J Mol Evol (2006) 63:602–611
DOI: 10.1007/s00239-005-0146-8

JOURNAL OF MOLECULAR EVOLUTION

Tandem Repetitive D Domains of the Sperm Ligand Zonadhesin Evolve Faster in the Paralogue Than in the Orthologue Comparison

Holger Herlyn, Hans Zischler

Institute of Anthropology, University of Mainz, Colonel-Kleinmann-Weg 2 (SB II), D-55099, Germany

Received: 14 June 2005 / Accepted: 20 April 2006 [*Reviewing Editor:* Dr. Yves Van de Peer]

Abstract. Gene duplication is regarded as an important evolutionary mechanism creating genetic and phenotypic novelty. At the same time, the evolutionary mechanisms following gene duplication have been a subject of much debate. Here we analyze the sequence evolution of zonadhesin, a mammalian sperm ligand that binds to the oocyte zona pellucida in a species-specific manner. In pig, rabbit, and primates, precursor zonadhesin comprises, among others, one partial and four complete tandem repetitive D domains. The mouse precursor is distinguished by 20 additional partial D3 domains consisting of 120 amino acids each. This gene structure allows sequence comparison in both paralogues and orthologues. Detailed sequence analysis reveals that D domains evolve faster across paralogues than orthologues. Moreover, at the codon level, partial D3 paralogues of mouse show evidence of positive selection, whereas the corresponding orthologues do not. Individual posttranslational motif patterns and positive selection point to neofunctionalization of partial D3 paralogues of mouse, rather than subfunctionalization. However, as we found additional evidence for homogenization by partial gene conversion, sequence evolution of partial D3 paralogues of mouse might be better described as a combination of divergent and convergent evolution. So far, the divergence at the codon level has outbalanced the convergence at the level of smaller fragments. The probable driving force behind the evolutionary patterns observed is sexual selection. We finally discuss whether the functional determination influences the evolutionary regime acting on sperm ligands and egg receptors, respectively.

Key words: Sperm-egg interaction — Neofunctionalization — Subfunctionalization — Positive selection — Motif prediction — Sexual selection

Correspondence to: Holger Herlyn; *email:* herlyn@uni-mainz.de

Introduction

The vast majoritiy of gene copies (= duplicate genes, paralogues) is silenced by time due to deleterious mutations (Lynch and Conery 2000), a process called nonfunctionalization or pseudogenization. However, during the last years it became obvious that the preservation of functional duplicate genes is more common than originally assumed. Gene duplication is, hence, regarded as one of the most important molecular mechanisms creating genetic and phenotypic novelty (Kimura, Ohta 1974; Lynch and Katju 2004; Hughes 2005). Supposing functional retention, paralogues can principally undergo two alternative evolutionary fates: either they evolve divergently, thus accumulating sequential differences, or they evolve concertedly by unequal crossing-over, intragenic gene conversion, and slippage-like processes (e.g., Zimmer et al. 1980; Thomas 1998; Hughes 1999; Ohta 2000). Among the competing models describing the divergent evolution of gene copies (for a summary see, e.g., Van de Peer et al. 2001; Aguileta et al. 2004; Lynch and Katju 2004; Hughes 2005), one extreme

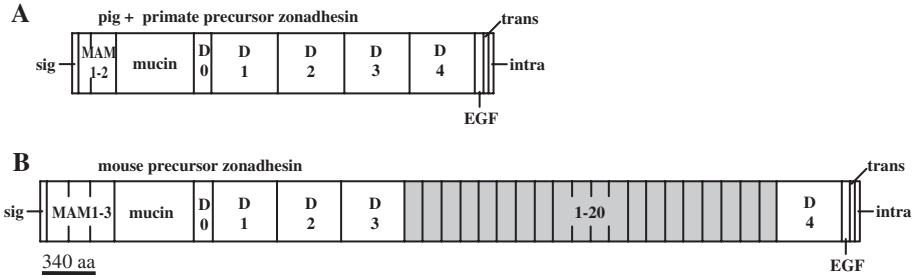

Fig. 1. Schematic domain structure of pig and primate (**A**) and mouse (**B**) precursor zonadhesin (modified after Gao and Garbers 1998). Partial D3 domains 1–20 (*1–20*) of mouse zonadhesin are highlighted in gray. D0–D4, domains D0–D4; EGF, EGF-like domain; intra, intracellular segment; MAM1-2, MAM domains 1 and 2; MAM1-3, MAM domains 1, 2, and 3; mucin, mucin-like repeat; sig, signal peptide; trans, transmembrane segment.

proposes that a succession of relaxed selective constraints and positive selection leads to the fixation of new or modified functional properties following gene duplication (''neofunctionalization'' [see Ohta 1988]). The alternative extreme, the duplication-degeneration-complementation model (Force et al. 1999), a recent modification of the model proposed by Jensen (1976) and Orgel (1977), assumes a partitioning of the original function among two paralogues under the influence of genetic drift (''subfunctionalization'' [Force et al. 1999]).

In the present study, we investigate the evolutionary patterns and processes of zonadhesin to elucidate the mechanisms following gene duplication in a tandem repetitive protein involved in sperm-egg interaction. Beyond pairwise distances, we take into account the ratio of nonsynonymous (or amino acid altering) to synonymous (or silent) nucleotide substitution rates ($d_n/d_s = K_a/K_s = p_n/p_s = \omega$) and the ratio of nonsynonymous radical to nonsynonymous conservative amino acid substitution rates (*pnr*/*pnc*). Theory predicts that both ω and *pnr*/*pnc* are < 1 when mutations cause a reduction of individual fitness (negative selection). Conversely, positive selection ($\omega > 1$; *pnr*/*pnc* > 1) is expected when amino acid changes increase the fitness of an individual. The term neutral evolution ($\omega = 1$; *pnr*/*pnc* $= 1$) consequently describes a situtation where substitutions have neither beneficial nor deleterious effects on individual fitness.

Zonadhesin is a sperm ligand for which sequence information is available from several eutherian representatives. In pig, rabbit, and human, precursor zonadhesin essentially consists of two meprin A5 antigen receptor tyrosine phosphatase mu (MAM) domains, mucin tandem repeats, and one partial (D0) and four complete (D1, D2, D3, D4) tandem repetitive von Willebrand adhesion domains (Hardy and Garbers 1995; Gao and Garbers 1998; Lea et al. 2001). Coding sequences of MAM and D domains point to an analogous structure of zonadhesin in nonhuman primates (Herlyn and Zischler 2005a, b). The structure of mouse zonadhesin differs from the outlined pattern by the presence of an extra MAM domain and 20 additional tandem repetitive partial domains derived from the C-terminus of the D3 domain (Fig. 1). The partial D3 repeats are each 120 codons long and represent more than 40% of the protein mass of mouse zonadhesin (Gao and Garbers 1998).

Here we analyse the evolution of zonadhesin D domains because of their tandem repetitive structure and their relevance in postacrosomal binding of zonadhesin to the zona pellucida of the egg (Hardy and Garbers 1994, 1995; Gao and Garbers 1998; Hickox et al. 2001; Lea et al. 2001; Bi et al. 2003), a combination that renders the zonadhesin D domains a unique subject for studying sequence evolution of a sperm ligand after intragenic duplication. We initially describe the sequence evolution of zonadhesin D domain paralogues in comparison to the corresponding orthologues. Subsequently, we focus on the evolution of the partial D3 repeats of mouse zonadhesin. The results are discussed in the light of the aforementioned evolutionary concepts.

Materials and Methods

Sampling and Alignments

Sequences coding for zonadhesin D domains of pig (*Sus scrofa*), house mouse (*Mus musculus*), European rabbit (*Oryctolagus cuniculus*), and human (*Homo sapiens*) were taken from GenBank (accession nos. U40024, NM_011741, AF244982, and AF332975). The nonhuman primate sample, consisting of sequences from the gray mouse lemur (*Microcebus murinus*), common squirrel monkey (*Saimiri sciureus*), cotton-top tamarin (*Saguinus oedipus*), white-tufted-ear marmoset (*Callithrix jacchus*), hamadryas baboon (*Papio hamadryas*), and crab-eating macaque (*Macaca fascicularis*), was published elsewhere (Herlyn and Zischler 2005b [accession nos. AY428845, AY428847, AY428849, AY428853, AY428855, and AY428857]).

Based on the coding sequences, 18 alignments were generated: one comprising the 20 partial D3 repeats of mouse (''partial D3 paralogues of mouse''), one including the corresponding D3 domain fragments of the 10 species listed above (''partial D3 orthologues''), and one combining paralogues of mouse and orthologues (''partial D3 paralogues of mouse + partial D3 orthologues''). Domains D0–D4 were separately aligned for each of the 10 species

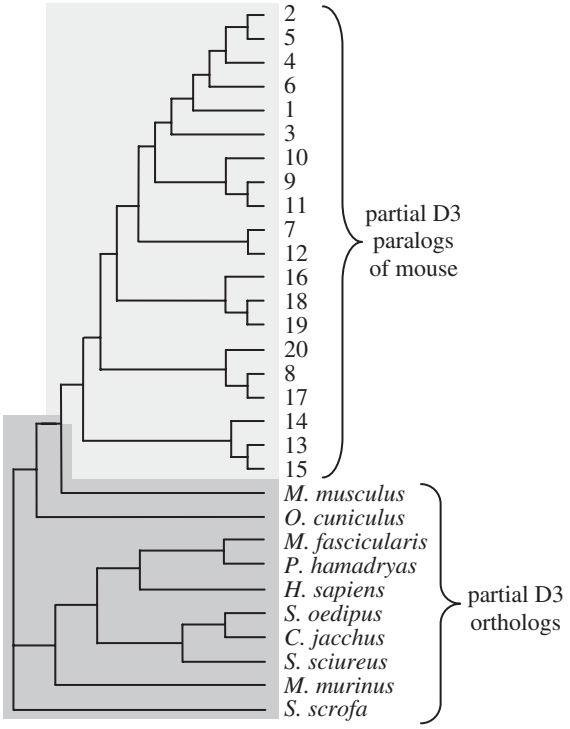

Fig. 2. Intree used for present branch-site analysis of partial D3 paralogues of mouse (*1–20*) and partial D3 orthologues of mouse (*M. musculus*), rabbit (*O. cuniculus*), etc. Different background shadings distinguish the paralogue clade (light gray) from the remaining branches (dark gray). The paralogue phylogeny represents the outcome of present tree reconstruction (MEGA, neighbor joining; see Fig. 3C for bootstrap values). The orthologue phylogeny represents the widely accepted view (Murphy et al. 2001; Smith and Cheverud 2002).

("D0–D4 paralogues of mouse," "D0–D4 paralogues of pig," etc.). Furthermore, datasets were assembled which each contained orthologues of a single domain ("D0 orthologues," "D1 orthologues," "D2 orthologues," "D3 orthologues," "D4 orthologues"). The nucleotide sequences of each dataset were translated, aligned, and retranslated using ClustalW (default settings) implemented in BioEdit version 7.0.1 (Hall 1999).

Pairwise Sequence Comparison and Saturation

To assess the evolutionary regime acting on zonadhesin D domains in the paralogue-orthologue comparison, we intitially determined the mean pairwise distance (d_n plus d_s) between the sequences of each of the 18 alignments, using the Kimura two-parameter model of sequence evolution implemented in MEGA 3.0 (Kumar et al. 2004). Gaps were deleted in pairwise comparisons only. Standard errors (SEs) were estimated from 1000 bootstrap replicates. Subsequently, we checked the alignments for saturation using DAMBE (Xia and Xie 2001). The test for saturation implemented in DAMBE (Xia et al. 2003) compares an entropy-based index of substitution saturation to a critical value inferred from simulation.

Further analysis focused on possible differences in the sequence evolution of partial D3 paralogues of mouse and partial D3 orthologues. MEGA 3.0 was used to infer mean pairwise d_n and d_s estimates, using the modified Nei and Gojobori method (Zhang et al. 1998) with a transition/transversion ratio = 3 (estimated by PAML; see below) and Jukes-Cantor correction. One thousand bootstrap replicates were generated to infer the SE for the mean of d_n and d_s. Gaps were deleted in pairwise comparisons. Furthermore, we calculated the mean d_n/d_s ($=\omega$) values of partial D3 paralogues of mouse and partial D3 orthologues. Finally, the program SCR3 was run to infer mean pairwise *pnr* and *pnc* on the basis of three different amino acid classifications (charge, polarity, and polarity and volume), applying the method of Hughes et al. (1990). For SCR3 gaps had to be stripped from the datasets. The transition/transversion ratio was set at 3.0 (estimated by PAML; see below). Standard errors for the mean of *pnr* and *pnc* were estimated applying the method of Nei and Jin (1989).

Gene Conversion and Phylogeny

We checked the partial D3 paralogues of mouse for complete and partial gene conversion using GENECONV (default settings [Sawyer 1989]). GENECONV performs statistical tests for detecting gene conversion on the basis of imbalances in the distribution of segments among homologous DNA sequences. Moreover, we recontructed the phylogeny among the partial D3 repeats of mouse using the neighbor-joining algorithm implemented in MEGA 3.0. The tree was rooted with the homologous fragment of the corresponding full D3 domain. Bootstrap values were calculated for each internal branch from 550 replicates. Gap containing positions were removed in pairwise comparisons only. The remaining settings were as default.

Branch-Site Analysis

In analogy to earlier studies on paralogue evolution (e.g., Torgerson and Singh 2004), we tested for lineage specificity of ω, using branch-site model B and model M3 ($K = 2$) implemented in the maximum likelihood framework of PAML 3.13d (Yang and Nielsen 2002). Model B allows for positive selection across a user-defined foreground and across all branches of a given phylogeny ("background"), while the corresponding null model M3 ($K = 2$) does not distinguish between fore- and background. Model B includes five freely estimated parameters, i.e., (1) proportion and (2) ω estimates of one site class conserved across the background, (3) proportion and (4) ω estimates of a second site class weakly constrained, neutral, or even positively selected across the background, and (5) ω of those sites that are under positive selection across the foreground. In contrast to model B, the null model M3 ($K = 2$) infers solely three freely estimated parameters from the entire dataset, i.e., (1) the proportion estimate of nonpositively selected sites plus (2) the according ω value and (3) the ω of the positively selected site class. The intree combined the paralogue phylogeny inferred from the present neighbor joining tree reconstruction with the widely accepted phylogeny among primates, Glires (rabbit and mouse), and pig (Murphy et al. 2001; Smith and Cheverud 2002) (Fig. 2). A principal drawback of PAML (and other programs) is that it does not include effective controls for the stochastic variation of d_n and d_s, which leads to the identification of false positives (Hughes and Friedman 2005). Moreover, we cannot rule out that uncertainties in the tree reconstruction promote the identification of false positives. To counteract both concerns, we considered only codon sites that got highly significant support ($p_{(\omega>1)} > 0.99$) as candidates for positive selection.

Based on an alignment combining partial D3 orthologues and paralogues ("partial D3 paralogues + orthologues"), we first defined the partial D3 paralogue clade as foreground. In the second run, we defined the remaining phylogeny covering the lineages leading to the orthologues as foreground (Fig. 2). Significance of the findings was assessed by a likelihood ratio test (LRT), comparing the log likelihood values (l) of model B and model M3 ($K = 2$). For LRT, twice the log likelihood difference ($2\Delta l$) of model B and model M3 ($K = 2$) was compared to critical values (cv) from a chi-square distribution equal to the difference in the

number of degrees of freedom between both models, i.e., 5 − 3 = 2. To correct for twofold testing (first test, paralogue phylogeny as foreground; second test, remaining branches as foreground), strict Bonferroni adjustment was carried out.

Motif Search

The translation products of partial D3 paralogues were scanned for amino acid motifs, using the PROSITE database implemented in the PredictProtein server (http://cubic.bioc.columbia.edu/predictprotein/). PROSITE identifies protein families and motifs by weighted comparison of DNA sequences with profiled database entries (Hofmann et al. 1999).

Results

Pairwise Sequence Comparison and Saturation

MEGA 3.0 infers high mean pairwise distances (d_n plus d_s) between D0–D4 paralogues of mouse, D0–D4 paralogues of rabbit, D0–D4 paralogues of pig, etc. Depending on the species (pig, rabbit, mouse, seven nonhuman primates, human), the values range from 0.877 to 1.063. Compared to this, the mean pairwise distances between D0 orthologues, D1 orthologues, D2 orthologues, D3 orthologues, and D4 orthologues are low (0.169–0.232). DAMBE indicates saturation for D0–D4 domains in the paralogue comparison, while little to no saturation is indicated for the orthologue comparison of domains D0–D4 ($p = 0.000$ each). As the sequence evolution of D domain orthologues has been described elsewhere (Swanson et al. 2003; Herlyn and Zischler 2005b), and a more detailed analysis of the D0–D4 paralogues of mouse, D0–D4 paralogues of rabbit, D0–D4 paralogues of pig, etc., does not appear reasonable given their apparent substitution saturation, subsequent analyses will focus on the sequence evolution of partial D3 paralogues of mouse and partial D3 orthologues.

DAMBE rules out noteworthy saturation for partial D3 paralogues of mouse and partial D3 orthologues, as well as for the merged dataset comprising both partial D3 paralogues and partial D3 orthologues ($p = 0.000$ each). In line with the results described in the previous paragraph, the mean pairwise MEGA distance (d_n plus d_s) of partial D3 paralogues of mouse (0.313) exceeds the corresponding mean of partial D3 orthologues (0.235). The relation even holds when comparing the difference of paralogue mean minus SE with the sum of orthologue mean plus SE (Fig. 3A). MEGA estimates of mean pairwise d_n and d_s confirm an increase in sequence diversity between partial D3 paralogues of mouse, compared to partial 3D orthologues (Fig. 3B). Final evidence for an acceleration of sequence evolution across paralogues comes from SCR estimates of *pnr* and *pnc*. Thus, the difference between paralogue mean and SE permanently exceeds the corresponding sum of orthologue mean plus SE, whichever amino acid classification is used (charge, polarity, polarity and volume) (Fig. 3B). Taken together, the different distance measures suggest that more substitutions accumulated in partial D3 paralogues of mouse than in partial D3 orthologues. Moreover, the difference between d_n of paralogues and orthologues is greater than the difference between d_s of paralogues and orthologues. Consistently, paralogues and orthologues differ more in *pnr* than in *pnc*, when underlying a charge- and polarity-based amino acid classification (Fig. 3B). The outlined accumulation of substitutions in paralogues is thus more pronounced regarding nonsynonymous and radical exchanges than with respect to synonymous and conservative exchanges. In line with this the mean pairwise ω value ($= d_n/d_s$) is higher in the paralogue ($= 0.511$) than in the orthologue comparison ($= 0.378$; not shown).

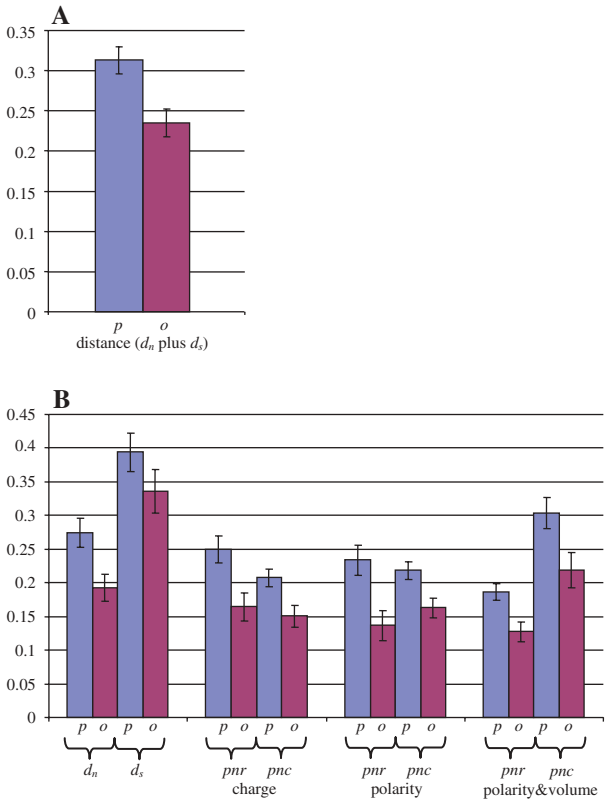

Fig. 3. Bar chart showing the means of different distance values inferred from pairwise sequence comparisons of partial D3 paralogues of mouse (*p*) and partial D3 orthologues of 10 mammalian species (*o*). The vertical lines represent standard errors. **A** Mean distances (d_n plus d_s) inferred from pairwise MEGA estimates. **B** Mean d_n, d_s, *pnr*, and *pnc* values inferred from pairwise MEGA and SCR3 estimates, respectively. Note the permanent increase in the values in paralogues, compared to orthologues, whichever parameter is taken. d_n, rate of nonsynonymous substitutions; d_s, rate of synonymous substitutions; *pnc*, rate of conservative nonsynonymous substitutions; *pnr*, rate of radical nonsynonymous substitutions. Note: *pnc* and *pnr* were inferred from amino acid classifications by charge, polarity, and polarity + volume.

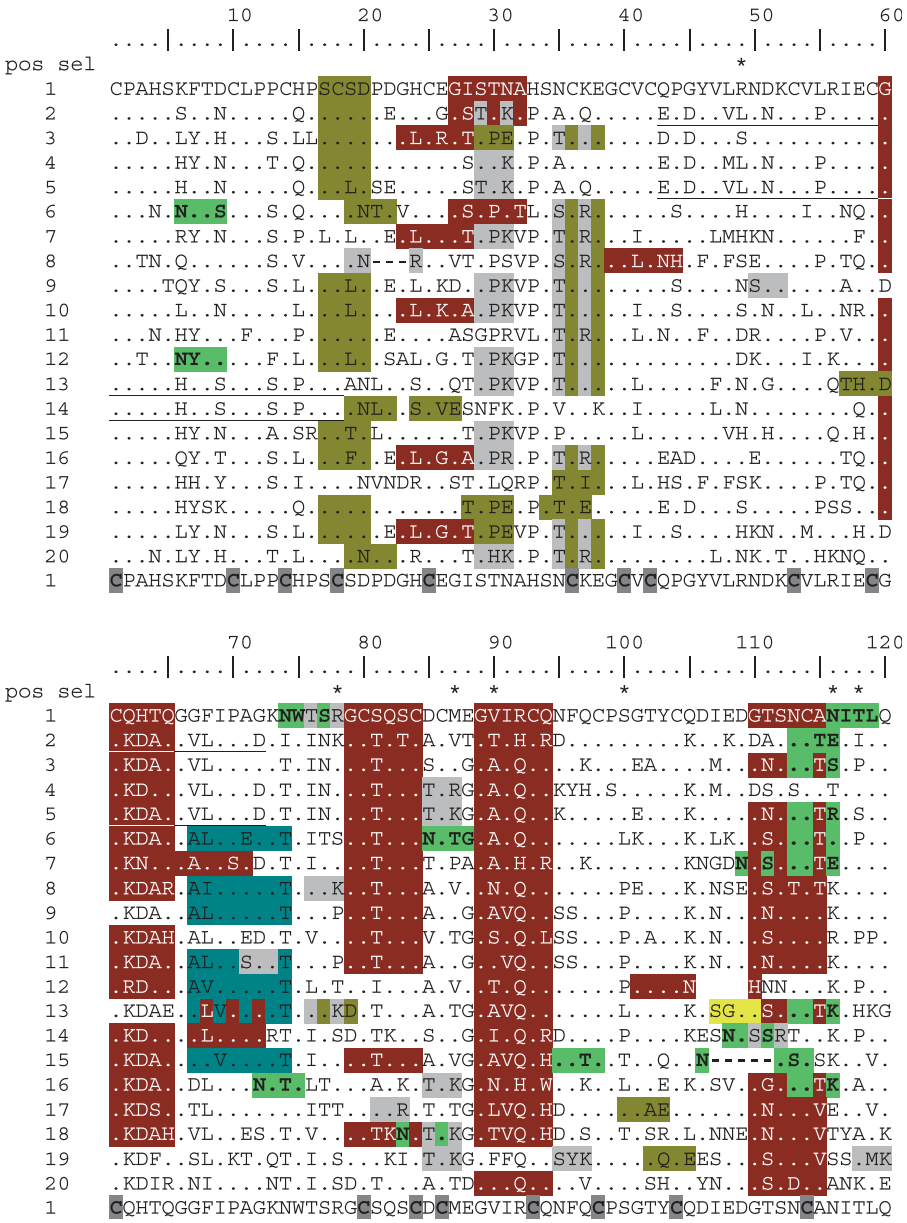

Fig. 4. Positive selection and motif distribution across partial D3 paralogs of mouse (*1-20*). Only the amino acid sequence of the reference repeat (*1*) is shown in detail. Dots in the alignment indicate congruencies with partial repeat 1. Amino acid sites suggested to be under positive selection with $p_{(\omega>1)} > 0.99$ (PAML model B) are highlighted by an asterisk above the alignment (*pos sel*). Shaded C's in the lower reference indicate conserved cysteins. Predicted motifs are highlighted by differential shading. Striations point to overlapping motifs (for instance, amino acid position 35 is the first site of a protein kinase C and a casein kinase II phosphorylation motif). Consensus sequences of motifs (PredictProtein/PROSITE): [AG].{4}GK[ST] ATP/GTP-binding site motif A (P-loop); [ST].{2}[DE] casein kinase II phosphorylation site; SG.G glycosaminoglycan attachment site; N[^P][ST][^P] N-glycosylation site; G[^EDR KHPFYW].{2}[STAGCN][^P] N-myristoylation site; [ST].[RK] protein kinase C phosphorylation site.

Gene Conversion and Phylogeny

GENECONV identifies two segments of 89- and 53-bp length that are identical between mouse partial D3 paralogues 2 and 5 ($p = 0.03$) and 13 and 14 ($p = 0.035$), respectively (see underlining in Fig. 4). As the entire 53-bp fragment and 94% of the 89-bp fragment are located in one exon, gene conversion indeed represents the probable explanation for the observed fragment identities. Judged from our data, gene conversion contributed to the homogenization of partial D3 paralogues, not to their diversification (Fig. 4). Neighbor joining tree reconstruction (MEGA 3.0) yields a fully resolved phylogeny for the partial D3 paralogues of mouse. Figure 5C illustrates that none of the terminal paralogue branches is particularly long and that support from bootstrap is rather low. In detail, only few sequences group together when underlying a minimum bootstrap support of 50%, i.e., repeats 13 and 15, 9 and 11, and 4 and 5 plus 2 (Fig. 5C). Both the possibility of gene conversion and the overall low support from bootstrapping reflect the incertainties of the presented tree reconstruction.

Branch-Site Analysis

PAML model B pinpoints sites under positive selection only when specifying the paralogue clade, and not when defining the remaining (orthologue) phylogeny as foreground (see different shadings in Fig. 2). In detail, model B suggests 21 codon sites to be under positive selection ($p_{(\omega>1)} > 0.5$; mean $\omega = 1.572$) across the paralogue foreground. To minimize a biasing effect of the potentially wrong

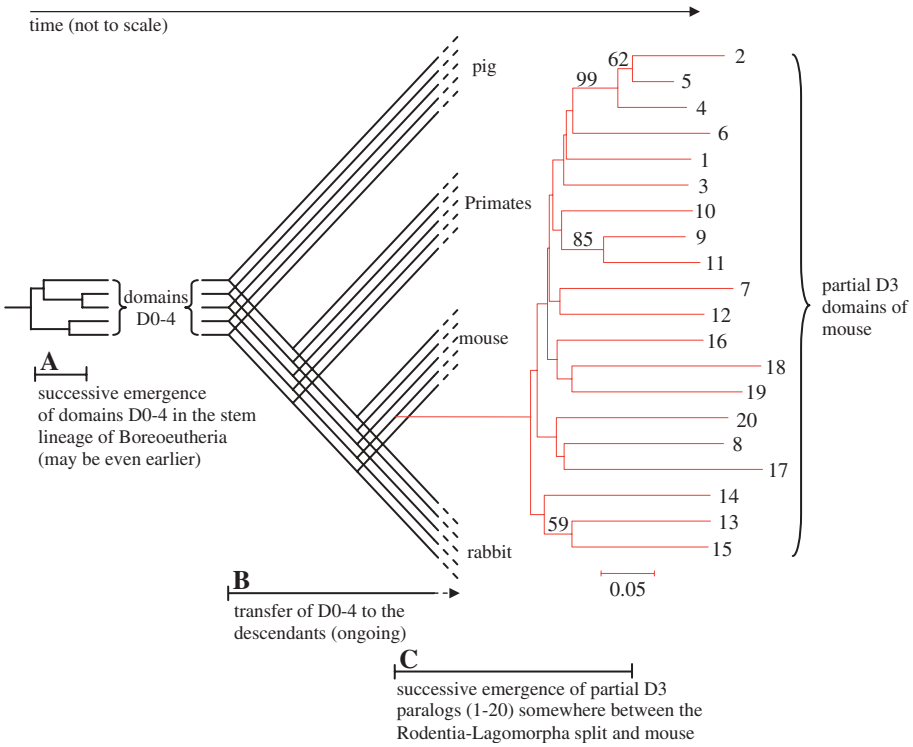

Fig. 5. Evolution of zonadhesin D domains. Rectangular trees are paralogue trees and visualize subsequent gene duplications (A, C). The tree in the center shows the phylogeny of the species and taxa used in the present study (B). The horizontal time axis above the scheme is not drawn to scale. **A** Successive emergence of domains D0–D4 in the stem lineage of Boreoeutheria or earlier. The pattern of subsequent gene duplications is hypothetical. **B** Once established, domains D0–D4 were transferred to the descendants of the last boreoeutherian stem species. To simplify visualization the species tree is truncated (dashed lines). In reality, the lineages of the orthologues reach present day. **C** Somewhere between the divergence of Glires (i.e., mouse-rabbit split in the figure) and mouse, an initial gene duplication of the C-terminus of domain D3 gave rise to the subsequent emergence of 20 partial D3 paralogues altogether (*1–20*). The tree was calculated using MEGA 3.0 (neighbor joining). Only bootstrap values >50 are shown. The bar below represents substitutions per nucleotide position.

paralogue phylogeny shown in Fig. 3C on the results, we accepted solely codon sites with highly significant support ($p_{(\omega>1)} > 0.99$) as candidates for positive selection, i.e., codon sites 49, 78, 87, 90, 100, 116, and 118 (see asterisks in Fig. 4). LRT supports the hypothesis of lineage specificity with high significance when specifying the paralogue phylogeny as foreground ($p \ll 0.01$). Conversely, taking the remaining phylogeny as foreground does not fit the data better than the null model M3 ($K = 2$) that does not distinguish between fore- and background (Table 1). As Fig. 3B shows an increase in all distance parameters in the paralogue-orthologue comparison, we rule out that the indication of positive selection across partial D3 paralogues results from a decrease in d_s and *pnc*. Thus, results of present branch-site analyses specify the findings inferred from pairwise sequence comparisons in that sequence evolution is enhanced in partial D3 paralogues of mouse due to positive selection of single codon sites.

Motif Search

Motif search (PredictProtein/PROSITE) reveals several phosphorylation, N-glycosylation, and myristylation motifs, as well as one glycosaminoglycan attachment and one ATP/GTP-binding site, across the partial D3 paralogues of mouse (Fig. 4). None of the motifs identified is conserved throughout all paralogues. In contrast to the nonconserved motifs, 18 cysteines that are also present in the C terminus of the complete D3 domain are conserved throughout the 20 partial D3 paralogues of mouse (see Cs in Fig. 4).

Discussion

Duplication Events and Sequence Evolution

The congruent presence of zonadhesin domains D0–D4 in pig, rabbit, mouse, and primates including humans (Hardy and Garbers 1995; Gao and Garbers 1998; Lea et al. 2001, Herlyn and Zischler 2005a, b) suggests an old phylogenetic origin of these paralogues. Presumably, they were realized already in the stem species of Boreoeutheria about 88 million–101 million years ago (Springer et al. 2003; see also Penny et al. 1999; Eizirik et al. 2001). Unfortunately, zonadhesin orthologues have not yet been published and annotated, respectively, for more basal taxa such as Xenarthra, Afrotheria, and Marsupialia (Murphy

Table 1. LRT statistics (PAML) based on the dataset comprising partial D3 paralogues and orthologues

Foreground	Model	*l*	LRT
Resolved paralogue phylogeny	M3 ($K = 2$)	−5704.950	$2\Delta l = 52.844$
	Model B	−5678.528	$cv = 9.21$, $p << 0.01$
Remaining phylogeny	M3 ($K = 2$)	−5704.950	$2\Delta l = 0$
	Model B	–5704.950	Not significant

Note. *cv*, critical value from chi-square distribution, given 2 df and correction for twofold testing (see below); $2\Delta l$, twofold difference between the *l* values of model B and M3 ($K = 2$); *l*, log likelihood; LRT, likelihood ratio test for positive selection; M3 ($K = 2$), discrete model with two site classes (null model); model B, branch-site model B (alternative model); *p*, significance level corrected for twofold testing (first test, paralogue phylogeny as foreground; second test, all branches except for those representing the paralogue phylogeny as foreground).

et al. 2001). Therefore, at the moment it is not possible to decide whether all five paralogues emerged in the boreoeutherian stem lineage or whether part or all of them evolved even earlier. In any case, it can be assumed that the D0–D4 domain paralogues evolved by subsequent gene duplication events before the last common ancestor of Boreoeutheria split into its descendants (Fig. 5A). The consequently older phylogenetic age of D0–D4 paralogues compared to D0–D4 orthologues (Figs. 5A and B) explains, at least partially, why the mean distance is higher in D0–D4 paralogues (0.877 to 1.063) than in D0–D4 orthologues (0.169 to 0.232).

Though we cannot assess to what extent different ages and substitution rates might have contributed to the higher distances among D0–D4 paralogues, compared to D0–D4 orthologues, a relative assessment is possible in the case of partial D3 paralogues and orthologues. Given the present sampling, partial D3 paralogues are confined to mouse. Thus, their origin can be estimated to be somewhere between the split of Rodentia and Lagomorpha 81 million–94 million years ago and today (Springer et al. 2003). Whatever the exact emergence time of each partial D3 paralogue may be, they are clearly younger than the D3 orthologues (Figs. 5B and C). Considering the increased pairwise distances among partial D3 paralogues, on the one hand, and the higher phylogenetic age of D3 orthologues, on the other, it becomes obvious that the D3 paralogues of mouse must have evolved with a higher substitution rate. Not only the pairwise distances (d_n plus d_s), but also ω, *pnr*, and *pnc* (see Fig. 3) underline an acceleration of sequence evolution across partial D3 paralogues of mouse compared to partial D3 orthologues. Despite this enhancement of sequence evolution, we found no noteworthy evidence for saturation due to multiple exchanges at single nucleotide positions in partial D3 paralogues of mouse. On the other hand, hints for occasional gene conversion events of shorter fragments encoded by one exon have been found (see underlining in Fig. 4). Gene conversion and short time intervals between single duplication events could thus explain the overall low support for the branches shown in figure 3C.

As outlined in the Materials and Methods, PAML tends to pinpoint a certain fraction of false positives (Huges and Friedman 2005). Moreover, the uncertainties of present tree reconstruction might have promoted the identification of false positives. For this reason we highlighted in Fig. 4 only those seven codon sites as candidates for positive selection across the paralogue foreground that got highly significant support ($p_{(\omega>1)} > 0.99$) under usage of the intree shown in Fig. 2. Irrespective of this, it is not decisive in the context of the present study whether each pinpointed candidate site is actually positively selected. Instead, it is essential to note that present branch-site analysis suggests positively selected sites solely for partial D3 paralogues of mouse, and not for partial D3 orthologues, even when applying less conservative conditions ($p_{(\omega>1)} > 0.5$). Thus, both pairwise sequence comparisons and branch-site analyses suggest an acceleration of sequence evolution in partial D3 paralogues of mouse, compared to partial D3 orthologues. We consider the reciprocal confirmation of the results as evidence for the validity of our conclusions.

Pattern and Process of Evolution

The presence of positively selected codon sites already points to the functional relevance of the partial D3 paralogues of mouse. Neither frame shift nor nonsense mutations occurred during their evolution (Fig. 4) (see also Gao and Garbers 1998). Furthermore, the absolute conservation of 18 cysteines suggests their involvement in tertiary and/or quarternary structure via disulfide bridges, thus providing additional evidence for functionality of partial D3 paralogues (see Cs in Fig. 4). Nonfunctionalization can thus be ruled out for each of the partial D3 paralogues of mouse. However, this conclusion is hardly surprising if one considers that all 20 partial D3 paralogues of mouse have been sequenced based on mRNA and cDNA, respectively (Gao and Garbers 1998).

Though the final proof for the functional relevance of each predicted motif still has to be provided, the large number of phosphorylation, N-glycosylation,

and myristylation motifs (Fig. 4) makes it appear probable that at least some of them contribute to the modulation of activity and binding properties of the individual D3 paralogues of mouse (see Jeromin et al. 2004; Bruce et al. 2004; Otto et al. 2004). Remarkably, there is not a single pair of partial D3 paralogues with an identical motif pattern (Fig. 4), a finding that implies functional divergence. Beyond this, the presence of positively selected codon sites suggests neofunctionalization (=gain of new functional properties [see Ohta 1988]) and not subfunctionalization (=partitioning of function among degenerated copies [Jensen 1976; Orgel 1977; Force et al. 1999; see also Piatigorski, Wistow 1991; Hughes 1994, 2005]). However, the term ''neofunctionalization'' has to be used with caution here. First, all partial D3 paralogues of mouse are involved in zona pellucida binding and their functional properties can be assumed to differ only gradually. Second, we cannot rule out that the signature of one or more early phases of subfunctionalization has been erased by time.

If one additionally takes into account the present evidence for partial gene conversion (see Fig. 4), the picture becomes even more complex. Though gene conversion can principally contribute to the diversification of paralogues (e.g., Pasquier 2005), we solely found evidence for homogenization of the partial D3 paralogues of mouse by gene conversion (see Fig. 4). Therefore, sequence evolution of partial D3 paralogues of mouse might be better described as a combination of divergent evolution at the codon level and convergent evolution at the level of smaller fragments. However the different forces might have interacted in the case of partial D3 paralogues of mouse zonadhesin, divergent evolution has so far outbalanced the probably equalizing effect of gene conversion. Given the much higher level of pairwise distances (d_n plus d_s) inferred from D0–D4 paralogues compared to D0–D4 orthologues (present study), the same holds true for domains D0–D4. We therefore conlude that divergence represents a general pattern in the evolution of zonadhesin D domains.

The Driving Force

Intragenic divergence as described here for zonadhesin D domains represents a common phenomenon in the evolution of tandem repeats (see, e.g., Muse et al. 1997; Thomas et al. 1997). However, prevailance of divergent evolution is not the only pattern realized. In the case of apolipoprotein(a), for instance, kringle domains evolve concertedly (Hughes 2000). In tenascin-X, not only exons coding for tandem repetitive domains but even neighboring introns are homogenized by concerted evolution (Ikuta et al. 1998; Hughes 1999). An example from sperm-egg interaction is the vitelline egg receptor for the sperm ligand lysin (VERL) of abalones. With 22–28 tandem repeats of 153 amino acids each (Swanson and Vacquier 2002; see Galindo et al. [2003] regarding the evolution of the N-terminus), the structure of VERL is similar to the tandem repetitive 20 partial D3 domains, each 120 amino acids long, in mouse zonadhesin (Gao and Garbers 1998) (Fig. 1). In contrast to zonadhesin D domains, the tandem repeats of VERL evolve concertedly by gene conversion and unequal crossing-over (Swanson and Vacquier 2002). The functional determination as a receptor or ligand might thus determine whether tandem repetitive structures involved in sperm-egg interaction evolve concertedly or divergently.

The concertedly evolving VERL repeats bind the positively evolving sperm ligand lysin. As it seems the binding partner lysin is forced to compensate the changes of VERL by complementary amino acid substitutions (Swanson and Vacquier 2002). Inasmuch, it is female cryptic choice by VERL and thus sexual selection that entails positive selection in lysin. Considering that zonadhesin D domains bind to zona pellucida (Hardy and Garbers 1994, 1995; Gao and Garbers 1998; Hickox et al. 2001; Lea et al. 2001; Bi et al. 2003), it is likely that female cryptic choice also contributes to the evolution of zonadhesin D domains (Swanson et al. 2003; Herlyn and Zischler 2005b; present study). The presence of positively selected codon sites is common in sex-related genes (reviewed by Swanson and Vacquier 2002). However, as far as we know the present findings provide for the first time evidence of an acceleration of sequence evolution in the paralogue-orthologue comparison regarding a protein involved in sperm-egg interaction.

Conclusions

1. A detailed analysis of sequences from pig, rabbit, mouse, and primates reveals higher pairwise distances between D domain paralogues than between D domain orthologues. Moreover, ω and *pnr*/*pnc* are increased in pairwise comparisons of partial D3 paralogues compared to partial D3 orthologues. Given the older phylogenetic origin of partial D3 orthologues, partial D3 paralogues of mouse evolved with a higher substitution rate and, especially, with a higher ratio of nonsynonymous to synonymous substitutions rates. The latter has been confirmed by branch-site analyses.
2. The individuality in primary structure and predicted motif pattern suggests functional differences among partial D3 paralogues of mouse. As we, furthermore, found hints for positive selection, partial D3 paralogues underwent neofunctional-

ization, rather than subfunctionalization. Irrespective of the terminology, all partial D3 repeats seem to be involved in postacrosomal sperm-zona pellucida binding and their functional properties can thus be assumed to differ only slightly. Furthermore, if one considers evidence for homogenization by partial gene conversion events, sequence evolution of partial D3 paralogues of mouse might be better described as a combination of divergent evolution at the codon level and convergent evolution at the level of smaller fragments. Irrespective of these considerations, divergent evolution has so far outbalanced the equalizing effect of gene conversion. A similar pattern can be assumed for the evolution of the complete D domains.

3. Presumably, constantly evolving zona pellucida receptor and thus female cryptic choice contribute to the divergent evolution of zonadhesin D domains. Comparing the present results with literature data suggests that the functional determination as a sperm ligand or egg receptor influences whether tandem repetitive structures involved in sperm-egg interaction evolve divergently or concertedly.

Acknowledgments. Many thanks go to two unknown reviewers whose suggestions improved the quality of the manuscript. The authors acknowledge financial support from the Deutsche Forschungsgemeinschaft (HE 3487/1-1).

References

Aguileta G, Bielawski JP, Yang Z (2004) Gene conversion and functional divergence in the β-globin gene family. J Mol Evol 59:177–189

Bi M, Hickox JR, Winfrey VP, Olson GE, Hardy DM (2003) Processing, localization and binding activity of zonadhesin suggest a function in sperm adhesion to the zona pellucida during exocytosis of the acrosome. Biochem J 375:477–488

Bruce YM, Mikolajczak SA, Yoshida T, Yoshida R, Kelvin DJ, Ochi A (2004) CD28 T cell costimulatory receptor function is negatively regulated by N-linked carbohydrates. Biochem Biophys Res Commun 317:60–67

Eizirik E, Murphy WJ, O'Brian SJ (2001) Molecular dating and biogeography of the early placental mammal radiation. J Hered 92:212–219

Force A, Lynch M, Pickett FB, Amores A, Van YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Galindo BE, Vacquier VD, Swanson WJ (2003) Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci USA 100:4639–4643

Gao Z, Garbers DL (1998) Species diversity in the structure of zonadhesin, a sperm-specific membrane protein containing multiple cell adhesion molecule-like domains. J Biol Chem 273:3415–2421

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hardy DM, Garbers DL (1994) Species-specific binding of sperm proteins to the extracellular matrix (zona pellucida) of the egg. J Biol Chem 269:19000–19004

Hardy DM, Garbers DL (1995) A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand Factor. J Biol Chem 270:26025–26028

Herlyn H, Zischler H (2005a) Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. Mol Phylogenet Evol 37:62–72

Herlyn H, Zischler H (2005b) Sequence evolution, processing, and posttranslational modification of zonadhesin D domains in primates, as inferred from cDNA data. Gene 362:85–97

Hickox JR, Bi M, Hardy DM (2001) Heterogeneous processing and zona pellucida binding activity of pig zonadhesin. J Biol Chem 276:41502–41509

Hofmann K, Bucher P, Falquet L, Bairoch A (1999) The PROSITE database, its status in 1999. Nucleic Acids Res 27: 215–219

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

Hughes AL (1999) Concerted evolution of exons and introns in the MHC-linked tenascin-X gene of mammals. Mol Biol Evol 16:1558–1567

Hughes AL (2000) Modes of evolution in the protease and kringle domains of the plasminogen-prothrombin family. Mol Phylogenet Evol 14:469–478

Hughes AL (2005) Gene duplication and the origin of novel proteins. Proc Natl Acad Sci USA 102:8791–8792

Hughes AL, Friedman R (2005) Variation in the pattern of synonymous and nonsynonymous difference between two fungal genomes. Mol Biol Evol 22:1320–1324

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol 7:515–524

Ikuta T, Sogawa N, Hiroyoshi A, Ikemura T, Matsumoto K (1998) Structural analysis of mouse tenascin-X: evolutionary aspects of reduplication of FNIII repeats in the tenascin gene family. Gene 217:1–13

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

Jeromin A, Muralidhar D, Parameswaran MN, Roder J, Fairwell T, Scarlata S, Dowal L, Mustafi SM, Chary KV, Sharma Y (2004) N-Terminal myristoylation regulates calcium-induced conformational changes in neuronal calcium sensor-1. J Biol Chem 279:27158–27167

Kimura M, Ohta T (1974) On some principles governing molecular evolution. Proc Natl Acad Sci USA 71:2848–2852

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150–163

Lea IA, Sivashanmugam P, O'Rand MG (2001) Zonadhesin: characterization, localization, and zona pellucida binding. Biol Reprod 65:1691–1700

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. Trends Genet 20:544–549

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294: 2348–2351

Muse SV, Clark AG, Thomas GH (1997) Comparisons of the nucleotide substitution process among repetitive segments of the α- and β-spectrin genes. J Mol Evol 44:492–500

Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6:290–300

Ohta T (1988) Evolution by gene duplication and compensatory advantageous mutations. Genetics 120:841–847

Ohta T (2000) Evolution of gene families. Gene 259:45–52

Orgel LE (1977) Gene-duplication and the origin of proteins with new function. J Theor Biol 67:773

Otto VI, Schlürpf T, Folkers G, Cummings RF (2004) Sialylated complex-type N-glycans enhance the signalling activity of soluble intercellular adhesion molecule-1 in mouse. J Biol Chem 279:35201–35209

Pasquier LD (2006) Germline and somatic diversification of immune recognition elements in Metazoa. Immunol Lett 104:2–17

Penny D, Hasegawa M, Waddell PJ, Hendy MD (1999) Mammalian evolution: timing and implications from using the LogDeterminant transform for proteins for differing amino acid composition. Syst Biol 48:76–93

Piatigorski J, Wistow G (1991) The recruitment of crystallins: new functions precede gene duplication. Science 252:1078–1079

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Smith RJ, Cheverud JM (2002) Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's Rule in primates. Int J Primatol 23:1095–1135

Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the cretaceous-tertiary boundary. Proc Natl Acad Sci USA 100:1056–1061

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144

Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20

Thomas GH (1998) Molecular evolution of spectrin repeats. BioEssays 20:600

Thomas GH, Newbern EC, Korte CC, Bales MA, Muse SV, Clark AG, Kiehart DP (1997) Intragenic duplication and divergence in the spectrin superfamily of proteins. Mol Biol Evol 14:1285–1295

Torgerson DG, Singh RS (2004) Rapid evolution through gene duplication and subfunctionalization of the testes-specific α4 proteasome subunits in *Drosophila*. Genetics 168:1421–1432

Van de Peer Y, Taylor JS, Braasch I, Meyer A (2001) The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J Mol Evol 53:436–446

Xia X, Xie Z (2001) DAMBE: data analysis in molecular biology and evolution. J Hered 92:371–373

Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. Mol Phylogenet Evol 26:1–7

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708–3713

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. Proc Natl Acad Sci USA 77:2158–2162