Zoosyst. Evol. 85 (2) 2009, 277–292 / DOI 10.1002/zoos.200900005

WILEY-VCH

museum für naturkunde

A comprehensive phylogeny of extant horses, rhinos and tapirs (Perissodactyla) through data combination

Samantha A. Price[1] and Olaf R. P. Bininda-Emonds*,[2]

[1] National Evolutionary Synthesis Center (NESCent), 2024 W. Main Street Suite A200, Durham, NC 27705, U.S.A.
[2] AG Systematik und Evolutionsbiologie, IBU – Fakultät V, Carl von Ossietzky Universität Oldenburg, Carl von Ossietzky Str. 9–11, 26111 Oldenburg, Germany

Received 14 July 2008
Accepted 21 November 2008
Published 24 September 2009

Abstract

We present the first phylogenies to include all extant species of Perissodactyla (odd-toed hoofed mammals) and the recently extinct quagga (*Equus quagga*). Two independent data sets were examined; one based on multiple genes and analyzed using both supertree and supermatrix approaches, and a second being a supertree constructed from trees collected from the scientific literature. All methods broadly confirmed the traditional view of perissodactyl interfamily relationships, with Equidae (= Hippomorpha) forming the sister-group to the clade Rhinocerotidae + Tapiridae (= Ceratomorpha). The contentious affinity of the Sumatran rhino (*Dicerorhinus sumatrensis*) is resolved in favour of it forming a clade with the two Asian rhinos (genus *Rhinoceros*). However, no data set or tree-building method managed to satisfactorily resolve the historically contentious relationships among extant equids; little agreement appears among the different trees for this group. In general, both the supertree and supermatrix approaches performed equally well, but both were hindered by the current paucity of data (e.g. no single gene has been sequenced to date for all 17 species) and its patchy distribution within Equidae. More data, both molecular and morphological, are required for all species to resolve the poorly supported nodes.

Key Words

Supertree
Supermatrix
Systematics
Global congruence

Introduction

The 17 extant species of perissodactyl (odd-toed hoofed mammals) are the relicts of a once large, diverse and widespread clade, members of which first appeared in the fossil record in the upper Palaeocene (Radinsky 1969). Molecular data, however, estimate that Perissodactyla diverged from its sister-group much earlier in the late Cretaceous: either from Cetartiodactyla 83.4 ± 0.7 (Bininda-Emonds et al. 2007) or 97.5–88.8 million years ago (Eizirik et al. 2001) or from Carnivora approximately 80 million years ago (Springer et al. 2003). Despite the severe decline in their species diversity, perissodactyls remain important for ecosystem function (e.g. Fragoso & Huffman 2000), and have also played an important role in human history and culture. The domestication of equids approximately 5,000 years ago revolutionized transportation and warfare by providing a swift and efficient way to move people and products over large distances (Vilà et al. 2001).

When originally described by Owen (1848), Perissodactyla included four extant families: Rhinocerotidae (rhinos), Tapiridae (tapirs), Equidae (horses, asses and zebras) and Hyracoidea (hyraxes). Shortly thereafter, hyraxes were elevated to a separate order, Hyracoidea (Huxley 1869). Despite isolated attempts to revive Owen's original definition of Perissodactyla (e.g. Prothero & Schoch 1989a, b, 2002), recent molecular evidence (e.g. Madsen et al. 2001; Murphy et al. 2001) overwhelmingly groups hyraxes with proboscideans (elephants) and sireniens (dugongs and manatees) in Afrotheria, whereas perissodactyls (*sensu stricto*) are placed within Laurasiatheria, often as the sister-taxon to Cetartiodactyla (even-toed hoofed mammals including whales) (see also Beck et al. 2006). Accordingly, this paper does not include the hyraxes within Perissodactyla.

The three extant perissodactyl families are divided traditionally into the two suborders Ceratomorpha (rhinos and tapirs) and Hippomorpha (horses, assess and zebras) (Wood 1937), which diverged from one another

* Corresponding author, e-mail: olaf.bininda@uni-oldenburg.de

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

WILEY InterScience® DISCOVER SOMETHING GREAT

prior to the early Eocene (e.g. Prothero & Schoch 1989a). Within each of the three families, however, there is little or no consensus concerning the species-level relationships. Only a single molecular data set exists that includes all four tapir species (Norman & Ashley 2000) and each of the two partial mitochondrial genes used (*MT-CO2* and *MT-CYB*) yielded different topologies. Within Rhinocerotidae, the placement of the Sumatran rhino (*Dicerorhinus sumatrensis*) is contentious, with different lines of evidence leading to different conclusions. Its possession of two horns leads some to place it as the sister-taxon of the similarly two-horned African rhinos (Dicerotini; *Ceratotherium simum* and *Diceros bicornis*) (Simpson 1945; Loose 1975). Others suggest instead that it is more closely related to the Asian *Rhinoceros* clade based on geography (Groves 1983), a placement that is also supported by the most recent molecular study (Tougard et al. 2001). A third viewpoint places *Dicerorhinus* as a separate lineage that is not more closely related to either Dicerotini or *Rhinoceros* (Guerin 1982; Cerdeno 1995).

The eight living representatives of Hippomorpha all belong to the genus *Equus*. Relationships within *Equus* remain unclear, although the wild relative of the domestic horse (*Equus caballus*) is held by many to be sister to the remaining species (e.g. Harris & Porter 1980; Lowenstein & Ryder 1985; George & Ryder 1986). The zebras are often split into the two distantly related subgenera *Dolicohippus* (*E. grevyi*) and *Hippotrigris* (*E. burchelli, E. zebra* and *E. quagga*), although it has been hypothesized that the three *Hippotrigris* species each have separate origins within the caballine horses of North America and Eurasia (Bennett 1980). The recently extinct quagga (*E. quagga*) – the last known individual died in Amsterdam Zoo in 1882 – is included in this study due to continued interest in its taxonomic status (e.g. Thackery 1997; Klein & Cruz-Uribe 1999; Groves & Bell 2004; Leonard et al. 2005) and the existence of the controversial selective breeding program to recreate the quagga phenotype from the plains zebra (*E. burchelli*) (www.quaggaproject.org).

A complete phylogeny of the perissodactyls is highly desirable, not only for systematic interest, but also to provide a framework for exploring evolutionary patterns and processes as well as potentially being an important tool for conservation biology (see Purvis et al. 2005). Yet, despite both perissodactyls being a clade of large charismatic mammals and the tractability of sequencing genes from just 17 species, no single gene has been sequenced for all perissodactyls and even the most recent review of perissodactyl phylogeny (Norman & Ashley 2000) includes only 10 species. To address this gap, we used two "competing" frameworks of data combination (supermatrix and supertree analysis) to yield the first, complete phylogenies of the order based upon robust analytical methods. In so doing, we also explore the problems and benefits associated with each method and their effects upon the phylogenetic hypotheses generated.

Material and methods

To address the question of perissodactyl phylogeny, we constructed and analyzed two independent data sets. The first comprises a multi-gene data set representing the "current systematic database" (*sensu* Gatesy et al. 2002) for Perissodactyla, which was amenable to analysis by both the supertree and supermatrix approaches. The second data set derived from a literature search of all previously postulated hypotheses of perissodactyl phylogeny. Because the data underlying these hypotheses were often incompatible or not available, this data set could only be analyzed in a (traditional) supertree framework (*sensu* Bininda-Emonds 2004).

Data collection

Molecular data set. This data set was prepared to provide an identical set of characters that could be analyzed using both supermatrix and supertree approaches so as to allow for direct comparison between them. All perissodactyl accessions were downloaded from GenBank on April 6, 2006 and passed through the Perl script GenBank-Strip v2.0 to retain only those genes that had been sequenced for at least three species (according to the NCBI taxonomy) and were longer than 200 bp (except for tRNA genes, where the lower limit was 50 bp). A total of 39 genes (Tab. 1; App. 1) met these criteria, for which the homologous sequences from the artiodactyl *Bos taurus* were added for outgroup analysis. The sequences for each gene were aligned using ClustalW (Thompson et al. 1994) or with transAlign (Bininda-Emonds 2005) in combination with ClustalW for the protein-coding sequences, and improved manually where needed. The Perl script seqCleaner v1.0.2 was used subsequently to standardize the species names according to Grubb (1993), to eliminate poor-quality sequences (i.e. those with > 5 % Ns), to prune flanking regions possessed by only a minority of species, and to ensure that all sequences overlapped pairwise by at least 100 bp (or 25 bp for the tRNA genes). The final data set comprised 39 genes with an aligned length of 19,260 bp. All but six of the genes derive from the mitochondrial genome. Of the latter genes, 18 encode tRNAs with the remainder encoding protein-coding genes, the control region ("MT-control region"), and the two rRNAs, *MT-RNR1* and *MT-RNR2*.

Literature-based supertree analysis

Potential sources of phylogenetic information were identified from the literature by searching Web of Science and BioAbstracts using the term perissodactyl*; after collection, the bibliographies of all relevant articles were searched to find any additional papers. All methods of phylogenetic estimation were accepted initially (including informal techniques with no algorithm), although only a single taxonomy (that of Grubb 1993) was included. The taxonomy, which is complete at the species-level, overlaps with all source trees, thereby acting as a 'seed tree' to provide the backbone for the analysis, which has been shown to improve accuracy in simulation (Bininda-Emonds & Sanderson 2001). The highly unresolved nature of the taxonomy also ensures that it does not unduly influence the topology of the supertree (see Beck et al. 2006). Because the duplication of data sets can potentially bias the resultant supertree (Springer & de Jong 2001; Gatesy et al. 2002), no other taxonomy was included because there is no way to ascertain how much duplication of source information occurs between taxonomies. To further reduce data set duplication, the protocol of Bininda-Emonds et al. (2004) was followed. Briefly, the protocol provides guidelines to help identify 'independent' phylogenetic hypotheses (largely based on the data source) suitable for inclusion in a supertree analysis. Although the protocol has been criticized (see Gatesy et al. 2004), an empirical case study by Beck et al. (2006) has shown that its use can indeed correct for errors in a supertree analysis arising from data duplication.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Selected characteristics of the 39 genes contributing to the molecular data set. Gene symbols, unless noted otherwise or in quotation marks, follow Wain et al. (2002).

Gene	Number of taxa	Number of bps
B2M	6	276
DRD4	7	504
HBE4 (Swiss-Prot)	8	318
MC1R	7	664
"MHC CLASS I ANTIGEN"	7	851
MT-ATP6	6	681
MT-ATP8	6	207
"MT-control region"	14	633
MT-CO1	7	1545
MT-CO2	9	684
MT-CO3	6	784
MT-CYB	12	1140
MT-ND1	6	957
MT-ND2	6	1044
MT-ND3	6	346
MT-ND4	6	1378
MT-ND4L	6	297
MT-ND5	6	1821
MT-ND6	6	528
MT-RNR1	17	985
MT-RNR2	7	1616
MT-TA	6	69
MT-TC	6	68
MT-TD	6	69
MT-TE	6	69
MT-TF	6	72
MT-TH	6	71
MT-TI	6	70
MT-TK	6	70
MT-TM	6	69
MT-TN	6	73
MT-TP	7	67
MT-TQ	6	73
MT-TR	6	70
MT-TT	6	81
MT-TV	7	67
MT-TW	6	74
MT-TY	6	68
PRNP	5	801
Combined supermatrix	19	19 260

All source trees were stored by inputting them exactly as presented in the original papers into MacClade (Maddison & Maddison 2003) and saving them to a single nexus-formatted treefile (Maddison et al. 1997). Thereafter, the Perl script synonoTree v2.1 (Bininda-Emonds et al. 2004) was used to standardize all species names among the source trees to those present in Grubb (1993). Any taxon that could not be unambiguously assigned to a species in Grubb (1993) was pruned from the source tree by synonoTree, which also accounts for any species that are rendered as non-monophyletic as a result of the synonymization process. All non-perissodactyl mammals were synonymized and reduced to a single terminal taxon ("outgroup") to yield a rooted source tree; source trees containing only perissodactyl species were considered to be unrooted. All trees that were pruned so as to be uninformative (i.e. fewer than three or four taxa for rooted versus unrooted trees, respectively) were deleted.

Tree building

All trees and their underlying data matrices have been deposited in TreeBASE (www.treebase.org; Sanderson et al., 1994) under the study accession number S2227 and the matrix accession numbers M4235 (literature-based MRP analysis) and M4234 and M4236 (molecular supermatrix).

Supertrees

Supertrees for both the molecular and literature-based data sets were derived using Matrix Representation with Parsimony (MRP; Baum 1992; Ragan 1992), which is by far the most commonly used supertree method to date (see Bininda-Emonds 2004). Briefly, MRP converts each source tree into its matrix equivalent, whereby its informative nodes are represented as a series of partial binary 'pseudocharacters'; each column in the matrix corresponds to one node in the tree. Taxa descended from a given node are coded as '1', taxa not descended from that node are coded as '0' and any taxa not present on a given tree are represented by '?'. Ordinarily, all source trees are effectively rooted by adding a hypothetical all-zero outgroup to the final MRP matrix. However, we used semi-rooted MRP coding as implemented in the Perl script SuperMRP v1.2.1 (Bininda-Emonds et al. 2005), in which the outgroup receives 0s only for those MRP characters pertaining to rooted source trees. For unrooted source trees, the outgroup instead receives ?s for the respective characters.

For the molecular data set, the gene tree for each individual gene was determined under a ML framework using RAxML-VI-HPC v2.2 (Stamatakis 2006) to serve as input into the supertree analysis. A GTR + G model of evolution was used in all cases; otherwise, the default search parameters were used. Bootstrap support values (Felsenstein 1985) for each gene tree were also determined at the same time based on 1,000 replicates. The outgroup *Bos taurus* was used to root all gene trees and was pruned thereafter. Two supertrees were built from the molecular data set, one ignoring differential signal strength within the source trees and the other using the bootstrap frequencies to weight each node (as implemented in SuperMRP). The latter procedure that has been shown in simulation to enable MRP supertree construction, on average, to slightly outperform supermatrix analysis in terms of the accuracy of reconstructing a known model tree (Bininda-Emonds & Sanderson 2001).

Two MRP matrices were also constructed for the source trees derived from the literature. The first contained all independent source trees identified, whereas the second examined the issue of source-tree quality by using the more stringent source-tree selection criterion advocated by Gatesy et al. (2004) and used by Price et al. (2005). For the latter analysis, poor-quality source trees (e.g. trees derived without the use of explicit tree-building algorithms or using methods that are now generally discredited) were excluded from the matrix. The inclusion of poor-quality source trees in supertree analyses remains highly criticized (Gatesy et al. 2002, 2003, 2004; Gatesy & Springer 2004) despite strong empirical evidence that they generally have little effect upon the supertree topology (Purvis 1995; Jones et al. 2002; Price et al. 2005; see also Bininda-Emonds 2000). Even so, MRP pseudocharacters derived from the taxonomy seed tree were given a weight of 0.001 (i.e., 1,000 $\times$ less than other pseudocharacters) to ensure that any single source tree could easily overrule the minimal information in the seed tree in the case of conflict.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All MRP matrices were analyzed under a parsimony criterion using a branch-and-bound search algorithm (Hendy & Penny 1982) as implemented in PAUP* v4.0b10 (Swofford 2003), thereby ensuring that all optimal solutions would be found. In all cases, the final supertree was held to be the strict consensus of all equally most parsimonious trees.

Support within each supertree was quantified using the supertree-specific reduced qualitative support (rQS) index (Bininda-Emonds 2003; Price et al. 2005). Unlike conventional bootstrap or Bremer support methods, the rQS indices calculate the degree of support for each individual node on the supertree among the set of source trees (with the supertree pruned to the taxon set of the source tree it is being compared to), thereby avoiding the inherent non-independence among the 'pseudocharacters' derived from a single tree in the MRP matrix. Each source tree either supports, conflicts or is equivocal with respect to a given node in the supertree. The results are summed across the set of source trees and normalized to fall between −1 (all source trees conflict with the supertree node) and +1 (all source trees support the supertree node). rQS values tend toward zero, the value for equivocal or non-applicable source trees. As such, any positive values indicate greater support among the set of source trees for a node than conflict, and are held to indicate well supported nodes.

Supermatrix

We concatenated all gene sequences into a single matrix that was analyzed using unweighted maximum parsimony (MP), neighbour joining (NJ) and minimum evolution (ME) (all using PAUP*), maximum likelihood (ML) (using RAxML) and Bayesian inference (BI) (using MrBayes v3.1.2; Ronquist & Huelsenbeck 2003) methods. MP analyses again used a branch-and-bound algorithm, whereas the ME analyses used a heuristic search algorithm with TBR branch-swapping. The NJ and ME analyses were based on GTR distances, whereas both likelihood-based analyses assumed a GTR + G model of evolution for the data set, with the parameters being free to vary between the individual genes. Again, the default search parameters were otherwise used in RAxML. BI searches employed a MC3 algorithm of two runs, each consisting of four chains (one heated, three cold) that were run for 10,000,000 generations with the first 5,000,000 generations being discarded as burn-in. Trees were sampled every 5,000 generations (for 1,000 trees total) to derive the final tree and estimates of the posterior probabilities. Support for the relationships within each tree except for the BI tree were quantified using 1,000 bootstrap replicates using the respective program (i.e. PAUP* or RAxML).

Tree comparison

Tree topologies were compared using two different congruence measures that differ in how they treat polytomies, the symmetric-difference metric (d_s; Robinson & Foulds 1979, 1981) and the consensus-fork index (CFI; Colless 1981). The CFI quantifies the resolution of the consensus of the trees being compared by dividing the number of non-trivial clusters (i.e. those containing two or more taxa) by their maximum possible number (= number of terminal taxa, n, minus 2). Because any polytomies will decrease the resolution of a strict consensus tree, they will result in decreased values of the CFI (i.e. any polytomy is loosely considered to be "wrong"). We therefore used a semi-strict consensus tree to retain information about nodes that were congruent (i.e. did not explicitly conflict) among the trees being compared. The resulting value is therefore probably also more comparable to the information provided by d_s, which calculates the number of clades that appear on one tree or the other but not on both. We standardized d_s by dividing it through its maximum value for a rooted tree of $2n-4$, and subtracted this value from 1 to yield a similarity measure comparable to the CFI.

Results and discussion

Data coverage and distribution

Molecular data set. The data set contained 39 genes, all but six of which are found on the mitochondrial genome. The gene with the greatest phylogenetic coverage was *MT-RNR1*, for which usable sequences were available for all but three of the 18 perissodactyl species examined (*Equus onager*, *Equus quagga*, and *Tapirus terrestris*). Only *Equus caballus* and *E. asinus* had sequence data available for all 39 genes, whereas *Ceratotherium simum*, *Rhinoceros unicornis* and *Tapirus terrestris* had sequence data available for over half of the genes. The remaining species had information for at most seven gene sequences. All members of the rhino family had sequences available for MT-control region, *MT-CYB* and *MT-RNR1*, whereas *MT-CO2* has been sequenced for all members of Tapiridae. By contrast, no single gene is available for all members of Equidae, although the MT-control region is available for eight of the nine species.

Literature-based data set. The full supertree was built from 19 source trees (including the seed taxonomy) derived from 15 published articles, of which two of the source trees representing eight MRP characters (see Tab. 2) were held to be derived from poor-quality data. Three articles used morphological data only, 12 used molecular data and one used both molecular and morphological data, although the method of phylogenetic construction that was used in the latter was unclear. The largest source tree (excluding the seed taxonomy) contained 10 perissodactyl species, with the majority of trees containing between four and eight species.

Perissodactyl phylogeny

The different data sets and analytical methods resulted in phylogenies that were all at least 80 % resolved, with those based on the molecular data set always being at or close to 100 % (Figs 1–3). As such, accounting for signal strength (i.e. bootstrap support) in the molecular supertree analysis had only a minimal, albeit positive effect on resolution. The analogous procedure in the literature-based supertree analysis (i.e. deleting poor-quality source trees) likewise produced only a small increase in resolution (from 81.2 % to 87.5 %). All areas of poor resolution were restricted to within Equidae.

Except for the NJ and ME supermatrix analyses, the trees produced by the different methods are all reasonably congruent (Figs 1–3; Tab. 4), with areas of conflict localized within Equidae. The NJ and ME analyses, by contrast, both reconstruct perissodactyl phylogenies in which none of Rhinocerotidae, Tapiridae or Equidae are monophyletic. This result illustrates the possible limitations of these methods for phylogenetic reconstruction using large, gap-laden data sets. In the present case, 14 of the 171 pairwise comparisons were unde-

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Source-tree information for the literature-based MRP supertrees. Source trees that were removed from the reduced tree are listed in italics. Source trees that were held to be rooted are indicated in bold face.

Source	Data type	Tree-building method	Number of MRP pseudo-characters
Grubb (1993)	**Taxonomy seed tree**	**Taxonomy**	**4**
Norman & Ashley (2000: fig. 1)	**COII mtRNA**	**MP**	**6**
Norman & Ashley (2000: fig. 3)	**12S mtRNA**	**MP**	**8**
Norman & Ashley (2000: fig. 4)	**COII & 12S mtRNA**	**MP**	**6**
Oakenfull & Clegg (1998: fig. 5D)	θ globin; nDNA	ML	3
Oakenfull & Clegg (1998: fig. 5A–C)	α1 and α2 globin gene	Mini-supertree[a] of three trees	3
George & Ryder (1986: fig. 1e)	Morphological (reanalysis of Eisenmann 1979)	MP	4
Tougard et al. (2001: fig. 3)	**12S & cytochrome b mtDNA**	**ML**	**6**
Bennett (1980: fig. 1)	Morphological	MP	7
Lowenstein & Ryder (1985)	*Immunological distances*	*Unspecified distance tree (UPGMA?)*	3
Perez-Barberia & Gordon (1999: fig. 3)	***Mixed molecular and morphological***	***Unspecified***	**5**
Amato et al. (1993: fig. 1)	mtDNA	MP	2
Ishida et al. (1995: fig. 3)	mtDNA D-loop	ML	3
Ishida et al. (1995: fig. 4)	Immunological distances	NJ	1
Flint et al. (1990: fig. 3)	Restriction maps	MP	4
Pitra & Veits (2000: fig 3I–III)	**Cytochrome b mtDNA**	**Mini-supertree[a] of three trees**	**5**
Morales & Melnick (1994)	Restriction maps	Mini-supertree[a] of two trees	2
Harris & Porter (1980: fig. 8)	Morphological	Correlation phenogram	6
Arnason & Janke (2002: fig. 2–3)	**Complete mtDNA genome**	**Mini-supertree[a] of two trees**	**2**

[a] To ensure the data matrix is not swamped by duplicated data, mini-supertrees are built when more than one tree is presented within the paper for the same data but no preference for a single topology is stated by the authors (for further details, see Bininda-Emonds et al. 2004).

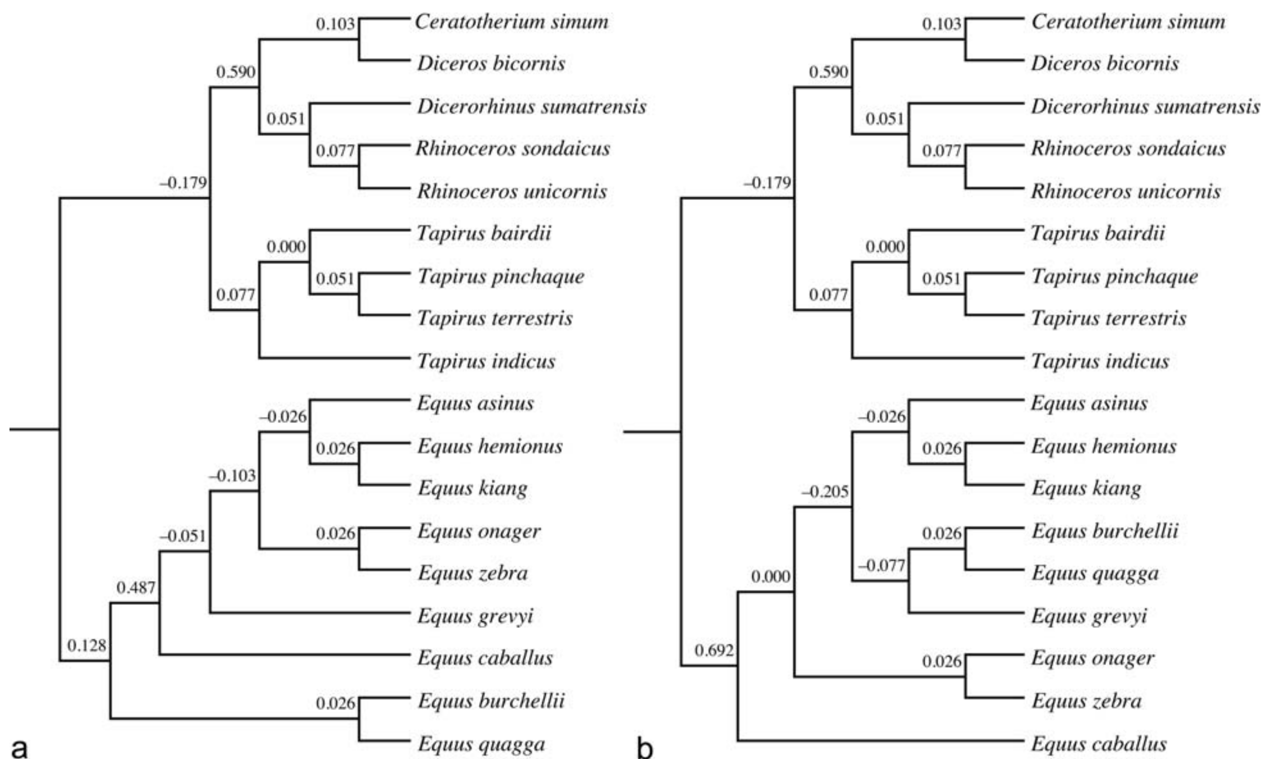

Figure 1. MRP supertrees derived from the analysis of the molecular data set, with source trees either **a.** weighted equally or **b.** weighted according to the bootstrap supports of their nodes. Tree statistics can be found in Table 3. Trees were rooted using the MRP outgroup (= *Bos taurus*), which was pruned subsequently. Numbers above branches represent rQS values.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

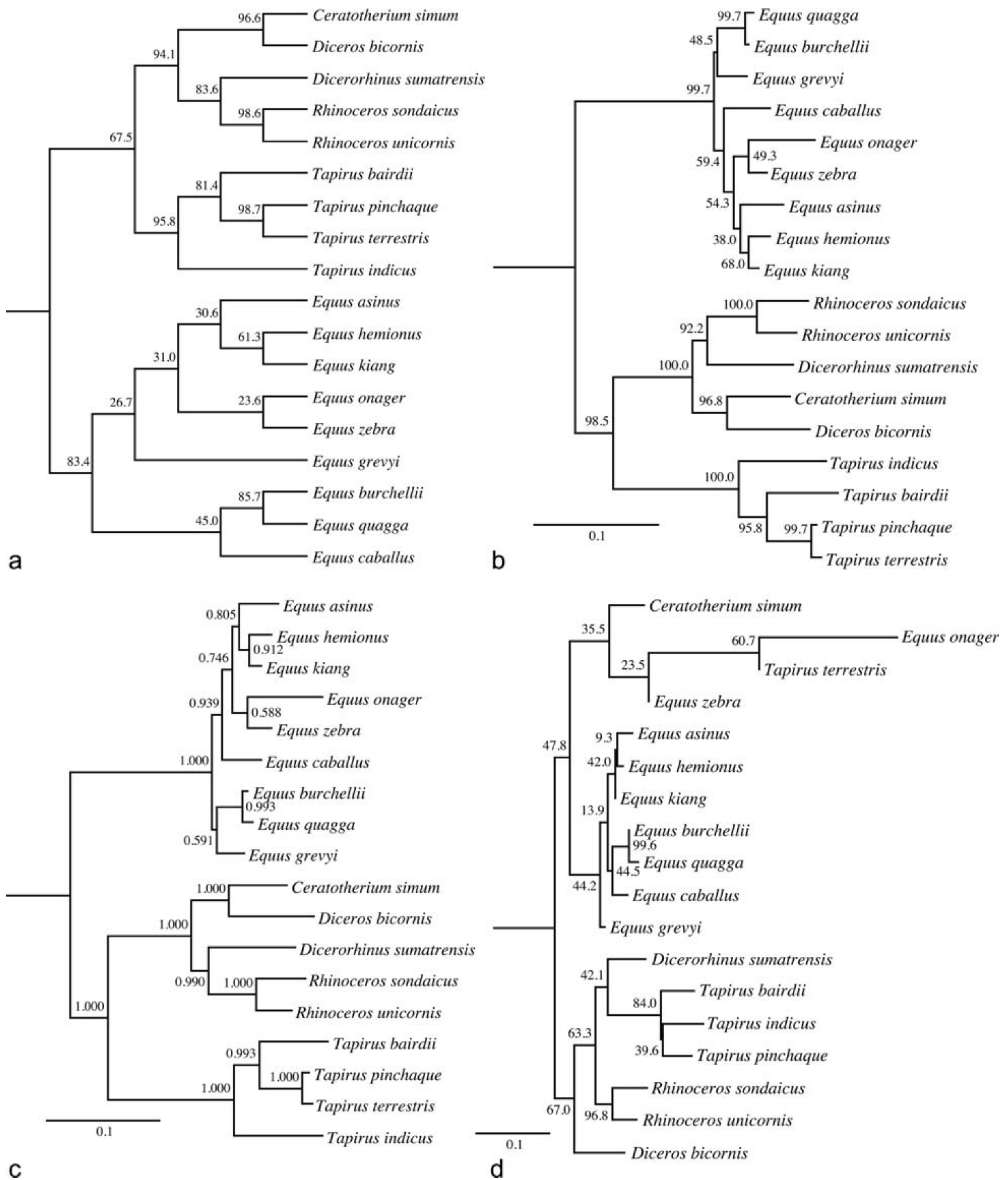

Figure 2. Supermatrix trees derived from the analysis of the molecular data set using **a.** MP, **b.** ML, **c.** BI, **d.** NJ or **e.** ME. Tree statistics can be found in Table 3. Trees were rooted using *Bos taurus*, which was pruned subsequently. Numbers above branches represent bootstrap frequencies (with 1,000 replicates) or posterior probabilities.

fined because of a lack of data for the two species; many other comparisons were undoubtedly based on highly limited amounts of data. Therefore, we will exclude the NJ and ME from further discussion of the tree topologies generated by this study. The ML and BI analyses produced identical topologies and henceforth will be referred to as the ML/BI topology.

All trees recovered each of Ceratomorpha, Tapiridae, Rhinocerotidae, and Equidae as monophyletic; these clades generally received strong support according to the appropriate measure. Moreover, all trees also reconstruct relationships within Tapiridae as (((*T. terrestris*, *T. pinchaque*), *T. bairdii*), *T. indicus*), although this result is admittedly based on a single data set (compris-

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

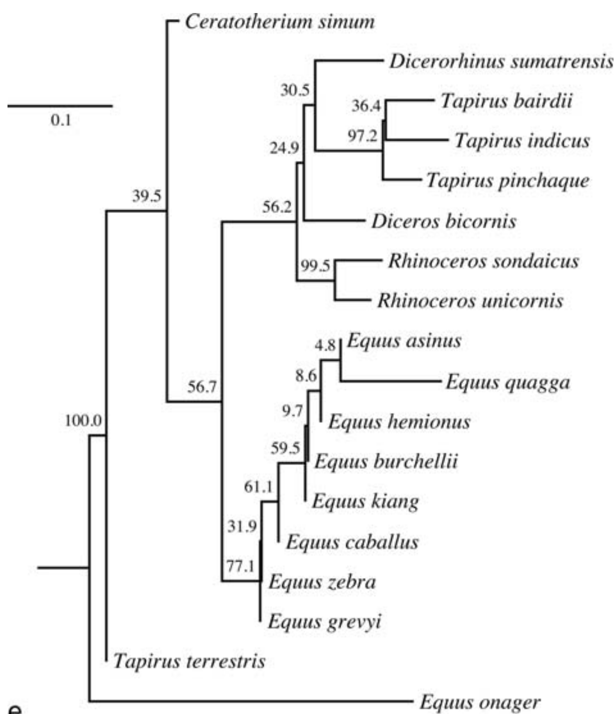

Figure 2. (Continued)

ing *MT-CO2* only) that was analyzed previously by Norman and Ashley (2000). The inferred relationships within Tapiridae are consistent with their current biogeography and what is understood of the evolutionary history of the group. The first split is between the Southeast Asian *T. indicus* from the remaining neotropical species, with the Central-South American *T. bairdii* then splitting from the two South American species (*T. pinchaque* and *T. terrestris*). The split between the two South American species is believed to be relatively recent (three million years ago), coinciding approximately with the time tapirs entered South America after the formation of the modern Panamanian land bridge (Ashley et al. 1996). Previous analyses of the tapir molecular data set (Ashley et al. 1996; Norman & Ashley 2000) also found a close relationship between the two South American species, although the position of *T. bairdii* was not resolved. It formed either part of a monophyletic Neotropical clade or a sister-taxon relationship with the Malayan tapir.

Broad consensus also exists with respect to relationships within Rhinocerotidae. The African *Ceratotherium* + *Diceros* (Dicerotini) and the two Asian *Rhinoceros* species are each consistently recovered as clades in agreement with current phylogenetic opinion (e.g. Tougard et al. 2001). The position of the historically problematic Sumatran rhino, *Dicerorhinus*, is also clarified with all analyses of the molecular data set unambiguously placing it in a clade with the Asian *Rhinoceros*, thereby supporting the biogeographic theory of rhino evolution (Tougard et al. 2001; Groves 1983). Three genes (MT-control region, *MT-CYB* and *MT-RNR1*) are available for all five species of rhino, the

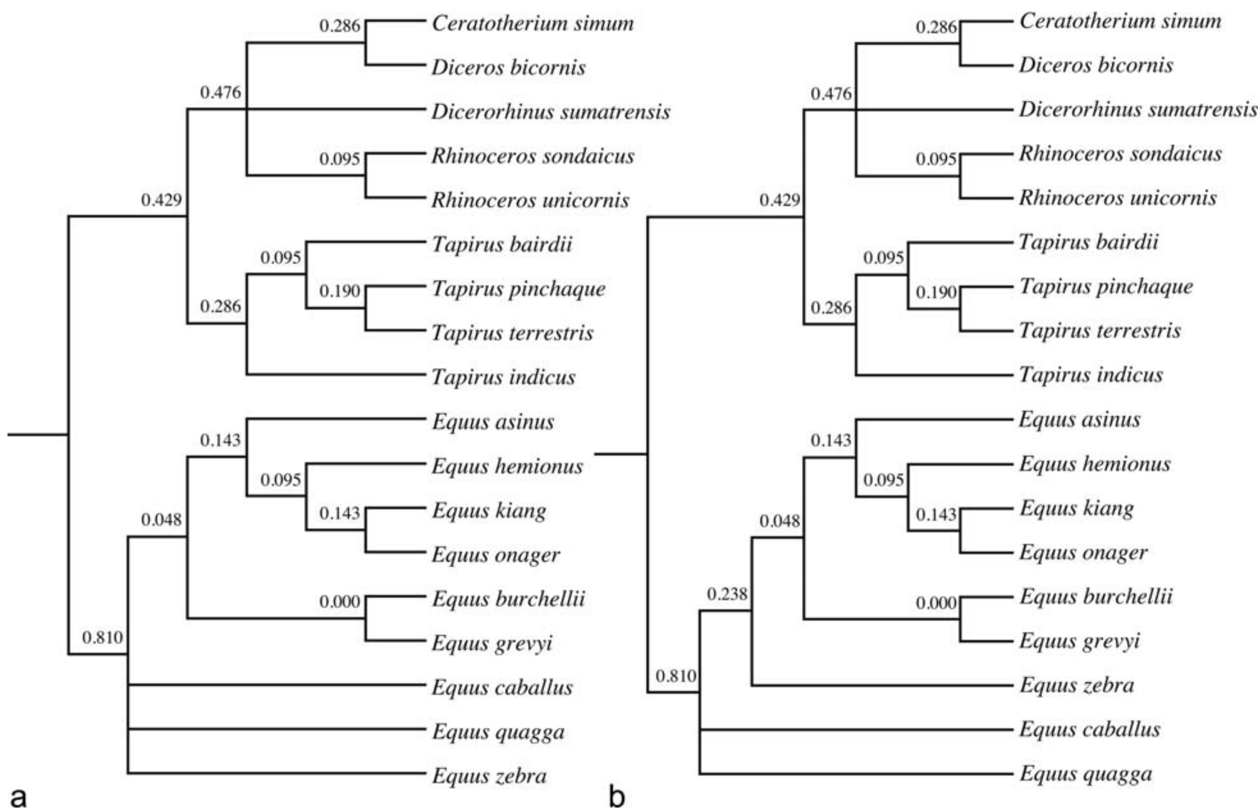

Figure 3. MRP supertrees derived from source trees obtained from the literature, with either **a.** all non-independent source trees being including or **b.** only good-quality, non-independent source trees. Tree statistics can be found in Table 3. Trees were rooted using the MRP outgroup, which was pruned subsequently. Numbers above branches represent rQS values.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 3. Selected statistics from the different combined data analyses for elucidating perissodactyl phylogeny.

Data set	Analysis method	Data quantity	Optimality score	Number of equally optimal solutions	Resolution
Molecular	Supertree (unweighted MRP)	153 MRP pseudo-characters	207 steps	2	93.8%
Molecular	Supertree (weighted MRP)	153 MRP pseudo-characters	1372.8 steps	1	100%
Molecular	Supermatrix (NJ)	19 260 bp	n/a	1	100%
Molecular	Supermatrix (ME)	19 260 bp	1.30127	1	100%
Molecular	Supermatrix (MP)	19 260 bp	9517 steps	1	100%
Molecular	Supermatrix (ML)	19 260 bp	−64851.863641	1	100%
Molecular	Supermatrix (BI)	19 260 bp	−64908.76 (run 1); −64939.25 (run 2)	n/a	n/a
Literature	All source trees	80 MRP pseudo-characters	94.004 steps	8	81.2%
Literature	Good-quality source trees	72 MRP pseudo-characters	83.004 steps	6	87.5%

latter two are from Tougard et al. (2001) who resolve the same relationship. The literature-based supertree analysis, however, places *Dicerorhinus* in a soft polytomy with respect to the remaining rhinoceros species. As such, this placement should not be taken as evidence for the separate-lineage hypothesis (Guerin 1982; Cerdeno 1995), which requires additional, independent evidence to support a multiple, simultaneous speciation event (hard polytomy). No support was found for the horn-number hypothesis (Simpson 1945; Loose 1975), which clusters *Dicerorhinus* with Dicerotini.

The major disagreement among the various methods and data sources concerns relationships within Equidae, in keeping with the historical uncertainty regarding the systematics of this group. Equidae is generally resolved as being monophyletic (with the exception of the molecular supermatrix analyzed using either NJ or ME; Figs 1d, e, respectively). Significant topological incongruence among the 39 gene trees in the supermatrix is indicated. For example, of the seven trees based on the molecular data set that include both *E. caballus* and *E. grevyi*, two trees place them as sister-taxa, three place *E. caballus* as basal to the rest of equids, one places *E. grevyi* as basal to all equids and one places *E. grevyi* as sister-taxon to *E. asinus* and *E. caballus* which is sister-taxon to *E. burchelli*. The difficulties in reconstructing relationships within *Equus* are also underscored by the low support values returned by all methods in this region of the tree. A possible explanation might be a rapid adaptive radiation within *Equus*, as hinted at by the short branch lengths recovered in the ML and BI analyses.

Nevertheless, *E. quagga* and *E. burchellii* are reliably recovered as sister-taxa by the different analyses of the supermatrix. This result, because it is at the species-level, is consistent with both the hypothesis that the quagga is a subspecies of the plains zebra (Groves & Bell 2004; Leonard et al. 2005) and with it being a separate species (Thackery 1997). Our results do exclude the possibility that the quagga is sister-taxon to the mountain zebra (*E. zebra*; Klein & Cruz-Uribe 1999). However, the placement of the quagga is essentially based on a single gene (MT-control region; Leonard et al. 2005), which is available for eight of the nine *Equus* species. A second gene for which sequence data for the quagga exist (111 bp of *MT-CO1*) does not contradict this placement; however, corresponding *MT-CO1* sequences were only available for five other perissodactyl species, with neither of the two equid sequence originating from zebras. Even so, the position of the *E. burchellii* and *E. quagga* clade is unstable across the supermatrix analyses. It is placed alternatively at the base of the equids (unweighted supertree), in a clade with *E. caballus* (MP) or *E. grevyi* (ML/BI) at the base of the equids, or as a clade with *E. grevyi* that forms a sister-taxon relationship with a clade comprising *E. asinus*, *E. hemionus* and *E. kiang* (weighted supertree).

The literature-based supertree analyses are even less informative about the placement of the quagga, with it forming a basal polytomy with *E. caballus* and also *E. zebra* in the unweighted analysis. Examination of the equally most parsimonious solutions reveals that the quagga is largely responsible for the polytomy, in part due to the limited and conflicting information available in the literature concerning its phylogenetic position (Bennett 1980; Harris & Porter 1980; Lowenstein & Ryder 1985).

Otherwise, our results show little support for the few traditional groupings that exist within equids. No tree supports the hypothesis that asses and zebras are monophyletic (e.g. Bennett 1980; Kaminski 1979), although support for this hypothesis within the literature is admittedly mixed (e.g. compare with Eisenmann 1979; Harris & Porter 1980; Flint et al. 1990). Zebras are never resolved as monophyletic, although all but *E. zebra* often tend to form a clade. The latter grouping, however, contradicts the monophyly of the zebra subgenus *Hippotrigris*, possibly lending support to the hypothesis of Bennett (1980) of multiple origins for the subgenus. Similarly, the asses form a clade only in the literature-based supertrees, although three of the four species (*E. asinus*, *E. hemionus*, and *E. kiang*) consistently cluster together. Moreover, except for the literature-based supertrees, there is little or no support for a

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 4. Topological congruence between pairs of the different combined data analyses as measured by either the CFI of the semi-strict consensus tree (above diagonal) or the inverted partition metric, $1-d_s$ (below diagonal).

Data set		Molecular supermatrix						Literature-based source trees		
	Analysis method	NJ	ME	MP	ML	BI	Unweighted MRP	Weighted MRP	Unweighted MRP	Weighted MRP
Molecular supermatrix	NJ	–	0.333	0.267	0.200	0.200	0.200	0.200	0.067	0.067
	ME	0.312	–	0.067	0.067	0.067	0.067	0.067	0.067	0.067
	MP	0.250	0.062	–	0.875	0.875	0.938	0.812	0.562	0.562
	ML	0.188	0.062	0.875	–	1.000	0.875	0.875	0.562	0.562
	BI	0.188	0.062	0.875	1.000	–	0.875	0.875	0.562	0.562
	Unweighted MRP	0.188	0.062	0.938	0.875	0.875	–	0.812	0.562	0.562
	Weighted MRP	0.188	0.062	0.812	0.875	0.875	0.812	–	0.625	0.625
Literature-based source trees	Unweighted MRP	0.156	0.156	0.594	0.594	0.594	0.594	0.594	–	0.875
	Weighted MRP	0.125	0.125	0.562	0.562	0.562	0.562	0.562	0.969	–

close relationship between *E. hemionus* and either *E. kiang* or *E. onager*, despite the latter two species historically being considered subspecies of *E. hemionus* (Schlawe 1986). In fact, *E. onager* is never placed as the sister species of *E. hemionus*, thereby contradicting its recent subordination as a subspecies of the latter (Grubb 2005).

Comparing approaches

Supertree versus supermatrix. The relative strengths of the supertree and supermatrix approaches have been vigorously debated (e.g. Rodrigo 1993; Springer & de Jong 2001; Gatesy et al. 2002, 2003, 2004; Bininda-Emonds et al. 2002, 2003, 2004). However, few direct comparisons between the two have been made in an empirical framework (but see Gatesy et al. 2004; Fulton & Strobeck 2006; Higdon et al. 2007). In the current study, the results reached by the different approaches are highly congruent, with differences tending to reflect areas of historical and ongoing uncertainty about relationships within Equidae, rather than explicit shortcomings in any given method. This conclusion supports the findings of both Fulton & Strobeck (2006) and Higdon et al. (2007) who found that supertree and supermatrix approaches performed equally well when analyzing molecular data sets of Arctoidea (Carnivora) and that the only discrepancies occurred in areas of the tree that were not well resolved or supported by either method.

Our analyses of the molecular data set that derive monophyletic perissodactyl families (i.e. excluding the NJ and ME trees) are all highly congruent. The supertree topologies for this data set are highly congruent with the supermatrix trees obtained using ML, MP and BI (unweighted supertree, CFI or $d_s = 0.875–0.938$; bootstrap weighted supertree, CFI or $d_s = 0.812–0.875$; Tab. 4); these results compare well to the congruence among the supermatrix analyses themselves (ML, MP and BI, CFI or $d_s = 0.875–1.000$; Tab. 4).

Much of the congruence undoubtedly derives from the overwhelming preponderance of mtDNA genes in the molecular data set, all of which are linked on a single, clonally-inherited molecule. This overweighting of mtDNA information, whether at the level of genes (33 of 39; 84.6 %) or nucleotides (15,846 of 19,260; 82.3 %), is inherent to the available data for perissodactyls and affects both the supertree and supermatrix analyses. However, the different approaches accommodated for it (as well as the non-independence between the mtDNA genes) in subtly different, but important ways by analyzing either the gene trees or nucleotides directly. Moreover, the supermatrix analyses themselves also differed in how well they accounted for the different models of evolution for each gene: not at all for MP; the use of a single model, but one whose parameters could vary between genes for ML; and fully gene-specific models for BI. As such, like the recent empirical multi-gene study of pinnipeds (Higdon et al. 2007) that also was disproportionately dominated by mtDNA information, the high degree of congruence observed, especially between the supertree and supermatrix results, must go beyond simple mtDNA swamping of the molecular data set.

These results reinforce the idea that supertree and supermatrix methods in combination provide us with a way of looking for global congruence (*sensu* Lapointe et al. 1999; and as advocated by Bininda-Emonds 2004) where we have increased confidence in the relationships that are agreed upon by the different approaches.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Accordingly, we can here place a higher degree of confidence on the higher-level relationships and on the relationships within the Rhinocerotidae and Tapiridae that were obtained because all methods reconstruct the same topology from the molecular data set. A caveat, however, is that the relationships are based only on the few genes that have been sequenced for these two families (one gene for all tapirs and three for all rhinos) and may change when new data are added.

Neither the supertree nor supermatrix approaches currently manage to resolve the contentious equid relationships satisfactorily. The supermatrix analyses resolve different topologies depending on the method of phylogenetic reconstruction, and none are strongly supported. The NJ and ME analyses both reconstruct perissodactyl phylogenies that do not resolve the three extant families as monophyletic, indicating the limitations of these methods. The supertree approach also generates different topologies depending on whether bootstrap support is used to weight the gene trees obtained from the molecular data set. Both the weighted and unweighted supertrees are 100 % resolved, but the accuracy of the relationships amongst equids is similarly questionable due to low branch support values. They are, however, largely congruent with those of the supermatrix analyses, which admittedly are also poorly supported. The literature-based supertree analysis has the advantage of including more of the global phylogenetic database, but many of the entries in this database are highly incomplete. The resulting poor taxonomic overlap has been demonstrated in simulation to decrease the accuracy of both supertree and supermatrix analyses (Bininda-Emonds & Sanderson 2001), with the anomalous positioning of quagga in this study providing a cogent example of some of the artifacts that can arise under such circumstances.

Conclusions

Our analyses represent the most comprehensive attempt, both in terms of data and taxonomic coverage, to reconstruct the phylogeny of Perissodactyla to date. The combination of all available data sets, either by combining raw data (supermatrix approach) or tree topologies (supertree approach) provides clear support for the placement of the Sumatran rhino as sister-taxon to its Asian compatriots (*Rhinoceros*) and also for a monophyletic Neotropical tapir clade. Conflict between the supertree and supermatrix topologies only occurs within Equidae where the relationships are otherwise weakly supported; the MP and ML/BI supermatrix analyses also build conflicting topologies for this clade. Traditional groupings, such as a monophyletic zebra or ass clade or the monophyly of both groups together, are rarely indicated, although a majority of each of the zebra and ass species often do form clades. The problems posed by the molecular data set, namely the low degree of taxon overlap resulting from the patchy distribution of sequence data, leads to poor support and/or resolution within controversial clades for both the supermatrix and supertree analyses (matching the general findings of Bininda-Emonds & Sanderson 2001). To resolve the outstanding issues surrounding the evolutionary history of the equids, more raw data need to be collected.

Acknowledgements

We thank John Gittleman for his support and initial encouragement in investigating perissodactyl phylogeny. Funding was provided through a Heisenberg Scholarship of the German Science Foundation (DFG; BI 825/2-1) (OBE) and a NESCent (NSF #EF-0423641) postdoctoral fellowship (SAP). All Perl scripts were written by OBE and are freely available at http://www.uni-oldenburg.de/molekularesystematik/.

References

Amato, G. D., Ashley, M. & Gatesy, J. 1993. Molecular evolution in living species of rhinoceros: implications for conservation. *In* Ryder, O. A. (ed.). Rhinoceros biology and conservation: proceedings of an international conference. Zoological Society of San Diego: pp. 114–135.

Arnason, U. & Janke, A. 2002. Mitogenomic analyses of eutherian relationships. – Cytogenetic and Genome Research 96: 2–32.

Ashley, M., Norman, J. E. & Stross, L. 1996. Phylogenetic analysis of the perissodactylan family Tapiridae using mitochondrial cytochrome c oxidase (COII) sequences. – Journal of Mammalian Evolution 3: 315–326.

Baum, B. R. 1992. Combing trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. – Taxon 41: 3–10.

Beck, R. M. D., Bininda-Emonds, O. R. P., Cardillo, M., Liu, F.-G. R. & Purvis, A. 2006. Supertree analysis and the higher-level phylogeny of placental mammals. – BMC Evolutionary Biology 6(93).

Bennett, D. K. 1980. Stripes do not a zebra make, Part I, A cladistic analysis of *Equus*. – Systematic Zoology 29: 272–287.

Bininda-Emonds, O. R. P. 2000. Factors influencing phylogenetic inference, a case study using the mammalian carnivores. – Molecular Phylogenetics and Evolution 16: 113–126.

Bininda-Emonds, O. R. P. 2003. Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. – Systematic Biology 52: 839–848.

Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. – Trends in Ecology and Evolution 19: 315–322.

Bininda-Emonds, O. R. P. 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. – BMC Bioinformatics 6(156).

Bininda-Emonds, O. R. P., Beck, R. M. D. & Purvis, A. 2005. Getting to the roots of matrix representation. – Systematic Biology 54: 668–672.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L. & Purvis, A. 2007. The delayed rise of present-day mammals. – Nature 446: 507–512.

Bininda-Emonds, O. R. P., Gittleman, J. L. & Purvis, A. 1999. Building large trees by combining phylogenetic information, a complete phylogeny of the extant Carnivora (Mammalia). – Biological Reviews 74: 143–175.

Bininda-Emonds, O. R. P., Gittleman, J. L. & Steel, M. 2002. The (super)tree of life. – Annual Review of Ecology and Systematics 33: 265–289.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Bininda-Emonds, O. R. P., Jones, K. E., Price, S. A., Cardillo, M., Grenyer, R. & Purvis, A. 2004. Garbage in, garbage out: data issues in supertree construction. *In* Bininda-Emonds, O. R. P. (ed.). Phylogenetic supertrees: combining information to reveal the Tree of Life. Kluwer Academic, Dordrecht: pp. 267–280.

Bininda-Emonds, O. R. P., Jones, K. E., Price, S. A., Grenyer, R., Cardillo, M., Habib, M., Purvis, A. & Gittleman, J. L. 2003. Supertrees are a necessary not-so-evil: a comment on Gatesy et al. – Systematic Biology 52: 724–729.

Bininda-Emonds, O. R. P. & Sanderson, M. J. 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. – Systematic Biology 50: 565–579.

Cerdeno, E. 1995. Cladistic analysis of the family Rhinocerotidae (Perissodactyla). – American Museum Novitates 143: 1–25.

Colless D. 1981. Predictivity and stability in classifications: some comments on recent studies. – Systematic Biology 30: 325–331.

Eisenmann, V. 1979. Caracteres evolutifs et phylogenie du genre *Equus* (Mammalia, Perissodactyla). – Comptes Rendus des Seances de L'Academie des Sciences, Serie D – Sciences Naturelles 288: 497–500.

Eizirik, E., Murphy, W. J. & O'Brien S. J. 2001. Molecular dating and biogeography of the early placental mammal radiation. – The Journal of Heredity 92: 212–219.

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

Flint, J., Ryder, O. A. & Clegg, J. B. 1990. Comparison of the alpha-globin gene cluster structure in Perissodactyla. – Journal of Molecular Evolution 30: 36–42.

Fragoso, J. M. & Huffman, J. M. 2000. Seed-dispersal and seedling recruitment patterns by the last Neotropical megafaunal element in Amazonia, the tapir. – Journal of Tropical Ecology 16: 369–385.

Fulton, T. L. & Strobeck, C. 2006. Molecular phylogeny of the Arctoidea (Carnivora): effect of missing data on supertree and supermatrix analyses of multiple gene data sets. – Molecular Phylogenetics and Evolution 41: 165–181.

Gatesy, J., Amato, G., Norell, M. A., DeSalle, R. & Hayashi, C. 2003. Combined support for wholesale taxic atavism in gavialine crocodilians. – Systematic Biology 52: 403–422.

Gatesy, J., Baker, R. H. & Hayashi, C. 2004. Inconsistencies in arguments for the supertree approach, supermatrices versus supertrees of Crocodylia. – Systematic Biology 53: 324–355.

Gatesy, J., Matthee, C. A., Desalle, R. & Hayashi, C. 2002. Resolution of a supertree/supermatrix paradox. – Systematic Biology 51: 652–664.

Gatesy, J. & Springer M. S. 2004. A critique of matrix representation with parsimony supertrees. *In* Bininda-Emonds, O. R. P. (ed.). Phylogenetic supertrees: combining information to reveal the Tree of Life. Kluwer Academic Publishers, Dordrecht: pp. 369–388.

George, M. & Ryder O. A. 1986. Mitochondrial DNA evolution in the genus *Equus*. – Molecular Biology and Evolution 3: 535–546.

Groves, C. P. 1983. Phylogeny of the living species of rhinoceros. – Zeitschrift für Zoologische Systematik und Evolutionsforschung 21: 293–313.

Groves, C. P. & Bell, C. H. 2004. New investigations on the taxonomy of the zebras genus *Equus*, subgenus *Hippotigris*. – Säugetierkunde 69: 182–196.

Grubb, P. 1993. Perissodactyla. *In* Wilson, D. E. & Reeder, D. M. (eds.). Mammal Species of the World. 2nd ed. Smithsonian Institution Press, Washington, D.C.: pp. 369–372

Grubb, P. 2005. Order Perissodactyla. *In* Wilson, D. E. & Reeder, D. M. (eds.). Mammal Species of the World. 3rd ed. The Johns Hopkins University Institution Press, Baltimore: pp. 629–636

Guerin, C. 1982. Les Rhinicerotidae (Mammalia, Perissodactyla) du Miocene terminal au Pleistocene superieur d'Europe occidentale compares aux especes actuelles, Tendances evolutives et relations phyletiques. – Geobios 15: 599–605.

Harris, A. H. & Porter, L. S. 1980. Late Pleistocene horses of Dry Cave, Eddy County, New Mexico. Journal of Mammalogy 61: 46–65.

Hendy, M. D. & Penny, D. 1982. Branch and bound algorithms to determine minimal evolutionary trees. – Mathematical Biosciences 59: 277–290.

Higdon, J. W., Bininda-Emonds, O. R. P., Beck, R. M. D. & Ferguson, S. H. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. – BMC Evolutionary Biology 7(216).

Huxley, T. H. 1869. An introduction to the classification of animals. John Churchill and Sons, London.

Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H. & Saitou, N. 1995. Mitochondrial DNA sequences of various species of the genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. – Journal of Molecular Evolution 41: 180–188.

Jones, K. E., Purvis, A., MacLarnon, A., Bininda-Emonds, O. R. P. & Simmons, N. B. 2002. A phylogenetic supertree of the bats (Mammalia, Chiroptera). – Biological Reviews 77: 223–259

Kaminski, M. 1979. The biochemical evolution of the horse. – Comparative Biochemistry and Physiology 63 B: 175–178.

Klein, R. G. & Cruz-Uribe, K. 1999. Craniometry of the genus *Equus* and the taxonomic affinities of the extinct South African quagga. – South African Journal of Science 95: 81–86.

Lapointe F.-J., Kirsch J. A. W. & Hutcheon J. M. 1999. Total evidence, consensus, and bat phylogeny: a distance based approach. – Molecular Phylogenetics and Evolution 11: 55–66.

Leonard, J. A., Rohland, N., Glaberman, S., Fleischer, R. C., Caccone, A. & Hofreiter, M. 2005. A rapid loss of stripes, the evolutionary history of the extinct quagga. – Biology Letters 1: 291–295.

Loose, H. 1975. Pleistocene Rhinocerotidae of W Europe with reference to the recent two-horned species of Africa and SE Asia. – Scripta Geologica 33: 1–59.

Lowenstein J. M. & Ryder O. A. 1985. Immunological systematics of the extinct quagga (Equidae). – Experientia 41: 1192–1193.

Maddison, D. R. & Maddison, W. P. 2003. MacClade (version 4.0.6). Sinauer Associates, Sunderland, Massachusetts.

Maddison, D. R., Swofford, D. L., & Maddison, W. P. 1997. NEXUS: an extensible file format for systematics information. Systematic Biology 46: 590–621.

Madsen, O., Scally, M., Douady, C. J., Kao, D. J., DeBry, R. W., Adkins, R., Amrine, H. M., Stanhope, M. J., de Jong, W. W. & Springer, M. S. 2001. Parallel adaptive radiations in two major clades of placental mammals. – Nature 409: 610–614.

Merenlender, A. M., Woodruff, D. S., Ryder, O. A., Kock, R. & Vahala, J. 1989. Allozyme variation and differentiation in African and Indian rhinoceroses. – Journal of Heredity 80: 377–382.

Miyamoto, M. M. & Goodman, M. 1986. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. – Systematic Zoology 35: 230–240.

Morales, J. C. & Melnick. D. J. 1994. Molecular systematics of the living rhinoceros. – Molecular Phylogenetics and Evolution 3: 128–134.

Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A. & O'Brien, S. J. 2001. Molecular phylogenetics and the origins of placental mammals. – Nature 409: 614–618.

Norman, J. E. & Ashley, M. V. 2000. Phylogenetics of Perissodactyla and tests of the molecular clock. – Journal of Molecular Evolution 50: 11–21.

Oakenfull E. A. & Clegg J. B. 1998. Phylogenetic relationships within the genus *Equus* and the evolution of alpha and theta globin genes. – Journal of Molecular Evolution 47: 772–783.

Owen, R. 1848. Description of teeth and portions of jaws of two extinct anthracotheroid quadrapeds (*Hyopotamus vectianus* and *Hyop. bovinus*) discovered by the Marchioness of Hastings in the Eocene deposits on the NW coast of the Isle of Wight, with an

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

attempt to develop Cuvier's idea of the classification of pachyderms by the number of their toes. – The Quarterly Journal of the Geological Society of London 4: 104–141.

Perez-Barberia, F. J. & Gordon, I. J. 1999. The relative roles of phylogeny, body size and feeding style on the activity time of temperate ruminants: a reanalysis. – Oecologia 120: 193–197.

Pitra, C. & Veits, J. 2000. Use of mitochondrial DNA sequences to test the Ceratomorpha (Perissodactyla: Mammalia) hypothesis. – Zeitschrift für Zoologische Systematik und Evolutionsforschung 38: 65–72.

Price, S. A., Bininda-Emonds, O. R. P. & Gittleman, J. L. 2005. A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). – Biological Reviews 80: 445–473.

Prothero, D. R. & Schoch, R. M. 1989a. Origin and evolution of the Perissodactyla: summary and synthesis. *In* Prothero, D. R. & Schoch, R. M. (eds). The evolution of perissodactyls, Oxford University Press, Oxford: pp. 504–529.

Prothero, D. R. & Schoch, R. M. 1989b. Classification of the Perissodactyla. *In* Prothero, D. R. & Schoch, R. M. (eds). The evolution of perissodactyls, Oxford University Press, Oxford: pp. 530–537.

Prothero, D. R. & Schoch, R. M. 2002. Hoofs, tusks & flippers: evolution of hoofed mammals. Johns Hopkins University Press, Baltimore.

Purvis, A. 1995. A composite estimate of primate phylogeny. – Proceedings of the Royal Society of London B 248: 405–421.

Purvis. A., Gittleman J. L. & Brooks T. 2005. Phylogeny and Conservation. Cambridge University Press, Cambridge.

Radinsky, L. B. 1969. The early evolution of the Perissodactyla. – Evolution 23: 308–328.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. – Molecular Phylogenetics Evolution 1: 53–58.

Robinson, D. F. & Foulds, L. R. 1979. Comparison of weighted labeled trees. – Lecture Notes in Mathematics 748: 119–126.

Robinson, D. F. & Foulds L. R. 1981. Comparison of phylogenetic trees. – Mathematical Biosciences 53: 131–147.

Rodrigo, A. G. 1993. A comment on Baum's method for combining phylogenetic trees. – Taxon 43: 631–666.

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

Sanderson, M. J., Donoghue, M. J., Piel, W. & Eriksson, T. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. – American Journal of Botany 81: 183.

Schlawe, L. 1986. Seltene pfleglinge aus Dschunarei und Mongolei, kulane, *Equus hemionus hemionus* Pallas, 1775 – Zoologische Garten, Neue Folge 56: 299–323.

Simpson, G. G. 1945. The principles of classification and a classification of mammals. – Bulletin of the American Museum of Natural History 85: 1–350.

Springer, M. S. & de Jong, W. W. 2001. Which mammalian supertree to bark up? – Science 291: 1709–1711.

Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. – Proceedings of the National Academy of Science of the United States of America 100: 1056–1061.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. – Bioinformatics 22: 2688–2690.

Swofford, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods) (version 4) Sinauer Associates, Sunderland, Massachusetts.

Thackery, J. F. 1997. Morphometric, palaeoecological and taxonomic considerations of southern African zebras: attempts to distinguish the quagga. – South African Journal of Science 93: 89–93.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. – Nucleic Acids Research 22: 4673–4680.

Tougard, C., Delefosse, T., Hanni, C. & Montgelard, C. 2001. Phylogenetic relationships of the five extant rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial cytochrome b and 12s rRNA genes. – Molecular Phylogenetics and Evolution 13: 34–44.

Vilà, C., Leonard, J. A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R. K. & Ellegren, H. 2001. Widespread origins of domestic horse lineages. – Science 291: 474–477.

Wain, H. M., Lush, M., Ducluzeau, F. & Povey, S. 2002. Genew, the human gene nomenclature database. – Nucleic Acids Research 30: 169–171.

Wood, H. E. 1937. Perissodactyl suborders. – Journal of Mammalogy 18: 106.

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Appendix

Appendix. GenBank accessions for each of the 39 genes used to create the molecular data set. Gene symbols, unless noted otherwise or in quotation marks, follow Wain et al. (2002).

Gene	*Bos taurus*	*Ceratotherium simum*	*Dicerorhinus sumatrensis*	*Diceros bicornis*	*Equus asinus*	*Equus burchellii*	*Equus caballus*
B2M	NM_173893				AY124685	AY124688	AY124664
DRD4	AB069666				AB080629	AB080633	DQ277648
HBE4 *(Swiss-Prot)*	X03249	AF139502			AF140616		AF139506
MC1R	AF445642				AF141364		AF288357
"MHC CLASS I ANTIGEN"	BC109586		AJ133680		AJ133664		DQ083408
MTATP6	NC_006853	Y07726			X97337		NC_001640
MTATP8	NC_006853	Y07726			X97337		NC_001640
"MT-control region"	AF492440	NC_001808	AY742826	AY742831	NC_001788	AF220922	X79547
MTCO1	NC_006853	Y07726			X97337		NC_001640
MTCO2	NC_006853	Y07726			X97337		NC_001640
MTCO3	NC_006853	Y07726			NC_001788		NC_001640
MTCYB	NC_006853	Y07726	AJ245723	X56283	X97337	AY534349	DQ223533
MTND1	NC_006853	Y07726			X97337		NC_001640
MTND2	NC_006853	Y07726			X97337		NC_001640
MTND3	NC_006853	Y07726			NC_001788		NC_001640
MTND4	NC_006853	Y07726			NC_001788		NC_001640
MTND4L	NC_006853	Y07726			X97337		NC_001640
MTND5	NC_006853	Y07726			X97337		NC_001640
MTND6	NC_006853	Y07726			X97337		NC_001640
MTRNR1		X86942	AJ245722	AJ245721	X97337	AF221581	X79547
MTRNR2		Y07726			X97337		X79547
MT-TA	NC_006853	Y07726			X97337		NC_001640
MT-TC	NC_006853	Y07726			X97337		NC_001640
MT-TD	NC_006853	Y07726			X97337		NC_001640
MT-TE	NC_006853	Y07726			X97337		NC_001640
MT-TF	NC_006853	Y07726			X97337		NC_001640
MT-TH	NC_006853	Y07726			X97337		AY584828
MT-TI	NC_006853	Y07726			X97337		NC_001640
MT-TK	NC_006853	Y07726			X97337		NC_001640
MT-TM	NC_006853	Y07726			X97337		NC_001640
MT-TN	NC_006853	Y07726			X97337		NC_001640
MT-TP	NC_006853	Y07726		L22010	X97337		AF014411
MT-TQ	NC_006853	Y07726			X97337		NC_001640
MT-TR	NC_006853	Y07726			X97337		NC_001640
MT-TT	NC_006853	Y07726			X97337		NC_001640
MT-TV	NC_006853	Y07726			X97337		NC_001640
MT-TW	NC_006853	Y07726			X97337		NC_001640
MT-TY	NC_006853	Y07726			X97337		NC_001640
PRNP	NM_181015			AY133052	AY968590	AF117329	AY133051

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Appendix. continued

Gene	*Equus grevyi*	*Equus hemionus*	*Equus kiang*	*Equus onager*	*Equus quagga*	*Equus zebra*	*Rhinoceros sondaicus*
B2M	AY124697					AY124700	
DRD4	AB080634	AB080631				AB080635	
HBE4 *(Swiss-Prot)*	AF139504			AF139505		AF140615	
MC1R	AF141363	AF141365	AF141366			AF097749	
"MHC CLASS I ANTIGEN"	AJ133676	AJ133671					
MTATP6							
MTATP8							
"MT-control region"	AF220929	AF220934	AF220933		AY914322	AF220927	AY739627
MTCO1					M30383		
MTCO2							
MTCO3							
MTCYB	X56282						AJ245725
MTND1							
MTND2							
MTND3							
MTND4							
MTND4L							
MTND5							
MTND6							
MTRNR1	X86943	AF221590	AF221589			AF221586	AJ245724
MTRNR2							
MT-TA							
MT-TC							
MT-TD							
MT-TE							
MT-TF							
MT-TH							
MT-TI							
MT-TK							
MT-TM							
MT-TN							
MT-TP							
MT-TQ							
MT-TR							
MT-TT							
MT-TV							
MT-TW							
MT-TY							
PRNP							

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Appendix. continued

Gene	*Rhinoceros unicornis*	*Tapirus bairdii*	*Tapirus indicus*	*Tapirus pinchaque*	*Tapirus terrestris*
B2M					
DRD4					
HBE4 (Swiss-Prot)					AF139503
MC1R					
"MHC CLASS I ANTIGEN"	AJ133670				
MTATP6	NC_001779				NC_005130
MTATP8	NC_001779				NC_005130
"MT-control region"	NC_001779				
MTCO1	NC_001779				NC_005130
MTCO2	NC_001779	U83506	U83507	U83505	NC_005130
MTCO3	NC_001779				NC_005130
MTCYB	X97336		AF145734		AF056030
MTND1	NC_001779				NC_005130
MTND2	NC_001779				NC_005130
MTND3	NC_001779				NC_005130
MTND4	NC_001779				NC_005130
MTND4L	NC_001779				NC_005130
MTND5	NC_001779				NC_005130
MTND6	NC_001779				NC_005130
MTRNR1	X97336	AF191834	AY012148	AF038012	
MTRNR2	X97336		AY011182		
MT-TA	NC_001779				NC_005130
MT-TC	NC_001779				NC_005130
MT-TD	NC_001779				NC_005130
MT-TE	NC_001779				NC_005130
MT-TF	NC_001779				NC_005130
MT-TH	NC_001779				NC_005130
MT-TI	NC_001779				NC_005130
MT-TK	NC_001779				NC_005130
MT-TM	NC_001779				NC_005130
MT-TN	NC_001779				NC_005130
MT-TP	NC_001779				NC_005130
MT-TQ	NC_001779				NC_005130
MT-TR	NC_001779				NC_005130
MT-TT	NC_001779				NC_005130
MT-TV	NC_001779		AY012148		NC_005130
MT-TW	NC_001779				NC_005130
MT-TY	NC_001779				NC_005130
PRNP					

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim