

Assessing the effect on homoplasy and accounting for MCs. Before assessing the effect that characters supporting conflicting clades with exceptionally long temporal gaps have on data set homoplasy, these characters must be examined. If an analysis of these characters reveals possible coding errors, then character states must be redefined and data set incongruence must be evaluated again. On the other hand, if evidence of structural, functional or any other kind of constraint acting on those characters is gathered, chances are that conflicting clades result from the effect of MCs.

The next step is resetting to missing values the characters supporting conflicting clades with significantly longer temporal gaps. This is done for the taxa on those clades only. Then, modified data sets are separately analyzed again and new MPTs are calculated. At this point, the hypothesis is tested that resetting suspected MCs to missing values has a significantly stronger effect in reducing homoplasy than randomly resetting to missing values an equal number of character states. If the null hypothesis cannot be rejected for a given data set, we cannot have confidence that conflicting clades in such data set result from the effect of MCs. If the null hypothesis cannot be rejected for any data set, the testing process is stopped and data sets should not be combined. In this case, strong instances of incongruence among data sets are more likely the result of these data recording discordant evolutionary histories. Accepting the null hypothesis also indicates that exceptionally long temporal gaps must result from an incomplete and/or biased fossil record. On the other hand, if the null hypothesis is rejected for at least one data set, it suggests that instances of data set incongruence result from the effect of MCs.

The testing process is iterated until no more conflicting, well-supported clades among MPTs are found. Finally, modified data sets are combined and the MPT is calculated. An application of this approach to the canid phylogeny is presented. Canids were selected because they have a relatively good fossil record and because previous studies (Wayne *et al.* 1997) indicated significant incongruence between two available data sets, one morphological (Tedford *et al.* 1995) and one molecular (Wayne *et al.* 1997).

METHODS

There are several sources of phylogenetic information for the Caninae (Geffen *et al.* 1992, Tedford *et al.* 1995, Wayne *et al.* 1997, Zrzavý & Řičánková 2004, Bardeleben *et al.* 2005; Lindblad-Toh *et al.* 2005, Prevosti 2010), however there is only strong incongruence between two of these sources (Muñoz-Durán & Fuentes, in prep.). The first one includes 2001 base pairs of mitochondrial

DNA (mtDNA) used by Wayne *et al*. (1997) in a phylogenetic analysis of 23 canid species. This sequence includes 729 bp of cytochrome *b* (Cytb), 588 bp of cytochrome *c* oxidase I (COI), and 684 bp of cytochrome *c* oxidase II (COII). Since analyses by Wayne *et al*. (1997) and Bardeleben *et al*. (2005) argued for congruent evolutionary histories among these three mitochondrial sequences, they were included in a single data partition (mtDNA). The second is a data set of 57 morphological characters for 14 extant canid genera studied by Tedford *et al*. (1995).

The analysis was performed at the genus level; however, *Canis adustus* Sundevall 1847 was included separately from the genus *Canis* Linnaeus 1758. The reasons for the above include: 1) most of the living canid genera are monotypic, 2) the monophyletic status of the polytypic genera *Pseudalopex* Burmeister 1856 and *Vulpes* Frisch 1775 have been confirmed by mtDNA and nuclear genes analyses (Geffen *et al.* 1992, Wayne *et al.* 1997, Zrzavý & Řičánková 2004, Bardeleben *et al.* 2005; Lindblad-Toh *et al.* 2005), 3) *Canis* can be considered a coherent unit for analyses on morphological grounds because of the level of similarity among its species (Tedford *et al.* 1995); however, *Canis adustus* renders *Canis* polyphyletic on mtDNA analyses (Wayne *et al.* 1997, Zrzavý & Řičánková 2004).

The selection of different species representing polytypic canid genera could affect the results of this analysis. Therefore, sets of conflicting clades obtained when different species were selected to represent polytypic genera were compared. Complete mtDNA sequences used in this analysis are available for both *Vulpes vulpes* Linnaeus 1758 and *V. macrotis* Merriam 1888, thus the analysis was done once with each one of these species. Separate analyses were done for the genus *Canis* using *C. mesomelas* Schreber 1775 and C. *latrans* Say 1823, which are the most basal and one of the most derived species within this genus, respectively, as suggested by mtDNA studies (Wayne *et al.* 1997). Analyses were also performed using the most basal species within the genus *Pseudalopex* Burmeister 1856, *P. sechurae* Thomas 1900, and one of the most derived species, *P. culpaeus* Molina 1782 (Wayne *et al.* 1997). *Urocyon cinereoargenteus* Schreber 1775 was selected to represent the genus *Urocyon* Baird 1858 because it is the only species in this genus for which the sequence of mtDNA used here is available.

Because the morphological data set of Tedford *et al*. (1995) and the molecular data set of Wayne *et al*. (1997) use different outgroup taxa, the combined tree in this analysis was rooted by combining the outgroup information. The phylogenetic tree of Tedford *et al*. (1995) was rooted using as outgroup taxa Hesperocyoninae, Borophaginae, and *Leptocyon* Matthew 1918. Since Hesperocyoninae is viewed by Tedford *et al*. (1995) as a completely plesiomorphic stem-group for all later canids, character states of this taxon were selected as the outgroup information for morphological traits. Character states of the harbor seal (*Phoca vitulina* Linnaeus 1758) were used as the outgroup information for the mtDNA data.

Separate and combined parsimony analyses of data sets were performed under PAUP 4.0b10 (Swofford 2003) in PowerBook G3. Heuristic search using the options of stepwise-addition, with swap on all if more than one starting tree existed, closest addition sequence, holding of 100 trees at each step, and the tree bisection-reconnection (TBR) swapping algorithm. The transition/transversion ratio was set to 1. Successive weighting (Farris 1969) was used as a tool to identify the most likely MPT when analyses yielded more than one MPT. In other words, when more than one MPT was obtained, characters were reweighted based on their rescaled consistency index (rc) and a new heuristic search was executed, keeping same parameters and options. This was done until the topology was stabilized. This procedure never took more than one reweighting cycle.

Phylogenetic signal value (g1; 1,000,000 random trees) was used as an estimation of the strength of the phylogenetic signal in data set versus random noise (Hillis & Huelsenbeck 1992) for each data partition and for the combined data set. Data set incongruence was assessed using bootstrap values (Felsenstein 1985). Most parsimonious trees for each data set were obtained and bootstrap proportions were calculated using 1000 bootstrap replicates, using the same search conditions as described above. Most parsimonious trees from separate analyses of data sets were examined for conflicts involving sister taxa hypotheses supported by bootstrap proportions higher than 95%. For the sake of data set exploration, maximum likelihood analyses were also performed on the mtDNA. MODELTEST 3.7 (Posada & Buckley 2004) was used to estimate the best-fit model for DNA evolution. The ML tree was calculated using the same search parameters as in the parsimony analyses. I did not insist in systematically explore model based approaches for estimating the phylogeny the Caninae since MCs may be the result of differences in life history traits or of structural, functional and developmental constraints, among other factors that are not accommodated by any model of sequence evolution I am aware of.

Assessing the Accuracy of Conflicting Clades: Comparing Temporal Gaps

Temporal gaps (*T*), the difference between first known times of sister taxa origination, were calculated using information on Table 1. Unfortunately, it was not possible to estimate confidence intervals for the stratigraphic ranges of canids. Available methods not only require information on first and last recorded appearances, but also on all occurrences of the taxa between the first and last known records. This data is difficult to collect for canid taxa with a broad geographical distribution. The limitations are related to the lack of relevant rock outcrops in some areas, to low sampling effort for some taxa and geographical regions, and to the access to existing data that have not been published in international journals, among others.

Table 1. First known times of origination for canid genera. FKTO, first known time of origination.

Taxa	FKTO (my)	Ref.
Vulpes	12	1, 2, 21
Chrysocyon	4.7	1, 3, 4, 21
Speothos	0.3	3, 4, 5,
Cerdocyon	4.4	1, 2, 21
Pseudalopex	2.5	5, 15, 20
Lycaon	2.3	6, 7
Cuon	2.6	8, 9, 10,
Canis	4.6	8, 11, 12, 13, 14, 21
C. adustus	3.2	6
Urocyon	4.3	1, 2
Otocyon	1.9	16
Nyctereutes	6.2	8, 10, 11, 12, 13, 17, 18, 19
Atelocynus	0.01	

References: 1, (Munthe 1998); 2, (Alroy 2000); 3, (Berta 1988); 4, (Berta 1987); 5, (Berta & Marshall 1978); 6, (Turner & Wood 1993); 7, (Turner 1990); 8, (Werdelin & Turner 1996); 9, (Erdbrink 1991); 10, (Alberdi *et al.* 1997); 11, (Rook 1992); 12, (Mein 1989); 13, (Bernor *et al.* 1996); 14, (Werdelin 1996b); 15, (Tonni *et al.* 1992); 16, (Savage 1978); 17, (Werdelin 1996a); 18, (Rook *et al.* 1991);19, (Ginsburg 1998); 20, (Van Valkenburgh 1991); 21, (Tedford *et al.* 2009). There is no known fossil record for *Atelocynus*, thus the age of origination of this genus was set to the Recent (0.01 Myr).

Statistical test

One-tailed t-test was used to examine the following null hypotheses: first, that the temporal gap calculated for each conflicting clade is not significantly longer than the mean temporal gap of all other well-supported clades; second, that the retention index (RI) of the MPT derived from the data sets in which suspected MCs were reset to missing values was not significantly higher than the mean value of this index from the MPTs derived from randomly modified data sets. To test the latter hypothesis 100 replicates of the data sets were produced in which an equal number of characters states (with the same distribution by codon position in the case of the mtDNA data set) as the suspected MCs were randomly reset to missing values. Lastly, a G-test was used to examine the hypothesis that different codon positions and substitution types of the mtDNA genes did not differ in the proportion of suspected misleading characters found on them.

RESULTS

Data Set Incongruence

Phylogenetic analyses of separate morphological and mtDNA data sets led to the most parsimonious trees shown in Figures 1A-B. The topology of these trees is the most stable, not only because they are among the initial set of MPTs derived from the data sets but also because these are the resulting topologies after one cycle of reweighting. Data set incongruence as indicated by bootstrap support of clades led to the identification of the conflicting phylogenetic hypotheses presented in Table 2. This table also presents temporal gaps (*T*) for all clades (conflicting and non-conflicting) supported by BP>95%. Little or no bootstrap support for conflicting nodes in rival trees was found when tables of bootstrap partitions for each data set were examined. Conflicting clades referred to the hypothesized phylogenetic relationships among the wolf like canids genera and two South American genera *Speothos* Lund 1839 and *Chrysocyon* Hamilton-Smith 1839. The sister taxa hypothesis between these last genera (100% BP) was included in the set of conflicting clades since its grouping with *Lycaon* Brookes 1827 is conflictive. The mtDNA suggest that the *Lycaon* lineage and the hypothesized ancestor of *Speothos* and *Chrysocyon* split from a common ancestor about 4.7 Myr. To test this hypothesis it is necessary to test also the hypothesis that *Speothos* and *Chrysocyon* evolved from a common ancestor. In this manner, the internal consistency of the node is tested.

The selection of different species representing polytypic genera in the mtDNA data set did not affect the set of conflicting clades between the MPTs from morphology and mtDNA. The main effect of including different species from polytypic genera was the rearrangement of the phylogenetic relationships among the genera *Vulpes*, *Nyctereutes* Temminck 1839, *Urocyon* and *Otocyon* Müller 1836. However, none of these different phylogenetic arrangements was supported by >95% bootstrap proportions and did not lead to the identification of new conflicting clades. The mtDNA trees shown in this paper were derived from analyses in which *Canis mesomelas, Pseudalopex culpaeus*, and *Vulpes vulpes* represent their corresponding genera.

Accuracy of Conflicting Clades

Comparing Temporal Gaps. The sister taxa hypothesis between *Speothos* and *Chrysocyon* was the only conflicting clade whose temporal gap was significantly longer than the mean temporal gap of all other well-supported clades (t = -2.575, p = 0.012). Thus, the characters supporting this clade were further analyzed.

Table 2. Temporal gaps for all well-supported clades (conflicting and non-conflicting) on trees derived from the independent analyses of data sets. Conflicting clades are identified with an asterisk (*).

Data set	Well-supported Clades	BP%	*T*(my)
Morphology	*(Canis-C. adustus, (Cuon, Lycaon)) **	99	2.0
	*(Cuon, Lycaon) **	99	0.3
	(Urocyon, Otocyon)	99	2.4
	(WLC, SAC+Nyctereutes)	96	1.6
MtDNA	*((((Speothos, Chrysocyon),Lycaon),(Cuon, Canis)), C. adustus) **	99	1.5
	*(((Speothos, Chrysocyon),Lycaon),(Cuon, Canis)) **	98	0.1
	*(Cuon, Canis) **	100	2.0
	*((Speothos, Chrysocyon),Lycaon) **	99	2.4
	*(Speothos, Chrysocyon) **	100	4.4
	((WLC-Speothos-Chrysocyon),(SAC))	100	0.3
	(Cerdocyon,(Atelocynus,Pseudalopex))	100	1.9
	(Atelocynus, Pseudalopex)	98	2.5

BP%, bootstrap proportions. *T*, temporal gap. WLC, Wolf like canids; SAC, South American canids.

Suspected Misleading Characters (MCs). Contrary to expectations that most suspected MCs must be associated with third codon positions, the results of a G-test indicated that there are no significant differences in the proportion of suspected MCs among the three classes of codon positions ($df = 2$, $G = 1.5806$, $p > 0.05$). Although the great majority of suspected MCs corresponds to third positions, normalizing by the number of parsimony informative characters in each of the three codon positions shows that third codon positions have a relative low proportion (10 %) of parsimony informative characters in which suspected MCs were found, only 47 out of 472 (Table 3). First codon positions have the largest proportion (15.1%) of suspected MCs relative to the total number of parsimony informative characters on that codon position.

Suspected MCs in all three codon positions correspond mostly to transitions rather than transversions ($df = 1$, $G = 23.76$, $p < 0.001$). All suspected MCs in first and second codon positions result from transitions, as well as 42 out of 47 suspected MCs in third positions (Table 3). This pattern is in agreement with expectations on the relative frequency of substitution types.

Table 3. Suspected misleading characters (MCs) in the mtDNA data set supporting the *Chrysocyon* - *Speothos* clade.

Data Set	N	NPI	MC	TV	TI	NPI*	NPI-NPI*
MtDNA							
1st	667	73	11	0	11	69	4
2nd	667	14	1	0	1	13	1
3rd	667	472	47	5	42	464	8
Total mtDNA	2001	554	59	5	54	546	13
Morphology	57	40	0				
Combined	2058	581					

N, total number of characters in data set. NPI, number of parsimony informative characters. MC, number of suspected misleading characters that were reset to missing values. TV, transversions. TI, transitions. NPI*, number of parsimony informative characters after correcting for misleading characters. NPI-NPI*, number of substitutions in slow evolving sites. 1st, 2nd, and 3rd stand for first, second, and third codon positions respectively.

About half of the suspected MCs may result from homoplasies in non-synonymous positions and in other slow evolving third codon positions. An inspection of the distribution of character states in other genera besides *Speothos* and *Chrysocyon* showed that in 29 out of 59 instances of suspected MCs, the character for most of the remaining genera was fixed in the ancestral state. In other words, about half of the suspected MCs may have resulted from homoplasies in characters with a slow rate of nucleotide substitution. This is associated with the observed reduction in the number of parsimony informative characters in the mtDNA data set after controlling for misleading characters (Table 3). The reduction in 13 parsimony informative characters corresponds to 13 MCs that were reset to missing values in sites where no more than one other genus, in addition to *Speothos* and *Chrysocyon*, expressed the derived character state. These 13 MCs correspond to four TIs in first position, one TI in second position, and eight TIs in third position.

Suspected MCs in first and second positions were associated with functional and structural constraints of mitochondrial proteins. All suspected MCs on first and second codon positions are related to amino acid replacements involving the hydrophobic residues leucine, isoleucine, valine, and threonine. Furthermore, based on a model of cytochrome *b* (Degli Esposti 1993), it was observed that all but two suspected MCs in first and second positions found in that protein are associated with transmembrane amino-acid residues. The remaining two suspected MCs correspond to first positions and are associated with amino acids on the intermembrane domain.

Assessing the effect on homoplasy and Controlling for Misleading Characters (MCs)

The previous analyses suggested that the *Speothos* - *Chrysocyon* clade was more likely an artifact of MCs in the mtDNA data set. Thus, the characters supporting this clade were reset to missing values in both genera and a new MPT was calculated (Figure 2). Controlling for MCs led to a reduction of only 0.4% in the size of the molecular data set, and an increase of 7% in the strength of its phylogenetic signal (g1 changed from -0.56 to -0.60).

One hundred copies of the mtDNA data set were prepared in which 118 character states (since 59 characters states were reset to missing values for both *Speothos* and *Chrysocyon*) with the same distribution by codon position as the suspected MCs were randomly reset to missing values. These data sets included only parsimony informative characters from the original mtDNA data set.

Resetting to missing values suspected MCs reduced homoplasy in the mtDNA data set in a manner that was significantly different from random. Results of t-tests ($df = 99$) for some tree statistics were as follow: tree length, $t = 17.365$, $p < 0.001$; consistency index, $t = -6.681$, $p <0.001$; retention index, $t = -5.779$, $p <0.001$; rescaled consistency index, $t = -6.631$, $p <0.001$. After controlling for misleading characters, a comparison of new MPT from the modified mtDNA data set with the MPT from the morphological data did not indicate the presence of new conflicting clades. Thus, both data sets were combined and a MPT was calculated.

Combined Analysis

The MPT from the combined morphological and mtDNA (Figure 3) shows the foxes at the base of the Caninae. *Urocyon* is the most basal species followed by *Nyctereutes*. A clade shared by *Otocyon* and *Vulpes* is placed as the sister taxon of a large clade including the wolflike canids and the South American canids. The latter form a polyphyletic group, with *Chrysocyon* at the base of the wolf-like canid clade and *Speothos* in a well supported clade (99% BP) with *Lycaon*. This tree also shows that the trenchant heel evolved once in the Caninae but later was lost in the *Canis* lineage. The same set of relationships is supported by the maximum likelihood tree after controlling for MCs (tree not shown, the best-fit model is GTR+I+G).

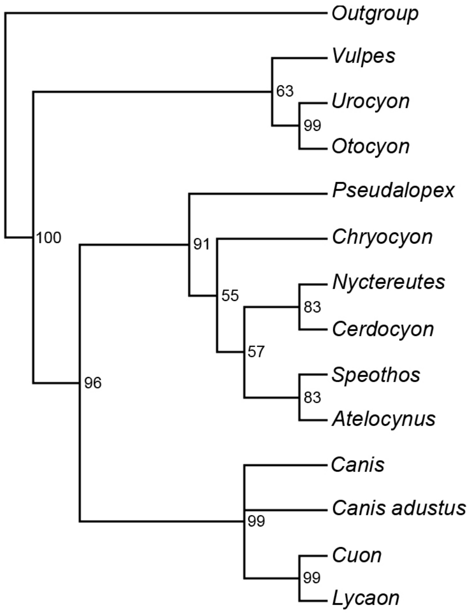

Figure 1A. Most parsimonious tree from morphological data set with bootstrap proportions.

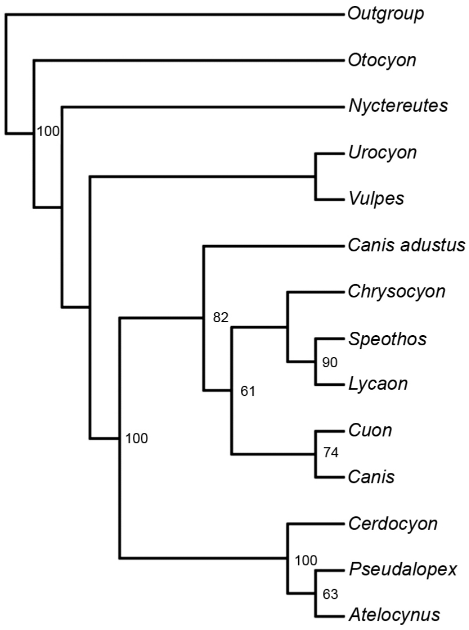

Figure 2. Most parsimonious tree from the modified mtDNA data set (after misleading characters were reset to missing values).

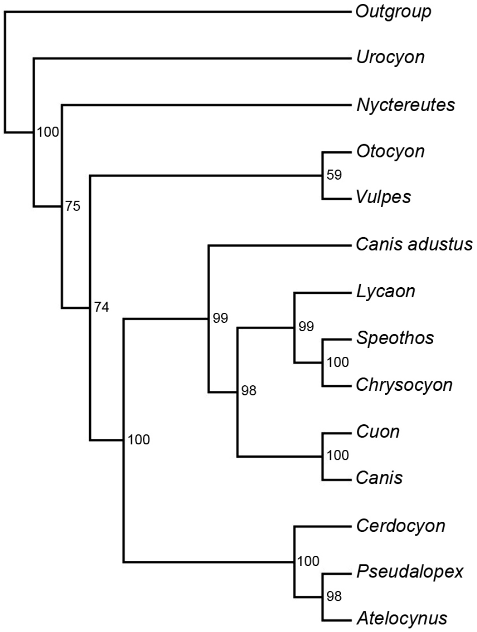

Figure 1B. Most parsimonious tree from mtDNA data set with bootstrap proportions.

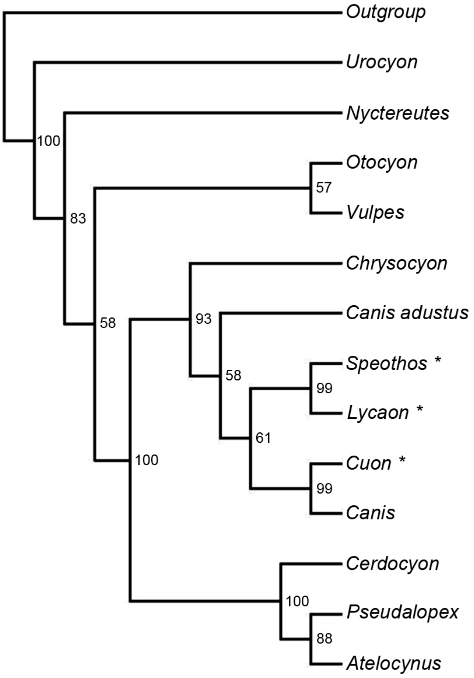

Figure 3. Most parsimonious tree from the combined morphological and mtDNA data set. Asterisks indicate the trenchant heeled species.

DISCUSSION

Incompleteness of the fossil record has been mistakenly equated to inadequacy (Paul 1992, Clyde & Fisher 1997, Paul 1998) of the fossil record in phylogenetic inference and in testing evolutionary hypotheses. However, it has been demonstrated (Norell & Novacek 1992, Benton 1995, Benton & Hitchin 1997, Clyde & Fisher 1997, Hitchin & Benton 1997, Hermsen *et al.* 2006, Rothwell & Nixon 2006, Cobbett *et al.* 2007, Wills *et al.* 2008, Magallón 2010) that fossil data and the stratigraphic sequence of fossil taxa carry a conspicuous phylogenetic signal. It is precisely this expected correspondence between the sequence of taxa origination in the rocks and estimates of phylogeny that has promoted the exploration of more comprehensive approaches to phylogenetic inference.

I use temporal data to identify well-supported, conflicting clades among MPTs that may result from the effect of misleading characters. Because conflicting clades with exceptional long temporal gaps may result either from incompleteness of the fossil record or from the effect of MCs, these clades are rejected only after evidence suggesting the homoplastic nature of the characters that support them is gathered. The main goal of this approach is to use stratigraphic data to identify and control for sources of incongruence among data sets previous to a "total evidence" analysis. This is different from previous methods in which known times of taxa origination are used either to test the reliability of entire phylogenetic trees or as an optimality criterion to estimate phylogenetic trees (Gauthier *et al.* 1988, Fisher 1992, Norell & Novacek 1992, Benton & Storrs 1994, Huelsenbeck 1994, Smith & Littlewood 1994, Wagner 1995, Clyde & Fisher 1997, Huelsenbeck & Rannala 1997, Siddall 1998, Wills 1999, Angielczyk 2002, Pol & Norell 2006, Wills *et al.* 2008). Furthermore, the approach proposed here and the results of this study support the claim (Wiens 2005, Cobbett *et al.* 2007, Magallón 2010) that fossil data, stratigraphic data in this particular case, can break up long branches. The sister taxa relationship between *Speothos* and *Chrysocyon*, implied by the mtDNA data set, arose from the structural and functional constrains. These constraints lead to among-site rate variation, a known cause of long-branch attraction (Poe 2003, Wiens 2005).

Misleading characters

Incongruence between the morphological and mtDNA data sets found in this study may be in part explained by differences in evolutionary rates among these two data sources. That MCs were detected only in the mtDNA data set could be explained by higher evolutionary rates in this data set, which make it more prone to saturation, and therefore homoplasy, in lineages that have long diverged from a common ancestor. This in turn could result in instances of long-branch attraction as the sister taxa relationship between *Speothos* and *Chrysocyon* suggested by the mtDNA data set.

The results of this study suggest that, similarly to systematic biases (Brinkmann *et al.* 2005, Jeffroy *et al.* 2006, Rodríguez-Ezpeleta *et al.* 2007), the presence of MCs in a dataset may result in well supported but inaccurate sister clade hypotheses. Furthermore, it is suggested that controlling for misleading characters must be accomplished after evidence indicating their misleading status has been gathered. A priori weighting schemes, such as giving higher weights to substitutions in slow evolving positions and down-weighting third position substitutions, could have had, if used in the present study, a doubly misleading effect. On the one hand, similarly to recent studies (Hansen *et al.* 2005, Simmons *et al.* 2006, Seo & Kishino 2008), results suggest that third positions are phylogenetically informative and that only a small fraction of all substitutions in this position may distort the phylogenetic signal in the mtDNA data set. This is in agreement with the low rate of saturation found in this mtDNA data set by Wayne *et al*. (1997). In addition to MCs, third codon substitutions may include other homoplasies that do not greatly affect the phylogenetic signal, and substitutions that do have phylogenetic content. More importantly, characters are neither informative nor misleading across all clades in a cladogram. On the contrary, a character can be both informative in some clades and misleading in others. Therefore, down-weighting or eliminating all third codon substitutions in this analysis would had overestimated the amount of MCs and produce a loss of information.

On the other hand, contrary to a priori expectations, results also showed that there are not significant differences in the proportion of MCs among different classes of codon positions. Therefore, not all substitutions in first and second positions will accurately reflect the pattern of descent in the molecular data set. A priori weighting schemes that assign higher weights to substitutions in slow evolving positions could actually enhance the misleading effect of misleading characters in those positions.

The approach proposed in this paper may prove a suitable alternative in identifying and controlling for characters that confuse the historical pattern in a data set, and that may lead to long-branch attraction. The homoplastic condition of the MCs identified through this approach is supported by evidence on the functional and structural constraints of mitochondrial proteins. The results of this analysis showed not only that non-synonymous sites have an important proportion of MCs, but also that those MCs are associated with substitutions between hydrophobic amino acid residues that, in the case of the Cytochrome *b* protein, tend to be located on transmembrane domains. This pattern is in agreement with functional and structural expectations, indicating that although the rate of nucleotide substitution in first and second positions is lower in comparison with the rate at third positions, the probability that nucleotide substitutions in these positions result in homoplastic characters is high. This is due to functional requirements of hydrophobicity that constrain the number of possible character states at first and second positions (Howell 1989, Degli Esposti *et al.* 1993, Naylor *et al.* 1995, Griffiths 1997, Naylor & Brown 1997, Hassanin *et al.* 1998, Naylor & Brown 1998). For instance, it has been found (Naylor *et al.* 1995, Hassanin *et al.* 1998) that the need of conserving hydrophobic properties constrains the type of nucleotide substitution in first positions mainly to A-G transitions and in second positions to C-T transitions.

The results of the present analysis may also be an example of heterotachy (Philippe & Lopez 2001, Lopez *et al.* 2002). This is because MCs affect only the *Chrysocyon* and *Speothos* lineages, and arose from biased codon substitutions that did not affect hydrophobic and functional properties of trasmembrane regions of the cytochrome *b* protein, as it could be expected for proteins that participate on fundamental metabolic processes and are under strong stabilizing selective pressures. Identified MCs indicate among lineage variation on the proportion and position of variable sites of the cytochrome *b* sequence.

The above suggests that paleontological data and the approach outlined in this paper not only allows for testing the phylogenetic accuracy of conflicting phylogenetic hypotheses, but it also may be a more suitable option than a priori weighting schemes in identifying and accounting for characters that confuse the phylogenetic signal in a data set. This approach may prove useful in reducing incongruence among data sets prior to a total evidence analysis.

Phylogeny of Caninae

Although the main objective of this analysis is not to estimate the phylogeny of the Caninae, I would like to comment on the implications of the topology of the resulting cladogram. The combined tree not only optimizes all available morphological and mtDNA characters, but it also conforms to paleontological data. In the combined tree neither the temporal gap for the *Speothos* - *Lycaon* clade (2 my) nor for the *Chrysocyon* - wolflike canid clade (0,1 my) are significantly longer compared to other well-supported clades. The presence of clades with BP <95% indicates that more data is still needed to resolve the exact phylogenetic relationships of some canid taxa, particularly the radiation pattern of basal genera *Urocyon*, *Otocyon*, *Nyctereutes*, and the South American genus *Chrysocyon*.

The combined tree shows *Urocyon* at the base of the radiation of the extant Caninae, followed by *Nyctereutes*, a clade shared by *Otocyon* and *Vulpes* and finally a large clade including all genera of dog-like canids (*Cerdocyon* + *Pseudalopex* + *Atelocynus* + *Chrysocyon* + *Cuon* + *Canis* + *Speothos* + *Lycaon*).

The MPT from the combined data set supports the hypothesis of a polyphyletic origin of the diverse group of South American canids (Geffen *et al.* 1996, Wayne *et al.* 1997, Lindblad-Toh *et al*. 2005, Prevosti 2010). Coincident with the results of recent analyses (Zrzavý & Řičánková 2004, Bardeleben *et al.* 2005), there is still uncertainty on the phylogenetic position of *Speothos* and *Chrysocyon*. Analyses including larger morphological, behavioral, and molecular data combined with paleontological information are needed to resolve long standing questions related to the evolutionary affinities of these South American canids. Particularly, the possibility that *Speothos* may belong to a clade of hypercarnivorous canids, together with *Lycaon* and *Cuon*, has received support from recent analyses of dental and osteological characters (Prevosti 2010). This is a hypothesis that deserves further consideration.

The tree indicates the successful colonization of South America by four canid lineages: *Chrysocyon, Speothos, Urocyon* and the common ancestor of remaining South American genera. Although the genus *Canis* does not inhabit in South America at the present time, there is a fossil record for this genus in Plio-Pleistocene formations in this continent (Prevosti, 2010), which indicates at least five independent colonization events of canid lineages into South America.

The resulting tree also shows that the wolf-like canids form a clade together with *Speothos* and *Chrysocyon.* The position of *C. adustus* still renders the genus *Canis* as polyphyletic, a feature that is common to most phylogenetic analyses of the Caninae (Wayne *et al.* 1997, Zrzavý & Řičánková 2004, Lindblad-Toh *et al*. 2005). The results of the present analysis support previous suggestions of including *C. adustus* in a separate genus *Lupulella* Hilzheimer 1906 (Zrzavý & Řičánková 2004).

The trenchant heel, a specialized cutting blade on the lower carnassial molar, represents an adaptation to increasing carnivory (Van Valkenburgh 1990) and is present in all social canines except *Canis lupus* Linnaeus 1758. The results of the present analysis suggest that this structure evolved only once in the common ancestor of the wolf-like canids and *Speothos*, but that it was later lost in the lineage leading to *Canis.* This is a simpler hypothesis compared to previous hypotheses suggesting that the trenchant heel has a single origin followed by multiple losses (Wayne *et al.* 1997), or that it evolved independently in two lineages and was later lost in one lineage (Tedford *et al.* 1995, Geffen *et al.* 1996, Wayne *et al.* 1997, Bardeleben *et al.* 2005, Lindblad-Toh *et al*. 2005) or that it has three independent origins and several loses (Clutton-Brock *et al.* 1976, Wayne *et al.* 1997).

CONCLUDING REMARKS

Paleontology provides information on approximate times of taxa origination which constitutes independent evidence on the evolutionary history of a group. These hints from the fossil record can be used to test the accuracy of conflicting cladistic hypotheses, and to identify and account for characters which confuse the pattern of shared ancestry in a data set. The approach proposed in this paper leads to a more comprehensive approach to phylogenetic inference through the incorporation of fossil data in the search for more accurate phylogenies. The application of this procedure to the phylogeny of the Caninae allowed a reduction in the level of incongruence among the morphological and mtDNA data sets, and led to a better tuning of their common phylogenetic signal. However, additional molecular and/or morphological data is still necessary to have a better estimate of the phylogenetic relationships of various canid taxa. Further phylogenetic analysis of other groups with a well preserved fossil record may reveal the extent to which this approach could be applied, as well as other limitations for its application.

ACKNOWLEDGMENTS

I am grateful to Jesualdo Arturo Fuentes and the three referees of this manuscript for their suggestions and thoughtful criticism of the manuscript. I thank Dorina Apahidean for her editorial assistance. I also thank the Colombian Agency for the Advance of Science and Technology-COLCIENCIAS (contract 192-2007) and the División de Investigaciones at the Universidad Nacional de Colombia - Sede Bogotá (DIB, Project 8373), for their financial support.

LITERATURE CITED

Alberdi, M. T., B. Azanza, E. Cerdeño & J. L. Prado. 1997. Similarity relationship between Mammal faunas and bichronology from Latest Miocene to Pleistocene in the Western Mediterranean area. Eclogae Geologicae Helvetiae 90: 115-132.

Alroy, J. 2000. Understanding the dynamics of trends within evolving lineages. Paleobiology 26: 319-329.

Ané, C., B. Larget, D. A. Baum, S. D. Smith & A. Rokas. 2007. Bayesian Estimation of Concordance among Gene Trees. Molecular Biology and Evolution 24: 412-426.

Angielczyk, K. D. 2002. A character-based method for measuring the fit of a cladogram to the fossil record. Systematic Biology 51: 176-191.

Asher, R. J., M. J. Novacek & J. H. Geisler. 2003. Relationships of Endemic African Mammals and Their Fossil Relatives Based on Morphological and Molecular Evidence. Journal of Mammalian Evolution 10: 131-194.

Azzaroli, A. 1995. The “Elephant-Equus” and the “End-Villafranchian” events in Eurasia. En: E.S. Vrba, G.H. Denton, T.C. Patridge & L.H. Burckle eds.). *Paleoclimate and evolution, with emphasis on human evolution*: 311-318. Yale University Press. New Haven and London.

Ballard, J. W. O., M. K. Thayer, A. F. Newton & E. R. Grismer. 1998. Data sets, partitions, and characters: philosophies and procedures for analyzing multiple data sets. Systematic Biology 47: 367-396.

Bardeleben, C., R. L. Moore & R. K. Wayne. 2005. A molecular phylogeny of the Canidae based on six nuclear loci. Molecular Phylogenetics and Evolution 37: 815-831.

Benton, M. J. 1995. Testing the time axis of phylogenies. Philosophical Transactions of the Royal Society of London B Biological Sciences 349: 5-10.

Benton, M. J. & R. Hitchin. 1997. Congruence between phylogenetic and stratigraphic data on the history of life. Proceedings of the Royal Society Biological Sciences Series B 264: 885-890.

Benton, M. J. & G. W. Storrs. 1994. Testing the quality of the fossil record: paleontological knowledge is improving. Geology 22: 111-114.

Bernor, R. L., V. Fahlbusch, P. Andrews, H. De Bruijn, M. Fortelius, F. Rogl, F. F. Steininger & L. Werdelin. 1996. The evolution of Western Eurasian Neogene Mammal faunas: a chronologic, systematic, biogeographic, and paleoenvironmental synthesis. En: L. Bernor, V. Fahlbusch & H.-W. Mittmann eds.). *The evolution of Western Eurasian Neogene Mammal Faunas.*: 449-469. Columbia University Press. New York.

Berta, A. 1987. Origin, diversification, and zoogeography of the South American Canidae. Fieldiana Zoology. 39: 455-472.

Berta, A. 1988. Quaternary evolution and biogeography of the large South American Canidae (Mammalia: Carnivora). University of California Publications on Geological Sciences 132: 1-160.

Berta, A. & L. Marshall. 1978. South American Carnivora. En: F. Westphal (ed. *Fossilium Catalogus l : Animalia*: 1-48. Dr. W. Junk b.v. Publishers. The Hague/ Boston/London.

Brinkmann, H., M. van der Giezen, Y. Zhou, G. t. Poncelin de Raucourt & H. Philippe. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Systematic Biology 54: 743-757.

Brochu, C. A. 1997. Morphology, fossils, divergence timing, and the phylogenetic relationships of Gavialis. Systematic Biology 46: 479-522.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford & P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384-397.

Clutton-Brock, J., G. B. Corbett & M. Hills. 1976. A review of the family Canidae with a classification by numerical methods. Bulletin of the British Museum (Natural History), Zoology 119-199.

Clyde, W. C. & D. C. Fisher. 1997. Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis. Paleobiology 23: 1-19.

Cobbett, A., M. Wilkinson & M. A. Wills. 2007. Fossils Impact as Hard as Living Taxa in Parsimony Analyses of Morphology. Systematic Biology 56: 753-766.

de Queiroz, A. 1993. For consensus (sometimes). Systematic Biology 42:368-372.

de Queiroz, A., M. J. Donoghue & K. Junhyong. 1995. Separate versus combined analysis of phylogenetic evidence. Annual Review of Ecology and Systematics 26: 657-681.

Degli Esposti, M., S. De Vries, M. Crimi, A. Ghelli, T. Patarnello & A. Meyer. 1993. Mitochondrial cytochrome *b*: evolution and structure of the protein. Biochimica et Biophysica Acta 1143: 243-271.

Donoghue, M. J., J. A. Doyle, J. Gauthier, A. G. Kluge & T. Rowe. 1989. The importance of fossils in phylogenetic reconstruction. Annual Review of Ecology and Systematics 20: 431-460.

Edwards, S. V., L. Liu & D. K. Pearl. 2007. High-Resolution Species Trees without Concatenation. Proceedings of the National Academy of Sciences of the United States of America 104: 5936-5941.

Eernisse, D. J. & A. G. Kluge. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Molecular Biology and Evolution 10: 1170-1195.

Efron, B., E. Halloran & S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. Proceedings of the National Academy of Sciences of the United States of America 93: 7085-7090.

Erdbrink, D. P. B. 1991. Dogs long gone: some fossil postcranial skeletal remains. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Biological Chemical Physical and Medical Sciences 94: 281-298.

Farris, J. S. 1969. A successive approximation approach to character weighting. Systematic Zoology 18: 374-385.

Farris, J. S., M. Källersjö, A. G. Kluge & C. Bult. 1995. Constructing a significance test for incongruence. Systematic Biology 44: 570-572.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.

Fisher, D. C. 1988. Stratocladistics: integrating stratigraphic and morphologic data in phylogenetic inference. Geological Society of America 20:A186.

Fisher, D. C. 1992. Stratigraphic Parsimony. En: W.P. Maddison & D.R. Maddison eds.). *MacClade: analysis of phylogeny and character evolution, Version 3*: 124-129. Sinauer Associates, Inc. Sunderland, Massachusetts.

Gauthier, J. A., A. G. Kluge & T. Rowe. 1988. Amniote phylogeny and the importance of fossils. Cladistics 4: 105-209.

Geffen, E., M. E. Gompper, J. L. Gittlemann, H.-K. Luh, D. W. Macdonal & R. K. Wayne. 1996. Size, life-history traits, and social organization in the Canidae: a reevaluation. American Naturalist 147: 140-160.

Geffen, E., A. Mercure, D. J. Girman, D. W. Macdonal & R. K. Wayne. 1992. Phylogenetic relationships of the fox-like canids: mitochondrial DNA restriction fragment, site and cytochrome *b* sequence analyses. Journal of Zoology (London) 27-39.

Ginsburg, L. 1998. Le gisement de vertébe's pliocenes de Calta, Ankara, Turquie.5.Carnivores. Geodiversitas 20: 379-396.

Griffiths, C. S. 1997. Correlation of functional domains and rates of nucleotide substitution in Cytochrome *b*. Molecular Phylogenetics and Evolution 7: 352-365.

Hansen, K., K. F. LoBuglio & D. H. Pfister. 2005. Evolutionary relationships of the cup-fungus genus Peziza and Pezizaceae inferred from multiple nuclear genes: RPB2, [beta]-tubulin, and LSU rDNA. Molecular Phylogenetics and Evolution 36: 1-23.

Hassanin, A., G. Lecointre & S. Tillier. 1998. The 'evolutionary signal' of homoplasy in protein coding gene sequences and its consequences for a priori weighting in phylogeny. Comptes Rendus de l'Academie des Sciences Serie III Sciences de la Vie 321: 611-620.

Hermsen, E., K. Nixon & W. Crepet. 2006. The impact of extinct taxa on understanding the early evolution of Angiosperm clades: an example incorporating fossil reproductive structures of Saxifragales. Plant Systematics & Evolution 260: 141-169.

Hillis, D. M. & J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182-192.

Hillis, D. M. & J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. Journal of Heredity 83: 189-195.

Hillis, D. M., B. K. Mable & C. Moritz. 1996. Applications of molecular systematics: the state of the field and a look of the future. En: D.M. Hillis, C. Moritz & B.K. Mable eds.). *Molecular Systematics*: 515-547. Sinauer Associates. Inc. Sunderland, Massachusetts.

Hitchin, R. & M. J. Benton. 1997. Congruence between parsimony and stratigraphy: Comparisons of three indices. Paleobiology 23: 20-32.

Howell, N. 1989. Evolutionary conservation of protein regions in the protonmotive cytochrome *b* and their possible roles in redox catalysis. Journal of Molecular Evolution 29: 157-169.

Huelsenbeck, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology 20: 470-483.

Huelsenbeck, J. P. & J. J. Bull. 1996. A likelihood ratio test for detection of conflicting phylogenetic signal. Systematic Biology 45: 92-98.

Huelsenbeck, J. P. & B. Rannala. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. Paleobiology 23: 174-180.

Jackson, J. A., C. S. Baker, M. Vant, D. J. Steel, L. Medrano-GonzÃ¡lez & S. R. Palumbi. 2009. Big and Slow: Phylogenetic Estimates of Molecular Evolution in Baleen Whales (Suborder Mysticeti). Molecular Biology and Evolution 26: 2427-2440.

Jeffroy, O., H. Brinkmann, F. Delsuc & H. Philippe. 2006. Phylogenomics: the beginning of incongruence? Trends in Genetics 22: 225-231.

Kishino, H. & M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution 29: 170-179.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Systematic Zoology 7-25.

Kolaczkowski, B. & J. W. Thornton. 2008. A Mixed Branch Length Model of Heterotachy Improves Phylogenetic Accuracy. Molecular Biology and Evolution 25: 1054-1066.

Lapointe, F.-J. 1998. How to validate phylogenetic trees? A stepwise procedure. En: C. Hayashi, H.H. Bock, K. Yajima, Y. Tanaka, N. Oshumi & Y. Baba eds.). *Data science, classifcation, and related methods: Studies in classifcation, data analysis, and knowledge optimization*: 71-88. Springer-Verlag. Tokyo.

Liang, L. & D. K. Pearl. 2007. Species Trees from Gene Trees: Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. Systematic Biology 56:504-514.

Lindblad-Toh K., C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, M. Kamal, M. Clamp, J.L.Chang, E.J. Kulbokas, M.C. Zody, E. Mauceli, X. Xie, M. Breen, R.K.Wayne, E.A. Ostrander, C.P. Ponting, F. Galibert, D.R. Smith, P.J. Dejong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C-W. Chin, A. Cook, J. Cuff, M.J. Daly, D. Decaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K-P. Koepfli, H.G. Parker, J.P. Pollinger, S.M.J. Searle, N.B. Sutter, R. Thomas, C. Webber, J. Baldwin, A. Abebe, A. Abouelleil, L. Aftuck, M. Ait-Zahra, T. Aldredge, N. Allen, P. An, S. Anderson, C. Antoine, H. Arachchi, A. Aslam, L. Ayotte, P. Bachantsang, A. Barry, T. Bayul, M. Benamara, A. Berlin, D. Bessette, B. Blitshteyn, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, A. Brown, P. Cahill, N. Calixte, J. Camarata, Y. Cheshatsang, J. Chu, M. Citroen, A. Collymore, P. Cooke , T. Dawoe, R. Daza, K. Decktor, S. Degray, N. Dhargay, K. Dooley, K. Dorjee K, L. Dorris, N. Duffey, A. Dupes, O. Egbiremolen, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, P. Ferreira, S. Fisher, M. Fitzgerald, K. Foley, C. Foley, A. Franke, D. Friedrich, D. Gage, M. Garber, G. Gearin, G. Giannoukos, T. Goode, A. Goyette, J. Graham, E. Grandbois, K. Gyaltsen, N. Hafez, D. Hagopian, B. Hagos, J. Hall, C. Healy, R. Hegarty, T. Honan, A. Horn, N. Houde, L. Hughes, L. Hunnicutt, M. Husby, B. Jester, C. Jones, A. Kamat, B. Kanga, C. Kells, D. Khazanovich, A.C. Kieu, P. Kisner, M. Kumar, K. Lance, T. Landers, M. Lara, W. Lee, J-P. Leger, N. Lennon, L. Leuper, S. Levine, J. Liu, X. Liu, Y. Lokyitsang, T. Lokyitsang, A. Lui, J. Macdonald, J. Major, R. Marabella, K. Maru, C. Matthews, S. Mcdonough, T. Mehta, J. Meldrim, A. Melnikov, L.

Meneus, A. Mihalev, T. Mihova, K. Miller, R. Mittelman, V. Mlenga, L. Mulrain, G. Munson, A. Navidi, J. Naylor, T. Nguyen, N. Nguyen, C. Nguyen, T. Nguyen, R. Nicol, N. Norbu, C. Norbu, N. Novod, T. Nyima, P. Olandt, B. O'neill, K. O'neill, S. Osman, L. Oyono, C. Patti, D. Perrin, P. Phunkhang, F. Pierre, M. Priest, A. Rachupka, S. Raghuraman, R. Rameau, V. Ray V, C. Raymond, F. Rege, C. Rise , J. Rogers, P. Rogov, J. Sahalie, S. Settipalli, T. Sharpe, T. Shea, M. Sheehan, N. Sherpa, J. Shi, D. Shih, J. Sloan, C. Smith, T. Sparrow, J. Stalker, N. Stange-T., S. Stavropoulos, C. Stone, S. Stone, S. Sykes, P. Tchuinga, P. Tenzing, S. Tesfaye, D. Thoulutsang, Y. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, V. Venkataraman, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, S. Yang, X. Yang, G. Young, Q. Yu, J. Zainoun, L. Zembek, A. Zimmer & E.S. Lander 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438(7069): 803-819.

Lopez, P., D. Casane & H. Philippe. 2002. Heterotachy, an Important Process of Protein Evolution. Molecular Biology and Evolution 19: 1-7.

Maddison, W. P. & L. L. Knowles. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. Systematic Biology 55: 21-30.

Magallón, S. 2010. Using Fossils to Break Long Branches in Molecular Dating: A Comparison of Relaxed Clocks Applied to the Origin of Angiosperms. Systematic Biology 59: 384-399.

Marshall, C. R. 1991. Estimation of taxonomic ranges from the fossil record. En: N. Gilinsky & P.W. Signor eds.). *Analytical Paleobiology*: 19-38. University of Tennessee. Knoxville.

Marshall, C. R. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. Paleobiology 23: 165-173.

Marshall, L. G. 1985. Geochronology and Land-Mammal biochronology of the transamerican faunal interchange. En: F.G. Stehli & D.S. Webb eds.). *The great American biotic interchange*: 49-85. Plenum Press. New York.

Martin, A. P. 1995. Mitochondrial DNA sequence evolution in sharks: rates, patterns, and phylogenetic inferences. Molecular Biology and Evolution 12: 1114-1123.

McKee, J. K. 1991. Palaeo-ecology of the Sterkfontein hominids: a review and synthesis. Palaeontologia Africana 28: 41-51.

Mein, P. 1989. Updating of MN Zones. En: E.H. Lindsay, V. Fahlbusch & P. Mein eds.). *European Neogene Mammal Chronology*: 73-90. Plenum Press and NATO Scientific Affairs Division. New York and London.

Mickevich, M. F. 1978. Taxonomic congruence. Systematic Zoology 27: 143-158.

Miyamoto, M. M. & W. M. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Systematic Biology 44: 64-76.

Munthe, K. 1998. Canidae. En: C.M. Janis, K.M. Scott & L.L. Jacobs eds.). *Evolution of Tertiary mammals of North America.*: 124-143. Cambridge University Press. Cambridge.

Naylor, G. J. P. & D. C. Adams. 2001. Are the fossil data really at odds with the molecular data? Morphological evidence for cetartiodactyla phylogeny reexamined. Systematic Biology 50: 444-453.

Naylor, G. J. P. & E. M. Brown. 1997. Structural biology and phylogenetic estimation. Nature 388: 527-528.

Naylor, G. J. P. & E. M. Brown. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of Inference based on comparisons of sequences. Systematic Biology 47: 61-76.

Naylor, G. J. P., T. M. Collins & W. M. Brown. 1995. Hydrophobicity and phylogeny. Nature 373: 565-566.

Norell, M. A. & M. J. Novacek. 1992. Congruence between superpositional and phylogenetic patterns: comparing cladistic patterns with fossil records. Cladistics 8: 319-337.

Novacek, M. J. 1992. Fossils as critical data for phylogeny. En: M.J. Novacek & Q.D. Wheeler eds.). *Extinction and phylogeny*: 46-88. Columbia University Press. New York.

Novacek, M. J. 1996. Paleontological data and the study of adaptation. En: M.R. Rose & G.V. Lauder eds.). *Adaptation*: 311-359. Academic Press. San Diego, California.

O'Keefe, R. F. & M. P. Sander. 1999. Paleontological paradigms and inferences of phylogenetic pattern: a case study. Paleobiology 25: 518-533.

Opdyke, N. D. 1995. Mammalian migration and climate over the last seven million years. En: E.S. Vrba, G.H. Denton, T.C. Patridge & B.L. H. eds.). *Paleoclimate and evolution, with emphasis on human evolution*: 109-114. Yale University Press. New Haven and London.

Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. Annual Review of Ecology and Systematics 12: 195-223.

Paul, C. R. C. 1992. How complete does the fossil record have to be? Revista Española de Paleontología 7: 127-133.

Paul, C. R. C. 1998. Adequacy, completeness and the fossil record. En: S.K. Donovan & C.R.C. Paul eds.). *The adequacy of the fossil record*: 1-22. John Wiley & Sons Ltd. Sussex.

Philippe, H. & P. Lopez. 2001. On the conservation of protein sequences in evolution. Trends in Biochemical Sciences 26: 414-416.

Poe, S. 2003. Evaluation of the Strategy of Long-Branch Subdivision to Improve the Accuracy of Phylogenetic Methods. Systematic Biology 52: 423.

Pol, D. & M. A. Norell. 2006. Uncertainty in the Age of Fossils and the Stratigraphic Fit to Phylogenies. Systematic Biology 55: 512-521.

Posada, D. & T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Systematic Biology 53: 793-808.

Prevosti, F.J. 2010. Phylogeny of the large extinct South American Canids (Mammalia, Carnivora, Canidae) using a "total evidence" approach. Cladistics 26: 456-481.

Pyron, A.R. 2010. A Likelihood Method for Assessing Molecular Divergence Time Estimates and the Placement of Fossil Calibrations. Systematic Biology 59: 185-194.

Rodrigo, A. G., M. Kelly-Borges, P. R. Bergquist & P. l. Bergquist. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. New Zealand Journal of Botany 31: 257-268.

Rodríguez-Ezpeleta, N., H. Brinkmann, B. Roure, N. Lartillot, B. F. Lang & H. Philippe. 2007. Detecting and Overcoming Systematic Errors in Genome-Scale Phylogenies. Systematic Biology 56: 389-399.

Rook, L. 1992. Canis monticinensis, new species, a new Canidae (Carnivora, Mammalia) from the late Messinian of Italy. Bollettino della Societa Paleontologica Italiana 31: 151-156.

Rook, L., G. Ficcarelli & D. Torre. 1991. Messinian carnivores from Italy. Bollettino della Societa Paleontologica Italiana 30: 7-22.

Rothwell, G. W. & K. C. Nixon. 2006. How does the inclusion of fossil data change our conclusions about the phylogenetic history of Euphyllophytes? International Journal of Plant Sciences 167: 737-749.

Ruano-Rubio, V. & M. A. Fares. 2007. Artifactual Phylogenies Caused by Correlated Distribution of Substitution Rates among Sites and Lineages: The Good, the Bad, and the Ugly. Systematic Biology 56: 68-82.

Santini, F. & J. C. Tyler. 2003. A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to Recent. Zoological Journal of the Linnean Society 565-617.

Savage, R. J. G. 1978. Carnivora. En: V.J. Maglio & H.B.S. Cook eds.). *Evolution of African Mammals*: 249-267. Harvard University Press. Cambridge Massachusetts.

Seo, T.-K. & H. Kishino. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Systematic Biology 57: 367-377.

Shavit Grievink, L., D. Penny, M. D. Hendy & B. R. Holland. 2010. Phylogenetic Tree Reconstruction Accuracy and Model Fit when Proportions of Variable Sites Change across the Tree. Systematic Biology 59: 288-297.

Siddall, M. E. 1998. Stratigraphic fit to phylogenies: A proposed solution. Cladistics 14: 201-208.

Simmons, M. P., Z. Li-Bing, C. T. Webb & A. Reeves. 2006. How Can Third Codon Positions Outperform First and Second Codon Positions in Phylogenetic Inference? An Empirical Example from the Seed Plants. Systematic Biology 55: 245-258.

Smith, A. B. 1998. What does palaeontology contributes to systematics in a molecular world? Molecular Phylogenetics and Evolution 9: 437-447.

Smith, A. B. & D. T. J. Littlewood. 1994. Paleontological data and molecular phylogenetic analysis. Paleobiology 20: 259-273.

Solow, A. R. 1996. Tests and confidence intervals for a common upper endpoint in fossil taxa. Paleobiology 22: 406-410.

Springer, M. S. 1990. The effect of random range truncations on patterns of evolution in the fossil record. Paleobiology 16: 512-520.

Strauss, D. & P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology 211: 411-427.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Takezaki, N. & G. Takashi. 1999. Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. Molecular Biology and Evolution 16: 590-601.

Tedford, R. H., B. E. Taylor & X. Wang. 1995. Phylogeny of the Caninae (Carnivora: Canidae): the living taxa. Am Mus Novit. 1-37.

Tedford, R. H., X. M. Wang & B. E. Taylor. 2009. Phylogenetic systematics of the north american fossil caninae (Carnivora: Canidae). Bulletin of the American Museum of Natural History. CP1-218.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37: 221-244.

Tonni, E. P., M. T. Alberdi, J. L. Prado, M. S. Bargo & A. L. Cione. 1992. Changes in mammal assemblages in the Pampean region (Argentina) and their relation with the Plio-Pleistocene boundary. Palaeogeography Palaeoclimatology Palaeoecology 95: 179-194.

Turner, A. 1990. The evolution of the guild of larger terrestrial carnivores during the Plio-Pleistocene in Africa. Geobios 349-368.

Turner, A. 1997. Further remains of Carnivora (Mammalia) from the Sterkfontein Hominid site. Palaeontologia Africana. 34:115-126.

Turner, A. & B. Wood. 1993. Taxonomic and geographic diversity in robust australopithecines and other African Plio-Pleistocene larger mammals. Journal of Human Evolution 24: 147-168.

Van Valkenburgh, B. 1990. Carnivore dental adaptations and diet: A study of trophic diversity in guilds. En: J.L. Gittleman (ed. *Carnivore behavior, ecology and evolution*: 410-436. Cornell University Press. Ithaca, New York.

Van Valkenburgh, B. 1991. Iterative evolution of hypercarnivory in canids (Mammalia: Carnivora): evolutionary interactions among sympatric predators. Paleobiology 17: 340-362.

Vrba, E. S. 1995. The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate. En: E.S. Vrba, G.H. Denton, T.C. Patridge & B.L. H. eds.). *Paleoclimate and evolution, with emphasis on human evolution*: 385-424. Yale University Press. New Haven and London.

Wagner, P. J. 1995. Stratigraphic test of cladistic hypotheses. Paleobiology 21: 153-178.

Wayne, R. K., E. Geffen, D. J. Girman, K. P. Koepfli, L. M. Lau & C. R. Marshall. 1997. Molecular systematics of the Canidae. Systematic Biology 46: 622-653.

Wendel, J.F. & J.J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. En: D.E. Soltis, P.S. Soltis & J.J. Doyle (eds.). *Molecular systematics of plants II - DNA sequencing*: 265-296. Norwell, Massachusetts, Kluwer Academic Publishers.

Wenzel, J. W. & M. E. Siddall. 1999. Noise. Cladistics 15: 51-64.

Werdelin, L. 1996a. Carnivores, exclusive of Hyaenidae, from the Later Miocene of Europe and Western Asia. En: L. Bernor, V. Fahlbusch & H.-W. Mittmann eds.). *The evolution of Western Eurasian Neogene Mammal Faunas*: 271-289. Columbia University Press. New York.

Werdelin, L. 1996b. The history of the felid classification. En: K. Nowell & P. Jackson eds.). *Wild Cats*: xviii-xxiii. IUCN. Gland, Switzerland.

Werdelin, L. & A. Turner. 1996. Turnover in the guild of larger carnivores in Eurasia across the Miocene-Pliocene boundary. Acta Zoologica Cracoviensia 39: 585-592.

Wheeler, W. C., P. Cartwright & C. Y. Hayashi. 1993. Arthropod phylogeny: a combined approach. Cladistics. 1-39.

Wiens, J. J. 1998. Combining data sets with different phylogenetic histories. Systematic Biology 47: 568-581.

Wiens, J. J. 2005. Can Incomplete Taxa Rescue Phylogenetic Analyses from Long-Branch Attraction? Systematic Biology 54: 731-742.

Wills, M. A. 1999. Congruence between phylogeny and stratigraphy: randomization tests and the gap excess ratio. Systematic Biology 48: 559-580.

Wills, M. A., P. M. Barrett & J. F. Heathcote. 2008. The Modified Gap Excess Ratio (GER*) and the Stratigraphic Congruence of Dinosaur Phylogenies. Systematic Biology 57: 891-904.

Zrzavý, J. & V. Řičánková. 2004. Phylogeny of Recent Canidae (Mammalia, Carnivora): relative reliability and utility of morphological and molecular datasets. Zoologica Scripta 33: 311-333.

Recibido: 01/02/2011
Aceptado: 15/09/2011