J Mol Evol (1990) 30:36–42

Journal of Molecular Evolution
© Springer-Verlag New York Inc. 1990

Comparison of the α-Globin Gene Cluster Structure in Perissodactyla

J. Flint,[1] O.A. Ryder,[2] and J.B. Clegg[1]

[1] MRC Molecular Haematology Unit, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, England
[2] Zoological Society of San Diego, PO Box 551, San Diego, California 92112, USA

Summary. To investigate molecular evolution in a mammalian order with a comprehensive fossil record, we have constructed α-globin-like gene cluster maps for members of the order Perissodactyla. Although the arrangement of genes is the same in the five Equidae examined, the tapir and rhinoceros differ from each other and the horse in the position and number of their ζ genes, but not in the arrangement of their α and θ genes. In contrast to morphological work, a dendrogram derived from restriction site maps associates the tapir with the horse rather than with the rhinoceros; however, this phylogeny is not statistically significant. Among the Equidae, *Equus caballus* emerges as an outgroup, in agreement with data from other disciplines.

Key Words: DNA evolution — perissodactyl evolution — α-globin gene — ζ-globin gene — Molecular phylogeny

Introduction

Living tapirs, horses, and rhinoceroses comprise the six genera in the order Perissodactyla, which has an extensive fossil record stretching back to the early Eocene (Radinsky 1969; Savage and Long 1986; Carroll 1987). The order's initial radiation is thought to have occurred in the late Paleocene [about 55 million years ago (Myr ago)]. A general consensus exists on the phylogeny of the Perissodactyla as based on fossil evidence; there is a basal dichotomy into two clades, one comprising *Hyracotherium* plus other members of the Equidae, and the other comprising *Homogalax,* the Tapiroidea, and the Rhinocerotoidea. Division of this latter clade occurred some time after the appearance of *Heptodon* in the middle to late Eocene (Carroll 1987), one lineage leading to the modern Tapiridae and the other to the Rhinocerotoidea. According to Radinsky (1969) the Rhinocerotoidea may be polyphyletic, but there is no argument that modern Rhinocerotidae derive from Hyracodontidae, one of the Rhinocerotoidea families (Prothero et al. 1986). The origin of the modern Equidae is also well documented and is indeed one of the classic examples of morphological evolution (Simpson 1951; McFadden 1988). Few biochemical data from extant species are available to complement these morphological analyses.

New techniques, making it possible to obtain information about DNA sequences, have allowed the construction of molecular phylogenies that, in some cases, have revised phylogenies based on anatomical data (Wilson et al. 1987). To provide molecular data on the perissodactyl relationships, we have constructed α-globin gene cluster maps of one species of tapir and rhinoceros, and of five species of Equidae. Using these data, we have investigated the evolutionary history of individual genes in the cluster over the last 55 million years, and compared phylogenies derived from nuclear restriction maps with those based on morphology.

Methods

DNA Sources. High molecular weight DNA was obtained from the peripheral leucocytes of 40 domestic horses (*Equus caballus*), 4 Grevy's zebras (*Equus grevyi*), 2 onagers (*Equus hemionus onager*), and 1 specimen each of Hartmann's mountain zebra (*Equus zebra hartmannae*), donkey (*Equus asinus*), southern white rhi-

Offprint requests to: J. Flint

noceros (*Ceratotherium simum simum*), and Malay tapir (*Tapir indicus*).

Map Construction. Twenty-eight kilobases of the horse ζ–α gene clusters of both BI and BII α-globin haplotypes have previously been cloned in phage (λ L47.1) and subcloned into a variety of plasmid vectors (Clegg et al. 1984; Clegg 1987a; Flint et al. 1988). Using multiple digests of these plasmids a detailed map with a resolution down to 100 bp was constructed for 10 enzymes (Bgl II, Bam HI, Eco RI, Xba I, Hind II, Hind III, Pvu II, Pst I, Sst I, and Nco I). This map was confirmed and extended on genomic DNA by Southern blotting (as described in Old and Higgs 1983) using six probes: an Eco RI–Pst I fragment containing 500 bp 5′ to the ζ gene and 200 bp of the first exon, the entire ζ gene in a 1.6-kb Bam HI fragment, a 300-bp Hind III–Bgl II fragment containing the $\psi\zeta$ gene, an Nco I–Xba I fragment containing the α2 gene, a Pst I–Eco RI fragment containing the 3′ end of the α1 gene, a Pst I–Eco RI containing the θ gene, and a Hind III–Bgl II fragment containing a noncoding sequence 3′ to the θ gene. A search has previously been carried out for polymorphic sites in the horse α cluster (Clegg 1987a). Having established a complete 10-enzyme map for the horse and demonstrated that only 3 of the 20 sites examined were polymorphic, we constructed maps for the remaining four equid species by Southern blotting. All combinations of nine enzymes were employed (Bgl II, Bam HI, Eco RI, Xba I, Hind II, Hind III, Pvu II, Pst I, and Sst I) using end-labeled Hind III-digested λ phage as a size marker. Three probes were used initially: the Bam HI ζ-gene-containing fragment, the Nco I–Xba I α2-gene fragment, and the Pst I–Eco RI θ-gene fragment. Where differences were detected at the 3′ and 5′ ends of the complex, the 3′ Eco RI–Hind III and 5′ Pst I–Eco RI fragments were used as hybridization probes to resolve the differences more accurately. Complete maps thus were constructed for two Grevy's zebras, two onagers, one Hartmann's zebra, and one donkey. Partial maps were also obtained for two other enzymes: Dra I and Sph I. DNA from all five equids was digested with four more enzymes (Taq I, Ava I, Rsa I, and Eco RV), run in adjacent lanes, and hybridized with the α2-gene probe alone. The position of these sites was not mapped with double digests. The presence of a few sites could not be confirmed in all species because where a species-specific site occurs, sites for the same enzyme beyond the specific site cannot be visualized by the same probe; thus, the presence in the donkey of a second Sst I site 5′ to the $\psi\zeta$ gene cannot be confirmed with the available ζ probes. With the exception of the extreme ends of the complex, this problem does not apply to the horse because cloned DNA allowed the construction of a complete map.

A similar procedure was employed in constructing maps for the rhinoceros and tapir. Double digests of eight enzymes were used (Bgl II, Bam HI, Eco RI, Xba I, Hind II, Hind III, Pvu II, and Pst I) in conjunction with the ζ-, α-, and θ-gene-specific hybridization probes described above. Filters were washed at 65°C and 55°C in order to detect genes with similar sequences to the α-like globin genes. The maps obtained indicated that only a few restriction size fragments were shared with the Equidae.

Calculations. The percent sequence divergence was calculated by the iterative method of Nei (1987); differences due to the presence of a polymorphic site in one species but not another were ignored. Sites occurring in insertion/deletion regions were not counted in interspecies comparisons. Previous work has shown that in the Equidae duplicated α and ζ genes have maintained homology through gene conversion (Clegg 1987a; Flint et al. 1988). The identical restriction maps of the duplicated α genes in the tapir and rhinoceros imply the same has occurred in these species. For this reason sites occurring in duplicated genes were only counted once. The great majority of the interspecies differences thus obtained were present in only one or another of the species examined. This meant that the sites were not informative for parsimony analysis and a distance matrix method was employed for tree construction (Nei et al. 1983; Sourdis and Krimbas 1987; Sourdis and Nei 1988).

Results

Structural Comparison

An alignment of the tapir, rhinoceros, and horse maps is shown in Fig. 1, and of the Equidae in Fig. 2.

α-Globin Genes

All species have duplicated α genes; in the Equidae and the rhinoceros these genes are 4.6 kb apart, but only 3.8 kb in the tapir. Comparison of the restriction sites indicates that the difference in size can be accounted for by an insertion/deletion event occurring between the Hind II and Pvu II sites in the tapir (Fig. 1). The tapir also differs in that hybridization at lower stringencies (55°C) revealed a third α-like gene which has an internal map quite unlike that of the α-gene pair and is impossible to tie into the complex with any of the enzymes employed. These features suggest that the gene may be unattached to the cluster and may have moved as a processed pseudogene.

θ Genes

In the Equidae and the rhinoceros the θ gene can be exactly located with respect to the α1 gene because of an internal Pst I site. The intergenic distance is then 4 kb in the rhinoceros and 2.0 kb in the horse. The maximum distance in the tapir from α1 to θ is 2.5 kb, but we have assumed that it is the same as in the horse (see Fig. 1).

ζ Genes

There is greater variability in the 5′ than the 3′ end of the cluster. The horse has a complete 5′ ζ2 gene, 11 kb from a $\psi\zeta$1 gene, which is truncated in the first intron. The latter is 2 kb from the α2 gene. All equids have this ζ-gene arrangement, but there are two insertion/deletions, one near each ζ gene. One of 200 bp lies between the α2 and $\psi\zeta$ genes and (regarded as a deletion) occurs in horses but not in other species; the other is 300 bp, lies 300 bp 5′ to ζ2, and occurs only in Hartmann's zebra (Fig. 2). The rhinoceros has a single ζ gene, 5 kb from α2, whereas the tapir has two, 11 kb apart and 4 kb from the α2 gene. Mapping with both 3′ and 5′ ζ-gene-specific probes shows both tapir ζ genes to be intact, but the internal maps and sizes of the two genes differ. Assuming the Bgl II sites to be homologous, the 3′ ζ gene is a minimum of 1.8 kb long; the size of the 5′ ζ2 gene is delimited by two Bam HI sites that are 1.45 kb apart. These differences and the analogy of nonfunctioning 3′ ζ genes in other

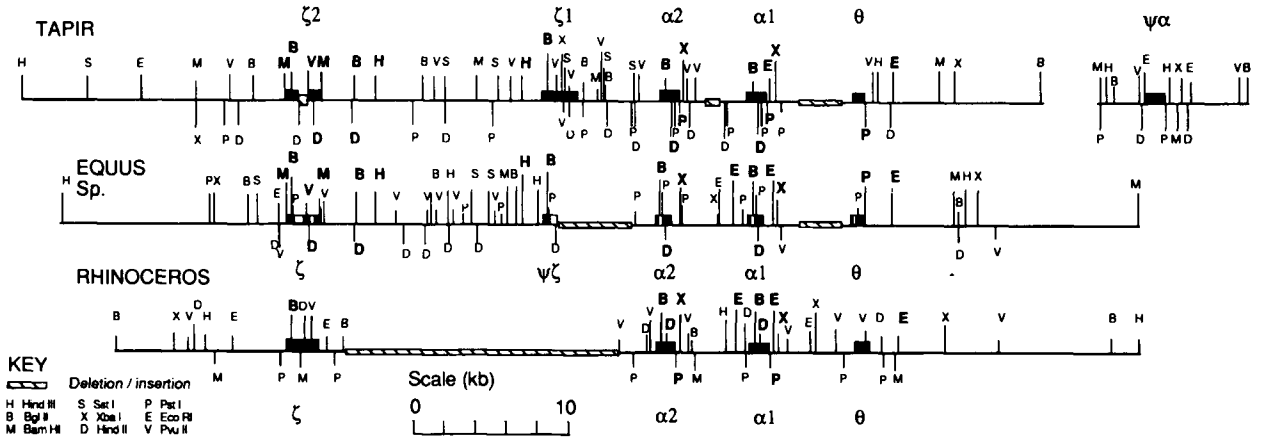

Fig. 1. Alignment of the tapir, rhinoceros, and equid maps. Insertions have been placed to align all genes and, where possible, two or more restriction sites. The equid map includes sites common to three or more of the Equidae (see Fig. 2). Sites specific to one species (and not included in this map) were not found in the tapir or rhinoceros. The equid map also includes data from mapping cloned DNA; not all the sites between the ζ genes could be visualized with the probes used, but they have been included in case they were found in other species. Sites held in common are marked in bold typeface. The position of introns is known for *E. caballus* from sequencing studies (Clegg 1984, 1987a,b; Flint et al. 1988) and is indicated by open boxes in the genes. It is not yet known where the introns occur in the rhinoceros and tapir genes, and they have therefore been omitted from the genes represented here.

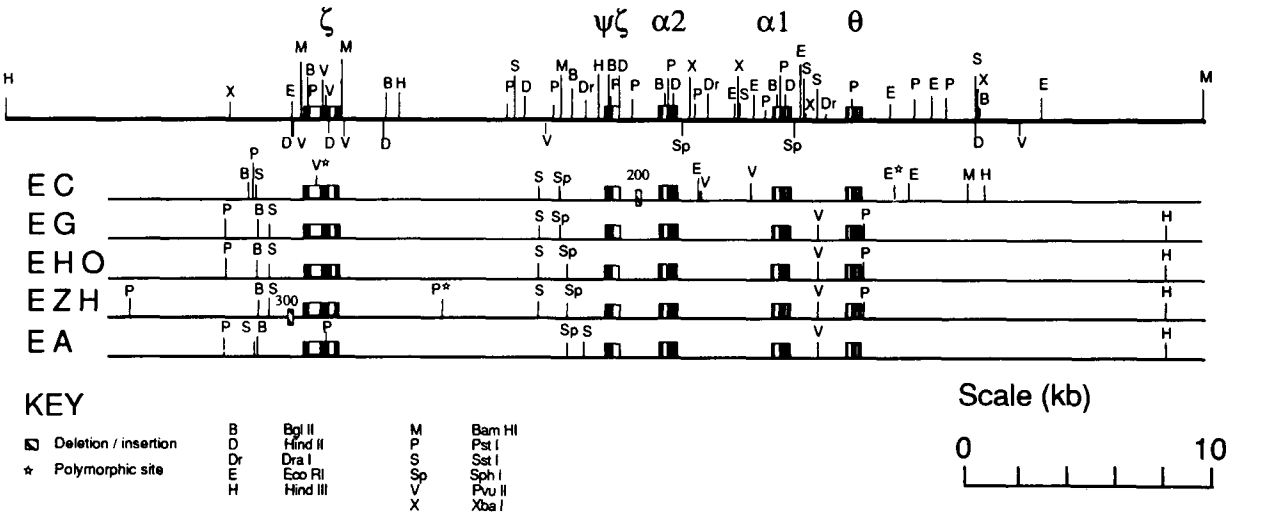

Fig. 2. Alignment of the five equid maps: sites held in common between all species are shown in the top map; sites that differ are shown in the species-specific maps below. Length differences are shown as insertions, and their size is given in base pairs. Only sites visualized by the probes used are shown. A more complete map for *E. caballus* is given in Fig. 1. Introns (shown as open boxes in the genes) are assumed to be in equivalent positions to those of *E. caballus*. Species abbreviations used: EC = *E. caballus*; EG = *E. grevyi*; EHO = *E. hemionus onager*; EZH = *E. zebra hartmannae*; EA = *E. asinus*.

species (primates and horses) suggest that the tapir ζ1 gene may be a pseudogene.

Comparison of Perissodactyl Restriction Maps

In order to see whether restriction sites were shared it was necessary to take into account deletion/insertion events that may produce different-sized electrophoretic bands, even when the sites involved are homologous. Insertions have been placed to align all genes and, where possible, two or more restriction sites. The shared sites are not randomly distributed over the whole cluster. Table 1 shows the percentage of shared sites found within 200 bp of coding regions (assuming the size of the genes is similar to the horse for which this information is known from sequencing). Averaging across all three species, 60% of genic sites are shared, whereas only 15% of sites outside genes are shared. In the comparison of the five equid species, for which 85 sites were mapped, 17 sites were different between species; only one such difference occurred in a coding region (Fig. 2).

A limited search for polymorphisms in the equid species other than the horse failed to detect the Pvu II and Eco RI polymorphisms described previously in horses (Clegg 1987a); however, one Pst I polymorphism appeared in the single Hartmann's zebra analyzed. The Eco RI site 3′ to the θ gene is linked in the horse to one of the two common haplotypes

Table 1. Position of shared sites in perissodactyl comparisons

		% shared	No. shared	Total no.
Tapir	Genes	60	12	21
	Intergenic	16	8	50
Horse	Genes	63	10	16
	Intergenic	18	9	50
Rhinoceros	Genes	64	7	11
	Intergenic	11	4	38

The first column gives the percentage of sites present in one species and also present in either of the other two species. This figure is divided into two according to whether the shared sites occur within genes or not. The second column gives the number of shared sites, and the third column the total number of sites examined; in each case the figure is again divided according to the location of sites. The size of each gene was estimated from the alignment of homologous sites to the horse (for example the Bgl II site in the horse ζ gene is 45 bp from the initiation codon; this is assumed to be the case also for the tapir and rhinoceros). Sites within 200 bp of initiation and stop codons are counted as genic

(the BI) (Clegg et al. 1984). All other equid species apparently possess only one haplotype, which is analogous to the equid BII, but have the Eco RI site; we assume that in the horse the BII haplotype has lost the Eco RI site.

Estimation of Sequence Divergence and Construction of Evolutionary Trees

An estimate of nucleotide divergence can be calculated from the fraction of sites shared between each species (Nei 1987). Results of such calculations are given in Table 2, where the horse is compared with the tapir and rhinoceros, and in Table 3, where all five equid species are compared. The fraction of shared sites is given in the lower left half of each matrix and the percentage sequence divergence in the upper right half. To achieve a fair estimate of the number of sites examined in the comparisons, sites were excluded when they occurred in DNA assumed to be deleted or inserted to achieve alignment. Thus, although 71 sites were examined in the tapir, only 44 could be counted for comparison with the horse and only 34 for comparison with the rhinoceros.

Table 2 shows that the horse and tapir are more similar than either is to the rhinoceros. Among the equid species (see Table 3), although approximately 2% sequence divergence exists between the horse and other equids, the zebras, the onager, and the donkey have less than 1% sequence divergence between them; however, because of the small number of differentiating sites, the standard errors of these estimates are large (Table 3).

In order to draw a phylogenetic tree from the data in Tables 2 and 3, we have used the unweighted

Table 2. Estimates of percentage sequence divergence between the horse, rhinoceros, and tapir

	Horse	Tapir	Rhinoceros
Horse		17.60 (3.27)	25.06 (5.36)
Tapir	13/35 3/9		23.24 (4.91)
Rhinoceros	7/29 1/4	8/28 1/6	

The lower left half of the matrix shows the number of shared sites over the average number of sites examined; the first fraction is for restriction enzymes with a 6-bp recognition sequence, and the second for Hind II, which cuts GTPyPuAC. The upper right half of the matrix gives the percentage estimate of nucleotide substitution calculated by the iterative method of Nei (equation 5.50, Nei 1987). The standard error for this figure is given in parentheses (Nei 1987)

pair-group method with arithmetic mean (UPGMA), a distance matrix method of dendrogram construction. Distance matrix methods have been shown to be superior to maximum parsimony analysis when the number of taxonomic units is small and the number of singular sites is large (in comparison to the number of sites present in two or more species) (Sourdis and Nei 1988). Despite the rate constancy assumption of UPGMA, it performs well in recovering the correct topology when there are large stochastic errors in the genetic distance estimates (Nei et al. 1983). Our data are therefore probably best interpreted by a UPGMA method. This has the additional important advantage of providing a rooted tree. The dendrogram is shown in Fig. 3.

The large standard errors of the sequence divergence estimates render the topology of the dendrogram unreliable, particularly for the branching points between the equid species where the standard errors are larger than the branch lengths. Assessment of the significance of the horse–rhinoceros–tapir triad indicates that the probability of the root being below the rhinoceros is between 0.7 and 0.8. Using a Monte Carlo simulation of possible triangles (2000 simulations) based on the means and standard errors of Table 2 gave 76.15% with the root below rhinoceros, 14.15% with the root below horse, and 9.7% with the root below tapir (Sneath 1986; P. Sneath, personal communication).

Where paleontological and other information are available to date the divergence nodes of the dendrogram, it can be used to calibrate the rate of DNA evolution. Where the DNA dendrogram agrees with the morphological, divergence times from the fossil record can be applied directly, as is the case for the origins of *E. caballus*. Horses are known to have separated from the other equids between 3 and 5 Myr ago (Simpson 1951; McFadden 1988) [mitochondrial DNA sequence divergence calculations estimate 3.7 Myr ago (George and Ryder 1986)]. Using the midpoint of this estimate gives a substitution rate of 2.5×10^{-9} bp /year. The relationship

Table 3. Estimates of percentage sequence divergence between Equidae

	EC	EG	EHO	EZH	EA
EC		2.07 (0.55)	2.23 (0.57)	1.67 (0.49)	2.39 (0.60)
EG	53/59 2/4 8/9		0.12 (0.12)	0.99 (0.36)	0.74 (0.31)
EHO	52/58 2/4 8/9	59/60 4/4 9/9		0.61 (0.28)	0.49 (0.25)
EZH	52/58 4/4 8/9	57/59 2/4 9/9	59/60 2/4 9/9		1.13 (0.39)
EA	51/58 2/4 8/9	55/58 4/4 9/9	57/59 4/4 9/9	56/59 2/4 9/9	

The lower left half of the matrix gives the fraction of shared sites for enzymes that have a 6-bp recognition sequence, those that have a 4-bp recognition sequence and the enzymes Hind II (recognition sequence GTPyPuAc) and Ava I (recognition sequence CPyCGPuG). The data come from Fig. 2 for the 6-bp recognition enzymes and for Hind II, and from four unmapped enzymes Eco RV (6-bp recognition site) Ava I, Taq I, and Rsa I. The last two enzymes have a 4-bp recognition sequence. The upper right half gives the estimated percentage sequence divergence and one standard error calculated by the method of Nei (1987). Species abbreviations used: EC = *E. caballus*; EG = *E. grevyi*; EHO = *E. hemionus onager*; EZH = *E. zebra hartmannae*; EA = *E. asinus*

between tapir, horse, and rhinoceros differs in the molecular and fossil phylogenies. We have taken the first branch of the DNA phylogeny to represent the initial radiation of the perissodactyls, thought to have occurred in the late Paleocene (Radinsky 1969), which is dated between 60.2 and 54.9 Myr ago (Harland et al. 1982). Again taking the midpoint (57.7 Myr ago), the substitution rate estimate is 2.1×10^{-9} bp/year. The good agreement between the two figures is evidence in favor of the constancy of DNA evolution in the perissodactyls, but as the errors involved in calculating the rates are large, the agreement may well have occurred by chance.

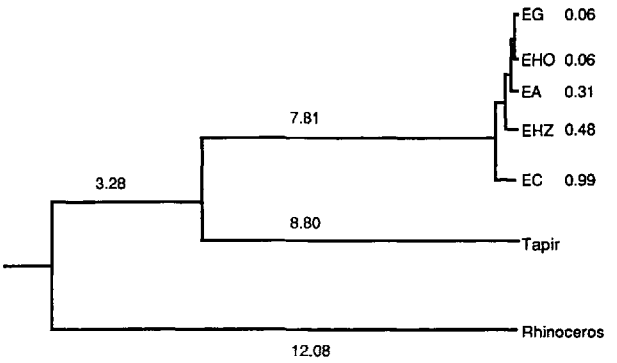

Fig. 3. The dendrogram is based on the sequence divergence estimates in Tables 3 and 4. The branching order was determined by an unweighted pair group method with arithmetic mean (Nei 1987). Species abbreviations used: EC = *E. caballus*; EG = *E. grevyi*; EHO = *E. hemionus onager*; EZH = *E. zebra hartmannae*; EA = *E. asinus*.

Discussion

Evolution of the Structure of the α Cluster in the Perissodactyla

Mammalian α-globin clusters have not been investigated as extensively as the β-globin clusters. Most attention has been devoted to the primates (the great apes and one lemur species; reviewed in Higgs et al. 1989); outside of this group, representatives of only four different orders of eutherian mammals have been examined: goat (Wernke and Lingrel 1986), horse (Clegg et al. 1984; Clegg 1987a; Flint et al. 1988), mouse (Leder et al. 1981, 1985), and rabbit (Cheng et al. 1987, 1988). In the five mammalian orders so far studied the α-globin cluster genes vary in relative position and number (Hardison and Gelinas 1986). Three different genes have been discovered: an embryonic ζ-globin gene, the adult α-globin gene and a θ gene (Clegg 1987b; Fischel-Ghodsian et al. 1987; Hsu et al. 1988). These genes are arranged in the order 5′-ζ–α–θ–3′ with duplications and triplications known for α- and ζ-globin genes but not so far for the θ-globin-like gene. The cluster's structure is further complicated by the presence of pseudogenes. The variability in globin gene arrangements, the lack of good fossil evidence linking the species investigated, and the fact that not all α-like genes have been mapped for each cluster (for instance the θ gene has not been sought in mouse or goat) has hindered evolutionary analysis (Hardison and Gelinas 1986). The perissodactyl maps described here provide an opportunity to compare directly the α-gene complex across 55 Myr.

Information on the comparative structure of different mammalian α clusters has suggested that the ancestral α-globin structure was ζ–α–α–θ (Cheng et al. 1987). However, sequence similarities between human, horse, and goat intergenic DNA indicate that the ζ-globin gene duplication probably predated the mammalian radiation (Flint et al. 1988). A ζ–ζ–α–α–θ "proto-cluster" is therefore more likely. Evidence from the perissodactyl comparisons supports this view. There are shared sites in the inter-ζ region in tapirs and horses, which indicate that a single duplication event explains why both species have two ζ genes. Unfortunately, it is impossible based on these data alone to decide whether the duplication existed before the separation of the rhinoceros; the dendrogram deduced from shared sites (Fig. 3) is equally parsimonious in terms of the number of ζ-gene duplications and losses whether there are one or two ζ genes in the stem lineage.

However, neither ζ1 gene in the horse or tapir is a perfect copy. The horse ζ1($\psi\zeta$) gene has suffered a partial deletion and the tapir ζ1 gene is bigger than the tapir ζ2 and has a different restriction map. Certainly the horse, and possibly the tapir as well, has a $\psi\zeta$1 gene. This is particularly intriguing in that the 3′ ζ1 gene in all mammalian species [with the possible exception of the rabbit (Cheng et al. 1988)] is either absent or nonfunctional (Proudfoot et al. 1982; Hill et al. 1985; Leder et al. 1985; Wernke and Lingrel 1986).

In contrast to the variability at the 5′ end of the cluster, the 3′ end is relatively stable. Thus, all lineages except rabbit possess duplicated α genes and a single θ gene. There is as yet no identifiable protein product for the θ gene, and there is still controversy as to whether it is a functional gene (Clegg 1987b; Fischel-Ghodsian et al. 1987). One indication that the gene could play a role is that it has been conserved in different species, as is found in our mapping work. Moreover, it is located in a homologous position in all the animals studied here (3′ to the α1 gene), and indeed in all mammals in which the gene has been found. Its conservation and equivalent position in relation to the α1 gene through 55 Myr supports the view that, at least in the Perissodactyla, it has a function (or had one recently), though probably not in producing a normal globin protein (Clegg 1987b; Leung et al. 1987).

Perissodactyl Phylogeny

Access to interspecies variation at the DNA level promises to provide enough differences to distinguish homology from analogy between species and hence to arrive at correct phylogenies. Sequence divergence estimates can be obtained from analysis of DNA sequences, which so far has been restricted to comparisons of only a few thousand base pairs at a single locus, or from restriction maps, which sample fewer bases, but from a larger stretch of DNA. Mitochondrial DNA has been used most for restriction enzyme mapping (Brown 1983). Although some misgivings have been voiced as to the randomness of variation present in mitochondrial DNA (Adams and Rothman 1982), the derived phylogenies have in general supported and refined morphological and biochemical studies. Nuclear DNA maps have been used in only a few interspecies comparisons, and their resolving power has not been fully assessed. One study of pheasant-like birds appears to confirm the usefulness of the technique (Helm-Bychowski and Wilson 1986). The evolutionary tree inferred from the maps agrees with that from other sources and the divergence statistics revealed a relatively constant rate of DNA substitution.

Our comparisons of the DNA maps of the perissodactyls do not, however, fit so snugly with the conclusions based on the data from pheasant-like birds, but instead confirm some of the difficulties envisaged by Adams and Rothman (1982) in their critique of the phenetic analysis of restriction enzyme data. Firstly, our derived phylogeny breaks with tradition. Previous morphological work suggested that the first branch was between a lineage leading to equids and one leading to rhinoceroses and tapirs (Radinsky 1969; Carroll 1987). DNA–DNA hybridization data suggest that rhinoceroses and tapirs are more closely related to each other than either is to horses, consistent with the conventional interpretation of the fossil data (R. Benveniste, personal communication). Because we estimate that the phylogeny given here is only about 75% reliable, we cannot, on these data alone, challenge the traditional view. The DNA data agree, however, with other biochemical comparisons (Kaminski 1979; George and Ryder 1986) when they place the horse as an outgroup.

A second problem raised by the perissodactyl nuclear DNA data is that although they may support the notion of a constant rate of evolution, this rate (2.3×10^{-9} bp/year) is different from that calculated from the pheasant data (3.7×10^{-9} bp/year) (Helm-Bychowski and Wilson 1986). There are a number of possible explanations for this difference. One is that the assumption of a constant substitution rate will not hold when comparisons are made between species with different generation times (as is the case with pheasant-like birds and perissodactyls). Thus, in an analysis of synonymous substitution rates of primates, rodents, and artiodactyls, rodents were found to have a substitution rate 4–10 times higher than primates and 2–4 times higher than artiodactyls (Li and Tanimura 1987). Examination of the substitution rates for a number of genes in primates has provided evidence for a rate slow-down in hominoid evolution (Britten 1986; Koop et al. 1986; Miyamoto et al. 1987; Sakoyama et al. 1987), and this has been attributed to a lengthening of generation time (Li and Tanimura 1987). Rhinoceroses and horses are more similar in their generation times than either is to tapirs. Tapirs become sexually mature in their second year, but female rhinoceroses do not conceive until they are between 4 and 6 years old. Obviously this situation is unlikely to have remained constant through the last 50 Myr of evolution, but the difference may have existed long enough to affect the calibration of DNA substitution rates.

Other factors may also be involved. One difficulty is that as the species diverge, the proportion of shared sites found within genes (and other conserved regions) increases relative to that in the intergenic regions. At some point, when no similarities exist

between intergenic areas, further differences can only arise in the conserved genic areas. Consequently, the substitution rate will appear to slow down. This situation has almost arisen in the rhinoceros/horse comparison (Table 1).

As the amount of DNA data increases, the difficulties mentioned here are likely to be resolved. We can then be more certain of separating analogy from homology and knowing under what circumstances the assumption of a constant substitution rate is justified, and what that rate may be. The advent of new techniques for rapidly sequencing DNA will also allow us to assess how accurate a picture of sequence divergence the restriction maps provide.

Acknowledgments We are very grateful to Dr. Michael Taylor for his comments, and to Professor P.H.A. Sneath for his advice as to the significance of our evolutionary tree. J.F. is a Foulkes Foundation fellow.

References

Adams J, Rothman ED (1982) Estimation of phylogenetic relationships from DNA restriction patterns and selection of endonuclease cleavage sites. Proc Natl Acad Sci USA 79: 3560–3564

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brown WA (1983) Evolution of animal mitochondrial DNA. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland MA, pp 52–88

Carroll RL (1987) Vertebrate paleontology and evolution. W.H. Freeman, Oxford

Cheng J-F, Raid L, Hardison RC (1987) Block duplications of a ζ–ζ–α–θ gene set in the rabbit α-like globin gene cluster. J Biol Chem 262:5414–5421

Cheng J-F, Krane DE, Hardison RC (1988) Nucleotide sequence and expression of rabbit-globin genes ζ1, ζ2 and ζ3: pseudogenes generated by block duplications are transcriptionally competent. J Biol Chem 263:9981–9993

Clegg JB (1987a) Gene conversions in the horse α globin gene complex. Mol Biol Evol 4:492–503

Clegg JB (1987b) Can the product of the θ gene be a real globin? Nature 329:465–466

Clegg JB, Goodbourn SEY, Braend M (1984) Genetic organization of the polymorphic equine α globin locus and sequence of the BII α1 gene. Nucleic Acids Res 12:7847–7858

Fischel-Ghodsian N, Higgs DR, Beyer EC (1987) Function of a new globin gene. Nature 329:397

Flint J, Taylor AM, Clegg JB (1988) Structure and evolution of the horse ζ globin locus. J Mol Biol 199:427–437

George M, Ryder OA (1986) Mitochrondial DNA evolution in the genus *Equus*. Mol Biol Evol 3:535–546

Hardison RG, Gelinas RE (1986) Assignment of orthologous relationships among mammalian α-globin genes by examining flanking regions reveals a rapid rate of evolution. Mol Biol Evol 3:243–261

Harland WB, Cox AV, Llewellyn PG, Pickton AG, Smith AG, Walters R (1982) A geologic time scale. Cambridge University Press, Cambridge

Helm-Bychowski KM, Wilson AC (1986) Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. Proc Natl Acad Sci USA 83:688–692

Higgs DR, Vickers MA, Wilkie AOM, Pretorius I-M, Jarman AP, Weatherall DJ (1989) A review of the molecular genetics of the human α-globin gene cluster. Blood 73:1081–1104

Hill AVS, Nicholls RD, Thein SL, Higgs DR (1985) Recombination within the human embryonic ζ-globin locus: a common ζ-ζ chromosome produced by gene conversion of the $\psi\zeta$ gene. Cell 42:809–819

Hsu SL, Marks J, Shaw J-P, Tam M, Higgs DR, Shen C-C, Shen C-KJ (1988) Structure and expression of the human θ globin gene. Nature 331:94–96

Kaminski M (1979) The biochemical evolution of the horse. Comp Biochem Physiol 63B:175–178

Koop BF, Goodman M, Xu P, Chan K, Slightom JL (1986) Primate eta-globin DNA sequences and man's place among the great apes. Nature 319:234–238

Leder A, Swan P, Ruddle F, D'Eustachio P, Leder P (1981) Dispersion of α-like globin genes of the mouse to three different chromosomes. Nature 293:196–200

Leder A, Weir L, Leder P (1985) Characterization, expression and evolution of the mouse embryonic ζ-globin gene. 1985. Mol Cell Biol 5:1025–1033

Leung SO, Proudfoot NJ, Whitelaw E (1987) The gene for θ globin is transcribed in human fetal erythroid tissues. Nature 329:551–554

Li WH, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

McFadden BJ (1988) Horses, the fossil record, and evolution. A current perspective. Evol Biol 22:131–158

Miyamoto MM, Slightom JL, Goodman M (1987) Phylogenetic relations of humans and African apes from DNA sequences in the ψ eta-globin region. Science 238:369–372

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II Gene frequency data. J Mol Evol 19:153–170

Old JM, Higgs DR (1983) Gene analysis. In: DJ Weatherall (ed) Methods in haematology, vol 6. pp 74–102

Prothero DR, Manning E, Hanson CB (1986) The phylogeny of the Rhinocerotidae. Zool J Linn Soc 87:341–366

Proudfoot NJ, Gil A, Maniatis T (1982) The structure of the human ζ globin gene and a closely linked, nearly identical pseudogene. Cell 31:553–563

Radinsky LB (1969) The early evolution of the Perissodactyla. Evolution 23:308–328

Sakoyama Y, Hong K-J, Byun S, Hisajima H, Ueda S, Yaoita Y, Hidenori H, Miyata T, Honjo T (1987) Nucleotide sequences of immunoglobulin ϵ genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution. Proc Natl Acad Sci USA 84:1080–1084

Savage RJG, Long MR (1986) Mammal evolution. British Museum (Natural History), London

Simpson GG (1951) Horses. Oxford University Press, New York

Sneath PHA (1986) Estimating uncertainty in evolutionary trees from Manhattan-distance triads. Syst Zool 35:470–488

Sourdis J, Krimbas C (1987) Accuracy of phylogenetic trees estimated from DNA sequence data. Mol Biol Evol 4:159–166

Sourdis J, Nei M (1988) Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol Biol Evol 5:298–311

Wernke SM, Lingrel JB (1986) Nucleotide sequence of the goat embryonic α globin gene (ζ) and linkage and evolutionary analysis of the complete α globin cluster. J Mol Biol 192:457–477

Wilson AC, Ochman H, Prager EM (1987) Molecular time scale for evolution. Trends Genet 3:241–247

Received March 29, 1989/Revised June 15, 1989