J Mol Evol (2001) 52:63–72
DOI: 10.1007/s002390010134

JOURNAL OF MOLECULAR EVOLUTION

Evolution of the Integrin α and β Protein Families

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, Columbia SC 29208, USA

Received: 22 June 2000 / Accepted: 11 September 2000

Abstract: A phylogenetic analysis of vertebrate and invertebrate α integrins supported the hypothesis that two major families of vertebrate α integrins originated prior to the divergence of deuterostomes and protostomes. These two families include, respectively, the αPS1 and αPS2 integrins of *Drosophila melanogaster*, and each family has duplicated repeatedly in vertebrates but not in *Drosophila*. In contrast, a third family (including αPS3) has duplicated in *Drosophila* but is absent from vertebrates. Vertebrate αPS1 and αPS2 family members are found on human chromosomes 2, 12, and 17. Linkage of these family members may have been conserved since prior to the origin of vertebrates, and the two genes duplicated simultaneously. A phylogenetic analysis of β integrins did not clearly resolve whether vertebrate β integrin genes duplicated prior to the origin of vertebrates, although it suggested that at least the gene encoding vertebrate β4 may have done so. In general, the phylogeny of neither α nor β integrins showed a close correspondence with patterns of α–β heterodimer formation or other functional characteristics. One major exception to this trend involved αL, αM, αX, and αD, a monophyletic group of immune cell-expressed α integrins, which share a number of common functional characteristics and have evolved in coordinated fashion with their β integrin partners.

Key words: Adhesion molecules — Gene duplication — *Hox* clusters — Integrin — Polyplodization — Protein phylogeny

Correspondence to: Austin L. Hughes, Ph.D.; *email:* austin@biol.sc.edu

Introduction

Integrins are adhesion receptors involved in cell–cell and cell–matrix interactions in animals (Bosman 1993; Gumbiner 1996; Piggott and Power 1993; Sonnenberg 1993). In vertebrates, integrins are involved in several immune system functions, including leukocyte adhesion to extracellular matrix proteins and binding of cytotoxic T lymphocytes to target cells (Shimuzu et al. 1999; Springer 1990; Wei et al. 1997). Integrins also function in signal transduction, providing a link between extracellular ligands and both cytoskeletal structures and intracellular signaling mechanisms (Ginsberg et al. 1992; Howe et al. 1998; Humphries 1996). Several integrins are known to recognize the tripeptide Arg–Gly–Asp (RGD), frequently found in matrix proteins (Sonnenberg 1993; Stark et al. 1997). The integrin receptor is a heterodimer consisting of noncovalently associated α and β chains, which are encoded by evolutionarily unrelated gene families. In mammals, there are numerous distinct α and β chains; however, only a small minority of the theoretically conceivable α–β heterodimers actually occurs in cells.

The purpose of the present paper is to use phylogenetic analysis of α and β integrin sequences to understand the evolution of these gene families. Three specific questions are addressed: (1) the time of origin of major groups of integrins relative to the origin of vertebrates, (2) the relationship between the map location of human integrin genes and their origin by gene duplication, and (3) the correspondence between gene phylogeny and integrin function.

Regarding the first of these questions, previous phylogenetic studies of the relationship between vertebrate and invertebrate β integrins have been inconclusive

Table 1. Sequences used in analyses[a]

α Integrin
Chordata
Human (*Homo sapiens*) α1 [5] (P56199), α2 [5] (P17301), α3 [17] (P26006), α4 [2] (P13612), α5 [12], α6 [2] (P08648), α7 [12] (AF032108), α8 (P53708), α9 [3] (Q13797), α10 [1] (AF074015), α11 [15] (AF109681), αV [2] (P06756), αIIB [17] (P08514), αE (P38570), αL [16] (P20701), αM [16] (P20701), αX [16] (P11215), αD (U37028)
Mouse (*Mus musculus*) α2 (Q62469), α3 (Q62470), α4 (Q00651), α5 (P11688), α6 (Q61739), α7 (I61186), α8 (AF041409), αV (P43406), αE (Q60677), αL (P24063), αM (P05555)
Rat (*Rattus norvegicus*) α1 (P18614), αD (AF021334)
Chicken (*Gallus gallus*) α1 (AB000470), α6 (P26007), α8 (P26009), αV (P26008)
Frog (*Xenopus laevis*) α3 (L43057), α4 (Q91687), α5 (Q06274), α6 (L35051)
Newt (*Pleuroldeles waltl*) αV (S60571)
Echinodermata
Lytechinus variegatus α-SU4, *Strongylocentrotus purpuratus* αP (AF177914)
Arthropoda
Drosophila melanogaster αPS1 (Q24247), αPS2 (P12080), αPS3 (U76605), CG16827 (AAF58154), CG5372 (AAF47029)
Nematoda
Caenorhabditis elegans α1 (P34446), α2 (Q03600)
Porifera
Geodia cydonium α (X97283)
β Integrin
Chordata
Human β1 [10] (P05556), β2 [21] (P05107), β3 [17] (P05106), β4 [17] (P16144), β5 (P18084), β6 [2] (P18564), β7 [12] (P26010), β8 (P26012)
Mouse β1 (P09055), β2 (P11835), β3 (O54890), β5 (AF022110), β7 (P26011)
Rat β4 (U60096)
Rabbit (*Oryctolagus cuniculus*) β8 (P26013)
Chicken β1 (P07228), β2 (S32659), β3 (S43534)
Frog β1A (P12606), β1B (P12607), β3 (I51530)
Catfish (*Ictalurus punctatus*) β2 (AF141656)
Echinodermata
Lytechinus variegatus βC (AF05907)
Strongylocentrotus purpuratus βC (AF0559607), βG (U77584), βL (AF078802)
Arthropoda
Signal crayfish (*Pacifastacus leniusculus*) β (X98852)
D. melanogaster βPS (P11584), β*ν* (L13305)
Mollusca
Planorb snail (*Biomphalaria glabrata*) β (AF060203)
Nematoda
C. elegans β-pat3 (U19744)
Cnidaria
Acropora millepora β (AF005356)
Porifera
G. cydonium β (Y18168)
Ophlitaspongia tenuis β-PO1 (AF005357)

[a] Database accession numbers are given in parentheses. Numbers in brackets are chromosome locations of human genes.

(Brower et al. 1997; Burke 1999). Such analyses have not been conducted in the case of the extensive set of vertebrate and invertebrate α integrins now available, including sequences newly discovered in the complete genome of *Drosophila melanogaster* (Rubin et al. 2000). Regarding the second question, mapping of human integrin genes has raised interesting questions regarding the mechanisms involved in these genes' origins. For example, Wang et al. (1995) noted the presence of α integrin gene clusters on human chromosomes 2, 12, and 17, which are three of the four human chromosomes bearing the developmentally important *Hox* clusters. These authors suggested that both α and β integrin genes "probably evolved in parallel with Hox genes" but did not specify what they meant by evolution "in parallel." Regarding the third question, Hughes (1992), in an analysis of the limited number of integrin sequences then available, found no close correspondence between the phylogeny of α integrins and that of β integrins with which they form heterodimers. However, many more sequences with which to address this question are now available.

Methods

Sequences used in analyses are listed in Table 1. For both α and β integrins, a representative group of available sequences was chosen, including human sequences, nonhuman mammals (usually rodent),

Table 2. Data matrix of dummy variables representing functional characteristics of α integrins[a]

	β1	β2	β3	β4	β7	Co	Ln	Fn	Vc	Ic	Fb	Vn	RGD	Im	P1	Ms	Ep
α1	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0
α2	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0
α3	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
α4	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
α5	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0
α6	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1
α7	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0
α8	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
α9	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1
α10	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
αV	0	0	1	0	0	0	0	1	0	0	0	1	1	1	1	0	0
αIIB	0	0	1	0	0	0	0	1	0	0	1	1	1	0	1	0	0
αE	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
αL	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
αX	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
αD	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
αM	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0

[a] Dummy variables are given a value of 1 if a characteristic is known to be true of a given α integrin and of 0 otherwise. Variables representing heterodimer formation with specific β integrin chains: β1, β2, β3, β4, β7. Variables representing ligands of integrins: Co (collagen), Ln (laminin), Fn (fibronectin), Vc (V-CAM), Ic (I-CAM), Fb (fibrinogen), Vn (vitronectin), RGD (binding RGD motif). Variables representing tissue expression: Im (immune system cells), Pl (platelets), Ms (muscle tissue), Ep (epithelial tissue). Based on information from Swissprot database annotations, Camper et al. (1998), Kumar (1998), Pigott and Power (1993), Sonnenberg (1993), Stark et al. (1997), and Van der Vieren et al. (1999).

nonmammalian vertebrates, and invertebrates. Amino acid sequences were aligned using the CLUSTAL W program (Thompson et al. 1994). Certain α integrins include a domain in the N-terminal portion of the mature protein that shows evidence of homology to domains in cartilage matrix protein and certain complement components (Takada and Hemler 1989) referred to as the "inserted domain" or "I-DOM." This region was deleted prior to alignment. The alignments are available from the author upon request. Because the alignment among distantly related sequences was poor in the signal peptide, transmembrane, and cytoplasmic portions, only the mature extracellular region was used in phylogenetic analyses. However, preliminary analyses using the entire sequence produced essentially identical results, although some bootstrap confidence percentages were lower (data not shown). In phylogenetic analyses, any position at which the alignment postulated a gap in any sequence was excluded from all pairwise comparisons.

Phylogenetic trees were constructed by the neighbor-joining (NJ) method (Saitou and Nei 1987) on the basis of three amino acid distances: (1) the uncorrected proportion of difference (*p*), (2) the Poisson-corrected amino acid distance, and (3) the gamma-corrected amino acid distance (Kumar et al. 1993; Ota and Nei 1994). The NJ method is known to be preferable to other commonly used methods when rates of evolution differ among branches of a phylogenetic tree (Nei 1991), as may often be true in the case of multigene families whose members have adapted to different functions. The three distances produced essentially identical results; therefore, we show only the results for *p*. When sequences are distantly related, phylogenetic tree reconstruction based on the uncorrected proportion of differences tends to outperform that based on corrections for multiple hits, presumably because the variance of correction formulas becomes high as the proportion of difference becomes large (Nei 1991).

The reliability of clustering pattern in all phylogenetic trees was tested by bootstrapping, which involves creating pseudo-samples by sampling sites from the data set with replacement and constructing phylogenetic trees based on the pseudosamples (Felsenstein 1985). In each case 1000 pseudo-samples were used.

To quantify functional similarity among integrin α chains, dummy variables were defined corresponding to three aspects of each α chain's functional biology: (1) the β chain or chains with which it can form heterodimers, (2) the major ligands of heterodimers involving that α chain, and (3) its tissue expression. The variable definitions, scores for individual α chains, and sources of information are summarized in Table 2. Principal components were extracted from the correlation matrix among the dummy variables, and principal-component scores of individual integrin α chains were used to form functional groupings.

Map locations of genes on the human *Hox* cluster-bearing chromosomes (2, 7, 12, and 17) were obtained from the Online Mendelian Inheritance in Man and Human Genome Map View resources at the National Center for Bioinformatics website.

Results

Phylogeny of α Integrins

Figure 1 shows the phylogenetic tree of α integrins. In a previous analysis with a small number of sequences, the α integrins having the I-DOM domain formed a monophyletic group apart from other α integrins (Hughes 1992). The same pattern was seen in the present analysis; the α integrins possessing this domain (the I-DOM family) clustered apart from all others, and the branch supporting this pattern received highly significant (100%) bootstrap support (Fig. 1). Therefore, although the α integrin tree itself was unrooted, the I-DOM family served as an outgroup to root the remainder of the tree, while the remaining α integrins served to root the I-DOM family tree.

Within the I-DOM family, there were two main clusters, each receiving 100% bootstrap support: (1) a cluster containing α1, α2, α10, and α11 and (2) a cluster containing αL, αM, αX, and αD (Fig. 1). αE clustered closer to the latter group, with 92% bootstrap support (Fig. 1).

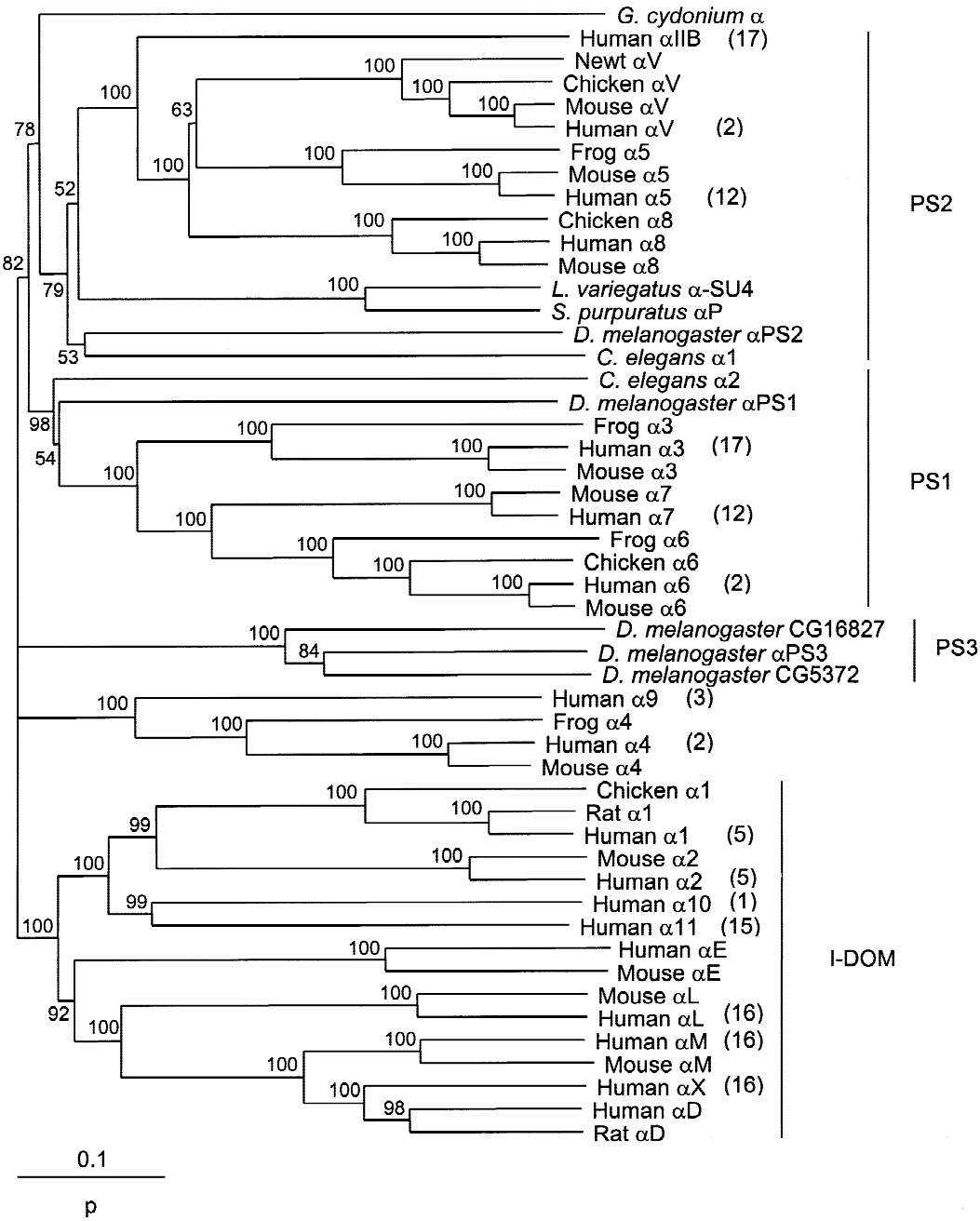

Fig. 1. Phylogenetic tree of α integrins, constructed by the NJ method on the basis of the proportion of amino acid difference (*p*). The *numbers* on the branches represent the percentage of 1000 bootstrap pseudo-samples supporting that branch; only values >50% are shown.

Because the I-DOM family contains no known invertebrate members, the phylogeny did not clarify the timing of duplication events within the I-DOM family relative to the origin of vetebrates. All members of the I-DOM family except α10 and α11 are known from both human and rodents. Because human and rodent homologues clustered together with 100% bootstrap support in every case (Fig. 1), the phylogeny supported the hypotheses that differentiation of the I-DOM family took place before the radiation of the orders of eutherian (placental) mammals. The only known nonmammalian member of the I-DOM family is α1 from chicken; the fact that chicken and mammalian α1 clustered together with 100% bootstrap support (Fig. 1) supported the hypothesis that this duplication occurred before the divergence of birds and mammals. Likewise, the phylogeny supported the hypothesis that the two major clusters of I-DOM sequences arose by a gene duplication prior to the divergence of birds and mammals.

Outside of the I-DOM family, α integrins formed four major clusters, but the relationships among these clusters were not well resolved (Fig. 1). Three of these clusters contain *Drosophila* sequences. Because these three clusters contain, respectively, the *Drosophila* sequences des-

ignated αPS1, αPS2, and αPS3, they were designated PS1, PS2, and PS3 families (Fig. 1). The PS1 family, which received 99% bootstrap support, contained vertebrate α3, α6, and α7, along with a *C. elegans* sequence and *Drosophila* αPS1 (Fig. 1). The PS2 family, which received 79% bootstrap support, contained α5, α8, αV, and αIIB, two echinoderm sequences, a *C. elegans* sequence, and *Drosophila* αPS2 (Fig. 1). The α integrin from the sponge *Geodia cydonium* also clustered close to the PS2 family (Fig. 1). In contrast, the PS3 family contained no sequences from vertebrates or any other organisms except *Drosophila*. In addition to *Drosophila* αPS3, there were two α integrin sequences revealed by complete sequencing of the *Drosophila* genome; these three *Drosophila* sequences clustered together with highly significant (100%) bootstrap support (Fig. 1). The internal branch separating the PS3 family from the PS1 and PS2 families received 82% bootstrap support (Fig. 1). The fourth cluster of α integrins lacking I-DOM contained only mammalian α4 and α9 (Fig. 1).

The phylogeny thus provided strong support for the hypothesis that the common ancestor of vertebrate α3, α6, and α7 diverged from other α integrins before deuterostomes (including vertebrates) diverged from protostomes (including insects). It also suggested, with somewhat weaker support, that the common ancestor of vertebrate α5, α8, αV, and αIIB diverged from other α integrins before deuterostomes diverged from protostomes. Likewise, it suggested that the PS3 family of *Drosophila* originated prior to the deuterostome–protostome split. The phylogeny thus suggested that the common ancestor of deuterostomes and protostomes possessed at least three types of α integrins, corresponding to αPS1, αPS2, and αPS3, and that, while the first two have diversified in vertebrates as a result of gene duplication, the last has been lost in the vertebrate lineage. In contrast, in *Drosophila*, only the PS3 family has diversified.

Within the PS1 family, the phylogenetic tree indicated that α3 was the first vertebrate gene to diverge, then α6 and α7; the branch supporting this pattern received 100% bootstrap support (Fig. 1). Because frog α3 and α6 sequences clustered with mammalian orthologues (Fig. 1), the phylogeny supported the hypothesis that vertebrate PS1 family genes duplicated before amniotes diverged from amphibians. In the case of the PS2 family, the phylogeny indicated with 100% bootstrap support that αIIB was the first to diverge (Fig. 1). However, the pattern of duplication of α5, α8, and αV was not well resolved. The tree indicated that α5 was the first of these to diverge, but bootstrap support was only 63% (Fig. 1).

Phylogeny of β *Integrins*

In the phylogenetic tree of β integrins, most vertebrate molecules were found in two main clusters, designated the β1 and β3 families in Fig. 2. The β1 family included vertebrate β1, β2, and β7 in a cluster that received 95% bootstrap support (Fig. 2). The β3 family, including vertebrate β3, β5, and β8, also received 98% bootstrap support (Fig. 2). Neither β1 nor β3 clusters included any invertebrate sequences (Fig. 2). The only vertebrate β integrin that fell outside of the β1 and β3 families was β4, and no clustering of vertebrate and invertebrate β integrins was observed in the phylogeny (Fig. 2). Thus, the phylogenetic analysis could not rule out the hypothesis that differentiation of β integrins has occurred independently in the deuterostome and protostome lineages.

Within the β1 integrin family, the phylogeny placed β1 outside β2 and β7, and this pattern received significant (96%) bootstrap support (Fig. 2). The catfish molecule identified as a β2 homologue clustered outside both β2 and β7 from mammals and birds (Fig. 2). However, the branch supporting this clustering pattern received only 81% bootstrap support. Thus, the relationships of this molecule were not definitively resolved; however, the phylogeny suggested the possibility that the gene duplication giving rise to β2 and β7 may have occurred in the tetrapod lineage after its divergence from bony fishes. On the other hand, the fact that frog homologues clustered with mammalian and avian β1 (Fig. 2) supported the hypothesis that duplication of the β1 gene occurred at least as early as prior to the amphibian–amniote divergence.

Within the β3 family, the phylogeny indicated, with 99% bootstrap support, that β8 was the first to diverge, followed by β6 (again with 99% bootstrap support), then, finally, β3 and β5 (Fig. 2). The fact that a frog homologue clustered with mammalian and avian β3 (Fig. 2) supported the hypothesis that all of these duplications preceded the amphibian–aminote divergence.

Phylogeny and Integrin Function

Figure 3 summarizes available information on integrin heterodimer formation in relation to vertebrate α and β integrin phylogenies (based on Figs. 1 and 2). As with an earlier analysis involving a smaller number of integrin chains (Hughes 1992), this figure shows that overall there is not a close correspondence between integrin phylogeny and heterodimer formation. For example, β1 forms heterodimers with certain α integrins from the I-DOM family and all three other major clusters of α integrins (Fig. 3). Likewise αV forms heterodimers with all members of the β3 family and also with β1 (Fig. 3). However, there are some cases in which phylogeny and heterodimer formation show some correlation. Most notably, the closely related αL, αM, αX, and αD all form heterodimers exclusively with β2 (Fig. 3).

When principal components were extracted from variables corresponding to functional aspects of α integrins (Table 1), the first two components accounted for 27.6

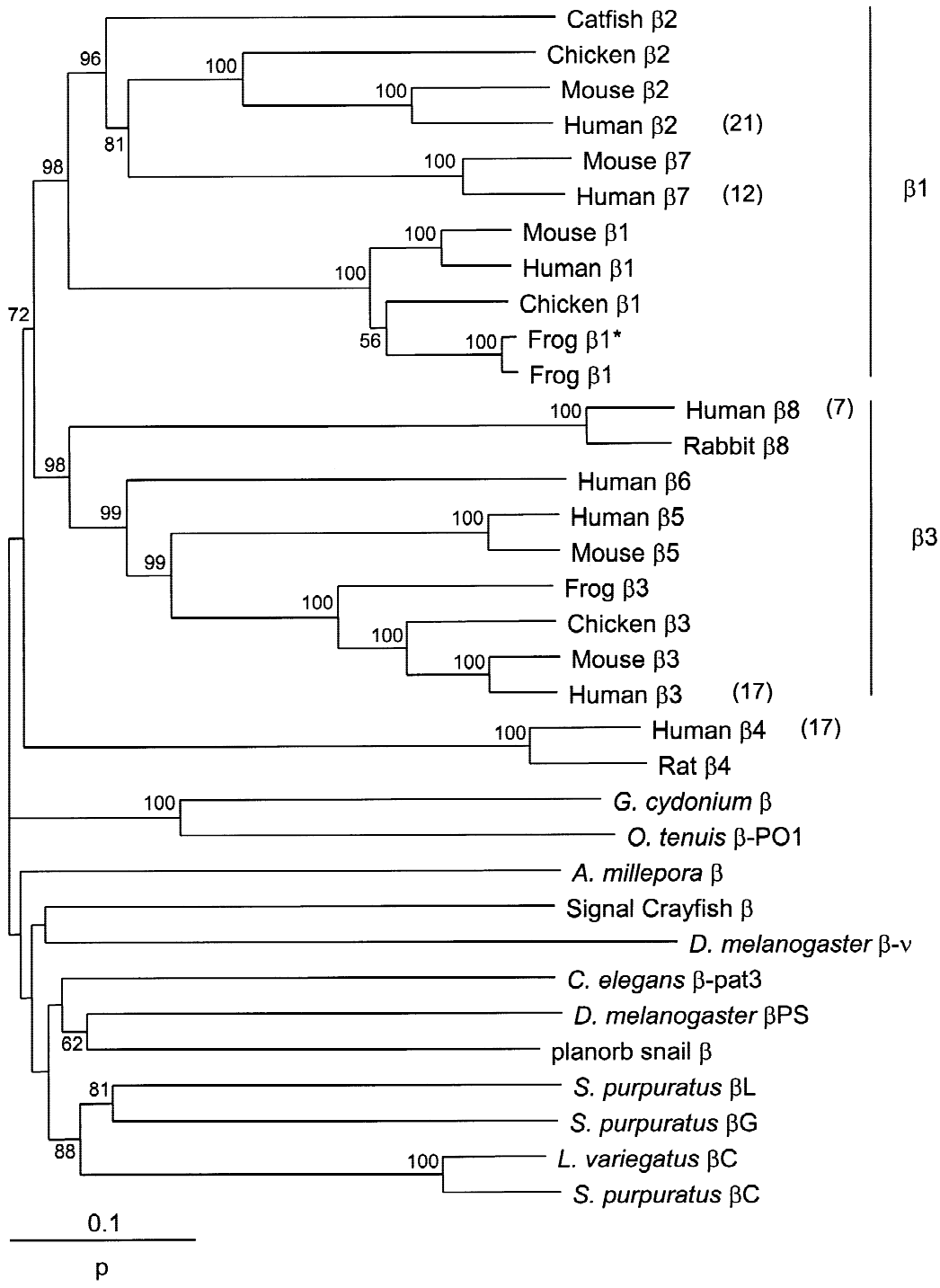

Fig. 2. Phylogenetic tree of β integrins, constructed by the NJ method on the basis of the proportion of amino acid difference (*p*). The *numbers* on the branches represent the percentage of 1000 bootstrap pseudo-samples supporting that branch; only values >50% are shown.

and 21.6% of the total variance, respectively. A plot of scores of individual a chains on these two components showed that the first component contrasted αV and αIIB with other α integrins (Fig. 4). These two α chains are unique in that they form heterodimers that bind vitronectin, and they share other characteristics as well (Table 2). The second component contrasted αL, αM, αX, and αD with other integrins. The latter constitute a functionally unique group in that they form heterodomers with β2 and have an immune system expression (Table 2). They also formed a monophyletic group in the phylogeny of α integrins (Fig. 1). This group thus represents the closest correspondence between phylogeny and function among the vertebrate α integrins.

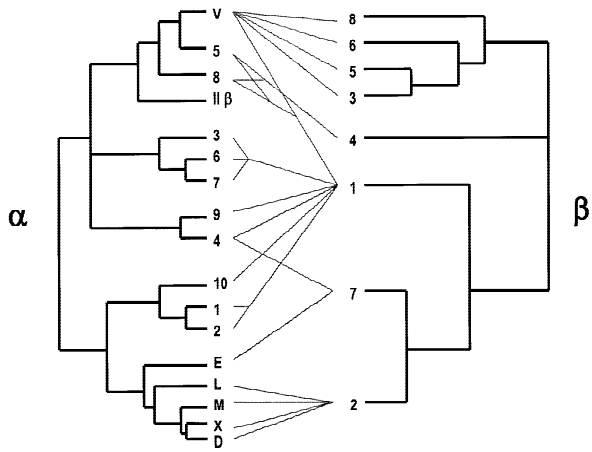

Fig. 3. Heterodimer formation among vertebrate α and β integrins; schematic α and β integrin phylogenies based on Figs. 1 and 2.

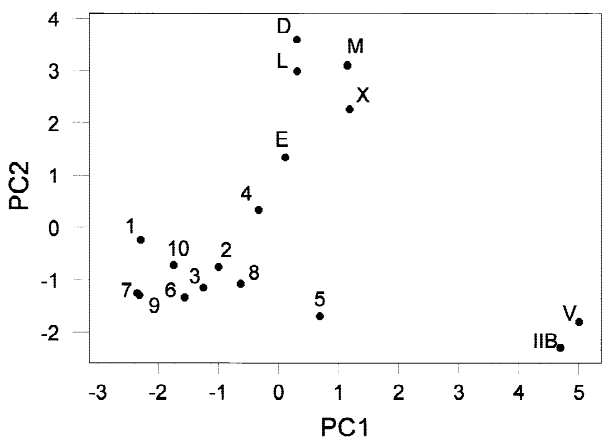

Fig. 4. Plot of scores on the first two principal components extracted from the correlation matrix of dummy variables in Table 2.

Discussion

The results of the phylogenetic analyses indicated that α integrins include ancient families that originated by gene duplication prior to the divergence of deuterostomes and protostomes, which is estimated to have occurred about 993 million years ago (Wang et al. 1999). The PS1 and PS2 families both include members in vertebrates and *Drosophila*, supporting the origin of at least these two families prior to the deuterostome–protostome divergence (Fig. 1). Although the PS3 family includes no known vertebrate members, it also probably originated prior to the deuterostome–protostome divergence. It is interesting that the PS3 family has undergone gene duplication in *Drosophila*, whereas the PS1 and PS2 families have not done so, in spite of the fact that the latter two families have extensively duplicated in the vertebrates. In contrast to the α integrins, the phylogenetic analysis did not clarify whether major groups of vertebrate β integrins arose prior to the origin of vertebrates.

α Integrins belonging to the PS1 and PS2 families are found on human chromosomes 2, 12, and 17 (Fig. 1). Members of the PS1 and PS2 families on these chromosomes show parallel phylogenies, with the gene on chrosomome 17 clustering outside those on chromosomes 2 and 12 (Fig. 1). This phylogenetic pattern suggests that, if the original duplication giving rise to the PS1 and PS2 families was a tandem duplication and if the duplicate genes remained linked, these linked pairs of genes may have been simultaneously duplicated as outlined in Fig. 5 to give rise to the present human genes. According to this scenario, the first duplication gave rise to the genes in the PS1 and PS2 families now on human chromosome 17, while the second duplication gave rise to the genes now on human chromosomes 2 and 12 (Fig. 5). This scenario gives a parsimonious explanation of the current linkage relationships among these genes in humans. As regards the assumption of a conserved linkage between PS1 and PS2 family members, it is of interest that αPS1 and αPS2 in *Drosophila* are both located on the X chromosome, though not closely linked. In contrast, the other three *Drosophila* α integrin genes are all located on 2R.

Because human chromosomes 2, 12, and 17 are among the four chromosomes bearing *Hox* clusters, it may be tempting to hypothesize that PS1 and PS2 family members on these chromosomes duplicated simultaneously with the *Hox* clusters. In the study by Zhang and Nei (1996) of *Hox* phylogeny, the relationships of the *Hox* clusters were not fully resolved, but there was strong bootstrap support for clustering of *HOXC* (on chromosome 2 in human) and *HOXD* (on chromosome 12 in human) (Fig. 6A). This phylogeny is consistent with that of human PS1 and PS2 family members (compare Figs. 5 and 6A). Thus these genes may have been duplicated simultaneously with the *Hox* clusters. Since PS1 and PS2 family members are not known from human chromosome 7, on this hypothesis it must be assumed that either (1) the duplication of the *Hox* cluster on chromosome 7 (*HOXA*) was not accompanied by duplication of α integrin genes or (2) α integrin genes originally linked to *HOXA* have been deleted.

Many biologists have cited the hypothesis that the vertebrate genome underwent two rounds of whole genome duplication by polyploidization early in vertebrate history (Ohno 1970; Sidow 1996). It is often assumed that the *Hox* clusters were duplicated as a part of these polyploidization events (Sidow 1996), and it might be argued that the duplication of α integrin genes occurred as a result of the same events. However, the hypothesis of polyploidization early in vertebrate history should be treated with some caution. Skrabanek and Wolfe, reviewing the evidence in 1998, concluded that as yet no firm support for the polyploidization hypothesis was available. More recently, Hughes (1999a) conducted the first rigorous test of a prediction of the hypothesis that

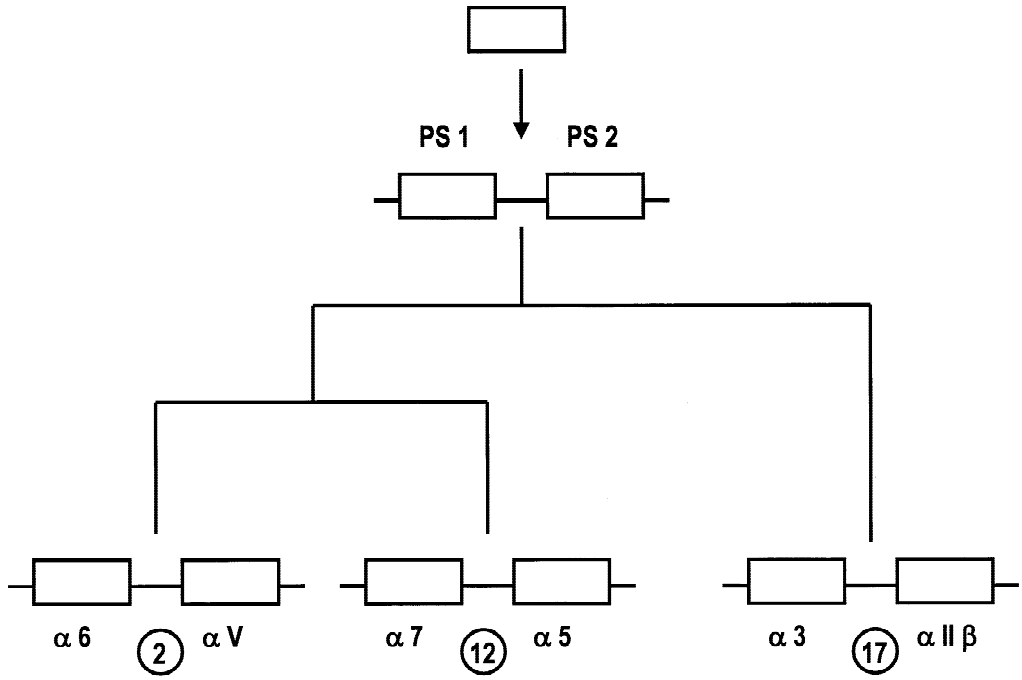

Fig. 5. Hypothetical scenario of joint duplication of α integrins on human chromosomes 2, 12, and 17 belonging to the PS1 and PS2 families.

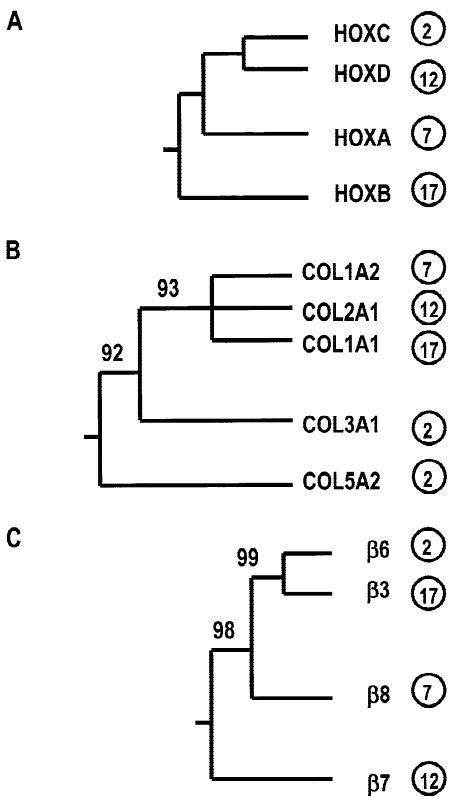

Fig. 6. Schematic phylogenies of some gene families including members on human chromosomes 2, 7, 12, and 17: (A) *Hox* clusters (Zhang and Nei 1996); (B) collagen genes (Bailey et al. 1997); (C) β integrin (based on Fig. 2).

the vertebrate genome underwent two rounds of polyploidization; the results failed to support this hypothesis.

As regards the genes on the *Hox* cluster-bearing chromosomes of humans, the evidence in support of polyploidization is very weak. As mentioned, the present phylogenetic analysis was consistent with the hypothesis that α integrins of the PS1 and PS2 families duplicated simultaneously with *Hox* clusters. However, the phylogenies of other genes on the *Hox* cluster-bearing chromosomes are not easily reconciled with the hypothesis of simultaneous duplication as would occur in polyploidization. For example, the phylogeny of the collagen genes closely linked to the *Hox* clusters (Bailey et al. 1997) revealed a phylogeny that is not easily reconciled with that of the *Hox* clusters themselves (compare Figs. 6A and B). Moreover, in the present study, the phylogeny of β integrins mapped to human chromosomes 2, 7, 12, and 17 yields a phylogeny that cannot be reconciled with either that of the α integrins or that of the *Hox* clusters (Fig. 6C). Thus, although the present phylogenetic analyses support the hypothesis that vertebrate PS1 and PS2 family α integrins were duplicated simultaneously, possibly at the same time as the *Hox* cluster duplications, they did not support the hypothesis that all β integrins on the same chromosomes were duplicated at the same time.

Map locations of most, but not all, α integrin genes on the human *Hox* cluster-bearing chromosomes show a tight linkage consistent with the hypothesis of simultaneous duplication. On chromosome 2, *HOXD* maps to

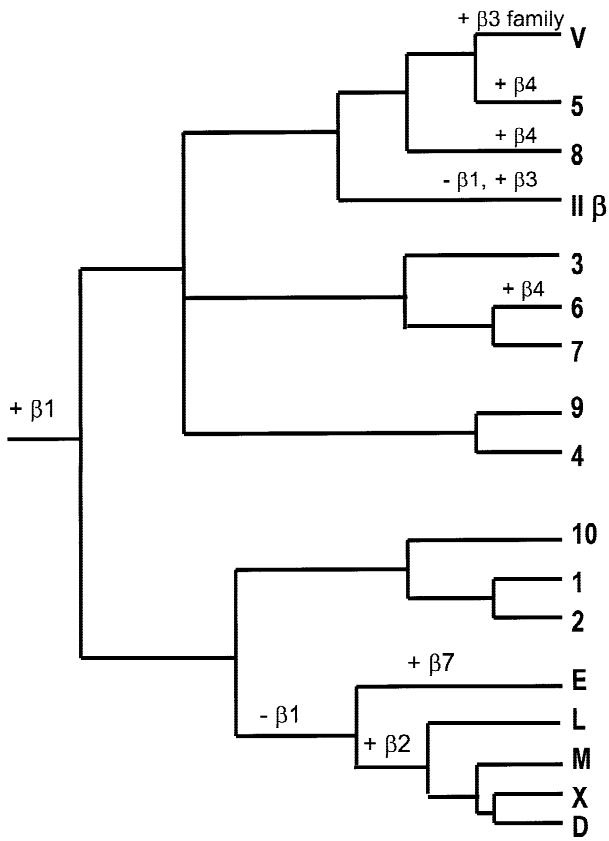

Fig. 7. Hypothetical events of gain and loss of heterodimer-forming capacity in the phylogeny of vertebrate α integrins.

2q31–q32, as does integrin αV. On the other hand, α6 maps to 2pter–p25.3. On chromosome 12, *HOXC* maps to 12q13, as does integrin α7, while integrin α5 maps to 12q11–q13. Likewise, on chromosome 17, *HOXB* maps to 17q21–q22, while integrin αIIB maps to 17q21.32. Mapping of an STS homologous to α3 suggests a position close to *HOXB.* However, there is evidence that close linkage need not reflect past simultaneous duplication events, since integrin α4, which is distantly related to the PS1 and PS2 families (Fig. 1), maps to 2q31–32, along with *HOXD.* Yet it is very unlikely that α4 duplicated along with the *Hox* clusters, and its closest relative, α9, maps to chromosome 3.

Map information on the β integrins reveals further complexities. Integrin β7 maps to 12q13.13, close to *HOXC,* while β3 maps to 17q21.32, close to *HOXB.* Thus, it seems possible that the ancestors of these two β integrin genes might have duplicated along with these two *Hox* clusters. Since, β7 belongs to the β1 family, and β3 belongs to the β3 family (Fig. 2), this would imply that these two families themselves arose through duplication of a β integrin gene along with the *Hox* clusters. This in turn is consistent with the phylogeny in Fig. 2, which suggests that these two families of β integrin genes arose within the vertebrates. On the other hand, the incongruity of *Hox* and integrin β phylogenies (Fig. 6) argues against the hypothesis that other integrin β genes were also duplicated along with the Hox clusters.

β6 has not been precisely mapped, but appears to occur somewhere in the region 2q23–q31, thus fairly close to *HOXD.* Similarly, β8 maps to 7p15–p21, close to *HOXA.* If it is true that the ancestors of the β1 and β3 families duplicated along with the *Hox* clusters, β6 and β8, both members of the β3 family, must have originated subsequently by tandem duplication. If so, the β6 gene must have been translocated from its presumed original location in linkage to *HOXB* to a new location in linkage with *HOXD.* Likewise, the β8 gene must have been translocated from linkage with *HOXB* to linkage with *HOXA.* On the other hand, β4, apparently quite distantly related to β3 (Fig. 2), was apparently translocated to chromosome 17 at some point in its evolutionary history, now mapping to 17q11–qter.

Overall, then, available map information on human α and β integrin genes suggests that, in addition to possible simultaneous duplication of certain genes with the *Hox* clusters, there have been additional independent events of duplication and translocation of integrin genes onto the *Hox* cluster-bearing chromosomes. These results thus support other results suggesting that the mere fact that members of two or more families are linked in two genomic regions need not imply that the genes involved were duplicated simultaneously (Hughes 1998).

If the α and β integrin families had coevolved closely, one might expect to find that the phylogenies of α chains corresponded to those of β chains with which they form heterodimers. Clearly this is not the case (Fig. 3). Rather, it seems that, as α-chain genes have duplicated, they have lost and acquired the ability to form heterodimers with different β chains in an opportunistic fashion. Figure 7 summarizes a parsimonious reconstruction of evolutionary changes in heterodimer formation patterns given the α integrin phylogeny of Fig. 1. According to this reconstruction, heterodimer formation with β1 was the ancestral state for known vertebrate α integrins. In the case of the I-DOM family members αE, αL, αM, αX, and αD, loss of the ability to associate with β1 seems to have coincided with duplication of β1 to form the common ancestor of β2 and β7. When the latter gene duplicated, αE was specialized for pairing with β7, while the others specialized in pairing with β2. This scenario of the evolution of heterodimer formation is thus consistent with the hypothesis that specialized gene functions arise after duplication of a more generalized ancestral gene (Jensen 1976; Hughes 1994, 1999b).

The present analyses suggest that phylogeny and gene function were not in general closely correlated in either α or β integrins. A multivariate analysis of functional traits of α integrins identified only one monophyletic group which was functionally well differentiated: the immune system-expressed I-DOM family members αL, αM, αX, and αD (Fig. 4). These molecules are also unique in that they originated in a coordinated duplication of the α- and β-chain genes (Fig. 7). Finally, these genes are unique among human integrins in that all except αD (which is unmapped) have been mapped to chromosome 16. Thus, in marked contrast to other integrin genes, this group of immune cell-expressed α integrin genes constitutes a functionally and phylogenetically distinctive group of physically linked genes.

Acknowledgments. This research was supported by grants from the National Institutes of Health and the South Carolina Commission on Higher Education.

References

Bailey WJ, Kim J, Wagner G, Ruddle FH (1997) Phylogenetic reconstruction of the vertebrate Hox cluster duplication. Mol Biol Evol 14:843–853

Bosman FT (1993) Integrins: Cell adhesives and modulators of cell function. Histochem J 25:469–477

Brower DL, Brower SM, Hayward DC, Ball EE (1997) Molecular evolution of integrins: Genes encoding integrin β subunits from a coral and a sponge. Proc Natl Acad Sci USA 94:9182–9187

Burke RD (1999) Invertebrate integrins: structure, function, and evolution. Int Rev Cytol 191:257–284

Camper L, Hellman U, Lundgren-Åkerlund E (1998) Isolation, cloning, and sequence analysis of the integrin subunit α10, a β1-associated collagen binding integrin expressed on chondrocytes. J Biol Chem 273:20383–20389

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Ginsberg MH, Du X, Plow EF (1992) Inside-out integrin signaling. Curr Opin Cell Biol 4:766–771

Gumbiner BM (1996) Cell adhesion: The molecular basis of tissue architecture and morphogenesis. Cell 84:345–357

Howe A, Aplin AE, Alahari SK, Juliano RL (1998) Integrin signaling and cell growth control. Curr Opin Cell Biol 10:220–221

Hughes AL (1992) Coevolution of the vertebrate integrin α and β chains. Mol Biol Evol 9:216–234

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond Biol Sci 256:119–124

Hughes AL (1998) Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. Mol Biol Evol 15:854–870

Hughes AL (1999a) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48:565–576

Hughes AL (1999b) Adaptive evolution of genes and genomes. Oxford University Press, New York

Humphries MJ (1996) Integrin activation: The link between ligand binding and signal transduction. Curr Opin Cell Biol 8:632–640

Jensen RA (1976) Enzyme recruitment in the evolution of new function. Annu Rev Microbiol 30:409–425

Kumar CC (1998) Signaling by integrin receptors. Oncogene 17:1365–1375

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetic analysis, version 1.01. Pennsylvania State University, University Park

Nei M (1991) Relative efficiencies of different tree-making methods for molecular data. In: Miyamoto M, Cracraft JL (eds) Phylogenetic analysis of DNA sequences. Academic Press, New York, pp 90–128

Ohno S (1970) Evolution by gene duplication. Springer Verlag, New York

Ota T, Nei M (1994) Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. J Mol Evol 38:642–643

Piggott R, Power C (1993) The adhesion molecule factsbook. Academic Press, London

Rubin GM, Yandell MD, Wortman JL, Miklos GLG, Nelson CR, Hariharan IK, Foortini ME, Li PW, Apweiler R, Fleischman W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LSB, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJM, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Zhong F, Zhong W, Gibbs R, Venter JC, Adams MD, Lewis S (2000) Comparative genomics of the eukaryotes. Science 287: 2204–2215

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Shimuzu Y, Rose DM, Ginsberg MH (1999) Integrins in the immune system. Adv Immunol 72:325–380

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715–722

Skrabanek L, Wolfe KH (1998) Eukaryotic genome duplication—Where's the evidence? Curr Opin Genet Dev 8:694–700

Sonnenberg A (1993) Integrins and their ligands. Curr Top Microbiol Immunol 184:7–35

Springer TA (1990) Adhesion receptors of the immune system. Nature 346:425–434

Stark KA, Yee GH, Roote CE, Williams EL, Zusman S, Hynes RO (1997) A novel α integrin subunit associates with βPS and functions in tissue morphogenesis and movement during *Drosophila* development. Development 124:4583–4594

Takada Y, Hemler ME (1989) The primary structure of the VLA-2/collagen receptor α_2 subunit (platelet GP1a): Homology to other integrins and the presence of a possible collagen-binding domain. J Cell Biol 109:397–407

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Van der Vieren M, Crowe DT, Hoekstra D, Vazeux R, Hoffman PA, Grayson MH, Bochner BS, Gallatin WM, Staunton DE (1999) The leucocyte integrin $\alpha_D\beta_2$ binds VCAM-1: Evidence for a binding interface between I domain and VCAM-1. J Immunol 163:1984–1990

Wang DY, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals, and fungi. Proc R Soc Lond Biol Sci 266:163–177

Wang W, Wu W, Desai T, Ward DC, Kaufman SJ (1995) Localization of the α7 integrin gene (ITGA7) on human chromosome 12q13: Clustering of integrin and hox genes implies parallel evolution of these gene families. Genomics 26:563–570

Wei J, Shaw LM, Mercurio AM (1997) Integrin signaling in leukocytes: Lessons from the $\alpha_6\beta_1$ integrin. J Leukocyte Biol 61:397–407

Zhang J, Nei M (1996) Evolution of the *Antennapedia*-class homeobox genes. Genetics 142:295–303