

Sequence evolution, processing, and posttranslational modification of zonadhesin D domains in primates, as inferred from cDNA data

Holger Herlyn*, Hans Zischler

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

Received 27 September 2004; received in revised form 3 May 2005; accepted 2 June 2005

Available online 26 September 2005

Received by W. Makalowski

Abstract

Zonadhesin is a mammalian transmembrane sperm ligand. Precursor zonadhesin essentially consists of MAM (mepripin/A5 antigen/mu receptor tyrosine phosphatase) domains, a mucin-like repeat, and D domains (homologous to von Willebrand D). Recent immunovisualization and binding assays indicate that zonadhesin D domains 1–3 bind postacrosomally to the zona pellucida. This feature has attracted considerable interest in the evolution of zonadhesin and its possible biological and biomedical implications. Previous molecular evolutionary analyses, however, were confined to cDNA sequences of only few distantly related species. Moreover, except for rabbit and pig, little is known about zonadhesin's processing. To delineate the situation in primates including humans, we analyze here the evolution of zonadhesin on the basis of D domain encoding cDNAs of about 4900 base pairs (bp) length from a representative primate sampling (1 Strepsirhini, 3 Cercopithecoidea, 3 Platyrrhini, and human; 7 new sequences) plus GenBank data from mouse, rabbit, and pig. Site-specific (CODEML and HyPhy) analysis indicates positive evolution of zonadhesin. Moreover, moving window analysis (CRANN) points to a positive correlation of sequence evolution and sperm-competition. Significant accumulations of positively selected sites across interspecifically variable motifs (identified by PROSITE) suggest that positive selection promotes differences between species by amino acid exchanges and changes in posttranslational modification. In the case of zonadhesin D domains, positive selection might thus contribute to the species-specific binding of zonadhesin and zona pellucida. A high conservation of processing and dimerization motifs of primate zonadhesin in analogy to pig, on the other hand, illustrates that zonadhesin's backbone needs to meet basic requirements in order to retain function.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Primates; Positive evolution; Moving window analysis; Motif search; Phylogenetic shadowing; Speciation

Abbreviations: bp, base pair(s); cDNA, DNA complementary to RNA; D domains, domains D0–4, domains homologous to von Willebrand D; Δl , log-likelihood difference; EGF, epidermal growth factor; f_b , fraction of the beta distributed site classes in M8; f_0, f_1 , fraction of the sites classes with $\omega \leq 1$ in M3; f_p , fraction of the positively selected site class in M3 and M8; f_s , fraction of the single site class in M0 (always = 1); l , log likelihood; LRT I+II, likelihood ratio tests for rate heterogeneity and positive selection, respectively; M0, M3, M7, M8, evolutionary models; MAM domain(s), mepripin/A5 antigen/mu receptor tyrosine phosphatase domain(s); $\omega = d_n/d_s$, rate ratio of non-synonymous to synonymous substitutions; $\omega_b, \omega_s, \omega_0, \omega_1, \omega_p$, mean ω of site classes with fractions f_b, f_s, f_0, f_1 , and f_p ; p , significance level; $p_{\text{beta}}, q_{\text{beta}}$, beta distribution parameters; PCR, polymerase chain reaction; $p_{(\omega > 1)}$, posterior probability that a certain codon site is under positive selection.

* Corresponding author. Tel.: +49 6131 3923179; fax: +49 6131 3923799.

E-mail address: herlyn@uni-mainz.de (H. Herlyn).

1. Introduction

Zonadhesin is a multidomain transmembrane sperm-ligand of mammals located in the acrosomal membrane. In pig, human, and rabbit, the unprocessed precursor essentially consists of two MAM domains, a mucin-like multiple tandem repeat, five D domains, a transmembrane segment, and a short cytoplasmic tail (Hardy and Garbers, 1995; Gao and Garbers, 1998; Lea et al., 2001; Fig. 1). Mouse precursor zonadhesin differs from this general pattern by the additional presence of one MAM domain and 20 partial D domains (Gao and Garbers, 1998). While the binding partners of zonadhesin MAM domains and mucin-like repeat are still to be determined, binding

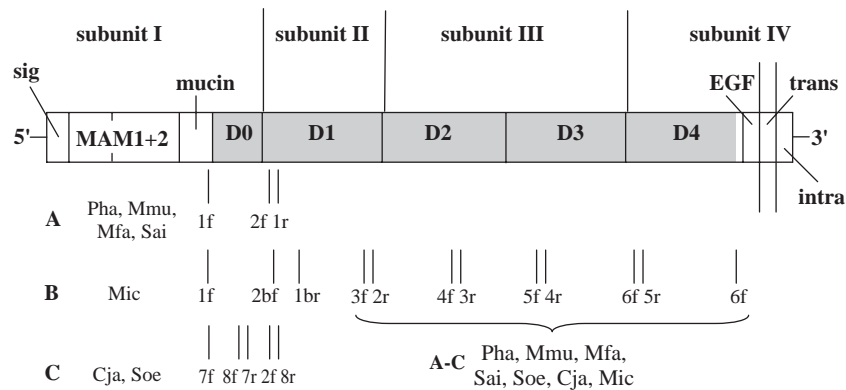


Fig. 1. Schematic domain structure and processing pattern into four subunits in pig and primate zonadhesin (modified after Bi et al., 2003). The mucin-like repeat (mucin) is arbitrarily truncated. The range of the present cDNA dataset is highlighted in gray (about 4900 bp). Below the domain structure, the approximate binding sites of PCR-primers (vertical bars) used to amplify overlapping fragments of (A) *Macaca fascicularis* (Mfa), *Macaca mulatta* (Mmu), *Papio hamadryas* (Pha), and *Saimiri sciureus* (Sai), (B) *Microcebus murinus* (Mic), and (C) *Callithrix jacchus* (Cja) and *Saguinus oedipus* (Soe) are shown. Codes of primers refer to Table 1. From primer 3f on, an identical set of primers was used for all species considered (A–C). Symbols are as follows: D0–D4, domains D0–4; EGF, EGF-like domain; intra, intracellular segment; MAM1+2, MAM domains 1 and 2; muc, mucin-like repeat (truncated); sig, signal peptide; trans, transmembrane segment. The figure is not drawn to scale.

assays and immunovisualization revealed that domains D1–3 bind post-acrosomally and species-specifically to the zona pellucida (Hardy and Garbers, 1995; Hickox et al., 2001; Lea et al., 2001; Bi et al., 2003). In analogy to the sperm ligand lysin of marine gastropods of the genera *Haliotis* and *Tegula*, the binding of zonadhesin might create a hole in the zona pellucida, thus enabling the spermatozoon to reach the egg cell membrane (see Olson et al., 2004).

Like other sex-related proteins of a broad range of taxa encompassing for instance the vagile algae *Chlamydomonas*, marine molluscs, sea urchins as well as Mammalia (reviewed in Swanson and Vacquier, 2002; see also Kamei and Glabe, 2003; Torgerson and Singh, 2003; Kouprina et al., 2004; Dorus et al., 2004), zonadhesin has been shown to be under positive selection (Swanson et al., 2003; for MAM domains see also Herlyn and Zischler, in press). However, possible correlations of amino acid replacements and posttranslational modification had only been studied for MAM domains so far (Herlyn and Zischler, in press). Moreover, previous analysis of the processing of the precursor zonadhesin focused on pig (Bi et al., 2003) and rabbit (Lea et al., 2001). Thus, still little is known about sequence evolution, processing and posttranslational modification of zonadhesin within humans and nonhuman primates. Especially, additional data on proteins involved in human reproduction appear desirable, as they might open up new perspectives in contraception and the treatment of sub- and infertility.

In pig and rabbit, posttranslational proteolytic processing of precursor zonadhesin into four subunits has been shown, with the cleavage sites being located in the D domains (Lea et al., 2001; Bi et al., 2003). Given this central importance of zonadhesin D domains for processing, we here report the sequence evolution, processing, and posttranslational modification of zonadhesin on the basis of D domain encoding cDNAs of about 4900 bp

length. As in evolutionary studies on other proteins with biomedical relevance to humans (see e.g., Sumiyama et al., 2002; Filip and Mundy, 2004; Wang and Su, 2004), the present study focuses on a representative primate sampling comprising new sequences from 1 Strepsirhini, 3 Platyrrhini, and 3 Cercopithecoidea, plus the human orthologue from GenBank. To assess sequence evolution on the level of single codon sites, we use CODEML (implemented in PAML version 3.14; see Yang, 1997; Nielsen and Yang, 1998; Yang et al., 2000) and HyPhy (Kosakovsky Pond et al., 2004) estimates of the rate ratio of non-synonymous to synonymous substitutions ($d_n/d_s = \omega$) as a measure. Thereby, ω values of 1.0 indicate neutral evolution, while $\omega < 1.0$ indicates negative or purifying selection, and $\omega > 1.0$ points at positive or adaptive evolution. To compare sequence evolution between species, moving window analysis (CRANN; Creevey and McInerney, 2002) is performed, using the primate ancestor sequence generated by BASEML (PAML package) as common reference. Subsequent predictions of processing and posttranslational modification are based on motif search carried out per eye and using PROSITE (Hofmann et al., 1999). The functional relevance of indicated motifs is then rated, based on three criteria: 1) High phylogenetic conservation of motifs (principle of “phylogenetic shadowing”; see e.g., Boffelli et al., 2003), 2) convergent evolution of motifs, and 3) overlaps of motifs with positively selected sites. Like in the case of semenogelin II (*SEM2*; Dorus et al., 2004), zonadhesin data suggest a positive correlation of sperm-competition and positive selection. Moreover, there is evidence that positive evolution on the level of single amino acid sites promotes changes in posttranslational modification. On the other hand, processing of the primate precursor zonadhesin is apparently analogous to pig (Bi et al., 2003), thus showing high conservation.

2. Materials and methods

2.1. RNA- and DNA-isolation, sequencing, and aligning

Testes were taken 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*), rhesus monkey (*Macaca mulatta*), and crab-eating macaque (*Macaca fascicularis*) during diagnostic necropsies, and stored at -80°C . Total RNA was isolated using TRI REAGENT (Molecular Research Center, Inc.). First-strand cDNA synthesis was carried out, using the SuperScript II and ThermoScript RT (both GibcoBRL) protocols and random primers. Subsequently, overlapping fragments of zonadhesin domains D0-4 were amplified by PCR using AmpliTaq (Applied Biosystems). Degenerated PCR-primers were designed from fragments conserved between pig and human zonadhesin from GenBank (for accession nos. see below). Based on the successively increasing amount of sequence data, additional primers were designed for *C. jacchus*, *S. oedipus*, and *M. murinus* (Table 1; Fig. 1). Electrophoretically separated fragments were isolated, ligated in pGEM T-vector (Promega), electroporated into *Escherichia coli* (TOP 10, Invitrogen), and sequenced in both directions on a LiCor system, using a Thermo Sequenase Cycle Sequencing Kit (Amersham). Consensus-sequences were deduced from forward and reverse sequences of at least two clones each (for sequencing primers see Table 1). Concatenation products of about 4900 bp coding for the complete D0-3 domains plus the nearly complete D4 domain

are available from GenBank (accession nos. AY428845, AY428847, AY428849, AY428851, AY428853, AY428855, AY428857). The concatenated D0-4 encoding sequences were visually aligned with homologues from pig (*Sus scrofa*), house mouse (*Mus musculus*), European rabbit (*Oryctolagus cuniculus*) and human (*Homo sapiens*) from GenBank (accession nos. U40024, NM_011741, AF244982, AF332975) using nucleotide and amino acid sequences.

2.2. Data analysis

For site-specific sequence analysis of the dataset, we used the maximum likelihood approach implemented in CODEML (PAML package version 3.14; Yang, 1997; Nielsen and Yang, 1998; Yang et al., 2000). Codon frequencies were estimated from the dataset using the F3x4 option. Other settings were as default. Thus, the threshold for minimal support for positive selection from posterior probability ($P_{(\omega>1)}$) was 0.5. CODEML was run on the basis of an unrooted intree representing the widely accepted phylogeny within primates (Fig. 2; Smith and Cheverud, 2002). To test for rate heterogeneity, we ran the one ratio model M0 and the discrete model M3 ($K=3$). Model M0 assumes one site class for all codon sites. M3 ($K=3$) distinguishes three ω site classes, one of which is allowed to exceed $\omega=1$. To assess the significance of the findings, a likelihood ratio test (LRT I) for rate heterogeneity was carried out. Therefore, twice the log likelihood difference ($2\Delta l$) of models M0 and M3 was compared with critical values from a chi-square distribution (cv in the

Table 1
Primers used to amplify overlapping fragments coding for zonadhesin domains D0-4

Species (three-letter code)	Forward (f) reverse (r)	Primer pairs	Fragment length [bp]
Pha, Mmu, Mfa	1f	5'-CCAAGTAC(C/T)(C/T)(C/T)(A/G)(C/T)GACCA(G/C)TG-3'	~ 600
Sai	1r	5'-TC(A/G)GGCAG(A/G)GTCACGTAGAC-3'	
Pha, Mmu, Mfa	2f	5'-GCCCC(A/G)CCCTGTGGCAACT(C/T)(A/G)AC-3'	~ 1020
Sai, Cja, Soe	2r	5'-CCAG(A/G)GC(A/T)(G/C)(A/G)CAGCGA(A/G)TTC-3'	
Pha, Mmu, Mfa	3f	5'-AAGGTAGGGGAGCGGTGGT(A/T)C-3'	~ 980
Sai, Cja, Soe, Mic	3r	5'-GCAC(C/T)TCA(G/T)AGGGCAGAAG-3'	
	4f	5'-CCTGTGCCGCTCCCTGCAGG-3'	~ 890
	4r	5'-TTGAAGTTC(C/C)(A/G)CACA(C/T)GCCGCAG-3'	
	5f	5'-GAG(A/G)TTGAAATCCCCA(A/C)A(A/G)C-3'	~ 1010
	5r	5'-CA(CT)CTGTGCGTCC(CT)TCCAC-3'	
	6f	5'-ATTTGA(C/T)GGC(C/T)TCAGCTACC-3'	~ 1000
	6r	5'-AGGCTCTGGG(C/T)(A/G)CCTACTGTAG-3'	
Cja, Soe	7f	5'-CGACCAGTGTGATCCTG-3'	~ 330
	7r	5'-AGAGA(C/T)CTGGCACTCGAC-3'	
	8f	5'-TTCAGCCCCAACTGCACAGAAC-3'	~ 340
	8r	5'-TTGGGCAGGGTACCGTAGAC-3'	
Mic	1f	5'-CCAAGTAC(C/T)(C/T)(C/T)(A/G)(C/T)GACCA(G/C)TG-3'	~ 750
	1br	5'-TCACCATCCCATCTCAC-3'	
	2bf	5'-A(A/G)GG(C/T)GTGTCCCTGCCT(A/G)A(A/G)-3'	~ 950
	2r	5'-CCAG(A/G)GC(A/T)(G/C)(A/G)CAGCGA(A/G)TTC-3'	
sequencing primer	forward	5'-AGG GTT TTC CCA GTC ACG ACG TT-3'	PCR-product
	reverse	5'-GAG CGG ATA ACA ATT TCA CAC AGG-3'	+ ~ 260

For the binding sites of the PCR primers along zonadhesin domains D0-4; see Fig. 1. Three-letter code: Cja, *Callithrix jacchus*; Mfa, *Macaca fascicularis*; Mic, *Microcebus murinus*; Mmu, *Macaca mulatta*; Pha, *Papio hamadryas*; Sai, *Saimiri sciureus*; Soe, *Saguinus oedipus*.

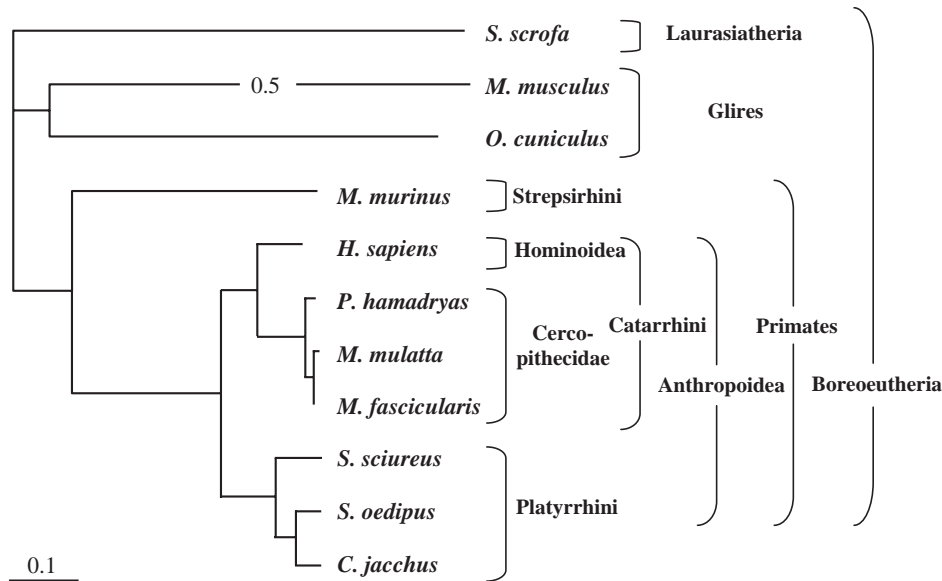


Fig. 2. Phylogram based on cDNA data from zonadhesin D domains (modified after Murphy et al., 2001; Smith and Cheverud, 2002). As the phylogenetic relations of Primates, Glires, and Laurasiatheria are still disputed, the base of the tree is unrooted. The long terminal branch leading to the mouse is truncated (the excised part represents 0.5 substitutions per nucleotide position). Scale bar: 0.1 substitutions per nucleotide position.

following) with the degrees of freedom ($d.f.$) equal to the difference in the number of freely estimated parameters between the two models, i.e. $d.f.=4$. Fraction estimates of the ω site classes (M3: f_0, f_1 , fractions of the two site classes with $\omega \leq 1$; f_p , fraction of the positively selected site class) and the ω values (M0: ω_s , mean ω of the single site class assumed; M3: ω_0, ω_1 , mean ω of the site classes with $\omega \leq 1$, ω_p , mean ω of the positively selected site class) are shown in Table 2.

Further analysis employed models M7 and M8. The null model M7 assumes a beta distribution of 10 discrete site classes with $\omega \leq 1$ but does not allow for positive selection. The beta and ω model M8 was used to identify sites under positive selection. Model M8 again assumes a beta distribution of 10 site classes with $\omega \leq 1$. Beyond this, however, M8 allows for a positively selected extra site class with $\omega > 1$. As CODEML M8 is known to be prone to local optima, it was run twice with different initial ω values (0.4 and 3.4). Both provided identical results. To assess the significance of the findings, a likelihood ratio test for positive selection (LRT II) was applied. Therefore, $2\Delta l$ of the beta& ω model M8 and the beta null model M7 was compared with cv with $d.f.=2$. For parameter estimates for the beta distribution (M7, M8: p_{beta} and q_{beta}), the proportion estimates of site classes in M8 (f_b , sum of all beta distributed site classes; f_p , as above), as well as the mean ω values for the positively selected extra site class (ω_p) see again Table 2. CODEML M8 implemented in PAML version 3.14 identifies candidates for positive selection by two separate empirical approaches, the Bayes empirical Bayes (BEB) approach that takes sampling errors into account, and the naive empirical Bayes (NEB) approach that ignores possible sampling errors. As the BEB approach represents a further develop-

ment of the NEB, the present study is based on the results of the BEB approach.

Subsequently, only those sites were regarded as candidates for positive selection that were identified by both the discrete model M3 and the beta& ω model M8. As an additional evidence for positive selection we took into account hints for convergent evolution at the amino

Table 2
Parameter estimates and test statistics of CODEML analysis

	Parameters
One ratio model M0	$l = -22,959.081$ $\omega_s = 0.250$ ($f_s = 1$)
Discrete model M3 ($K=3$)	$l = -22,581.896$ $f_0 = 0.382, f_1 = 0.485, f_p = 0.133$ $\omega_0 = 0.023, \omega_1 = 0.330, \omega_p = 1.128$
LRT I	$2\Delta l = 754.370, cv = 18.467, p < 0.001$
Beta null model M7	$l = -22,588.895$ $p_{\text{beta}} = 0.437, q_{\text{beta}} = 0.999$
Beta& ω model M8	$l = -22,583.253$ $p_{\text{beta}} = 0.465, q_{\text{beta}} = 1.112$ $f_b = 0.990$ ($f_p = 0.010$) $\omega_p = 2.749$
LRT II	$2\Delta l = 11.284, cv = 9.210, p < 0.01$

cv , critical value from chi-square distribution, given 2 degrees of freedom; $2\Delta l$, twofold difference between the l values of M7 and M8; f_b , summarized fraction of the beta distributed site classes in M8; f_0, f_1 , fraction of the two sites classes with $\omega \leq 1$ in M3; f_p , fraction of the positively selected site class in M3 and M8; f_s , fraction of the single site class in M0 (always=1); l , log likelihood; LRT I+II, likelihood ratio tests for rate heterogeneity and positive selection, respectively; M0, one ratio model M0; M3 discrete model M3 (with three site classes or $K=3$); M7, beta model M7 (null model); M8, beta& ω model M8 (alternative model); $\omega = d_n/d_s$ =rate ratio of non-synonymous to synonymous substitutions; $\omega_b, \omega_s, \omega_0, \omega_1, \omega_p$, mean ω of site classes with fractions f_b, f_s, f_0, f_1, f_p (see there); p , significance level; $p_{\text{beta}}, q_{\text{beta}}$, estimates of the beta distribution parameters p and q .

acid level. Moreover, codon sites with increased support from M8 and significant support from M3 ($p_{(\omega>1)} > 0.95$) were regarded as very likely candidates for positive selection. To assess the robustness of CODEML results, we performed an analogous analysis using HyPhy (Kosakovsky Pond et al., 2004). To run M0, M3, M7, and M8 under HyPhy, the codon frequency matrix $GY3 \times 4$ was chosen. Other settings (including initial ω values) were as default. Both CODEML and HyPhy employ maximum likelihood to infer rate heterogeneity and positive selection. Beyond this, however, HyPhy includes among others generalizations of the parsimony based approach of Suzuki and Gojobori (1999).

Moving window analysis was performed using CRANN (Creevey and McInerney, 2002) and setting window size and shift size at 60 and 30 codons, respectively. A window size of 60 codons was chosen, because this interval allowed for a high resolution without providing zero d_n and/or d_s values. Moving window analysis was performed using the method of Li (1993). To compare sequence evolution between primate species, moving window analysis was based on two sequence alignments, each comprising one primate representative plus the primate ancestor. The primate ancestor sequence was generated before using BASEML (implemented in the PAML package version 3.14). Except for the treatment of ambiguous data (no removal), the settings were set as default.

Motif search was performed using the PROSITE database. PROSITE identifies protein families and motifs by weighted comparison of DNA sequences with profiled database entries (Hofmann et al., 1999). Additional motifs (CXXC, DP) were identified by eye. Based on the present dataset and using unweighted parsimony criteria (i.e. loss and gain of motifs count equally), we plotted the probable emergence of phosphorylation, *N*-myristylation, *N*-glycosylation, and amidation motifs on trees depicting the phylogeny within the present sampling. To simplify matters, the motifs are termed according to their first codon position in the following, and not according to their entire range (e.g., *N*-myristylation motif 31, instead of *N*-myristylation motif 31–36). The functional relevance of predicted motifs was assessed adopting the principle of “phylogenetic shadowing” (see e.g., Boffelli et al., 2003) on functional motifs. According to this, the probability of actual functionality increases with increasing phylogenetic conservation. Moreover, convergent evolution of a motif and overlaps of motifs with positively selected sites was taken as evidence for functional relevance of a motif. Finally, a one-tailed Fisher’s exact test for independency was carried out to assess whether the number of candidate sites for positive selection across variable posttranslational motifs reflects the number of candidate sites for positive selection across the entire alignment. Calculations were done, using a web-based server (<http://www.physics.csbsju.edu/stats/fisher.form.html>).

3. Results and discussion

3.1. Identification of positively selected sites

Based on the present dataset which comprises orthologues of three non-primates (pig, rabbit, and mouse), the mouse lemur *M. murinus*, 3 Platyrrhini, 3 Cercopithecidae, and human, CODEML null models M0 and M7 represent the sequence evolution of domains D0–4 worse than the alternative models M3 and M8 (see l values in Table 2). LRT’s I and II indicate rate heterogeneity and positive selection (significance levels = $p < 0.001$ and $p < 0.01$, respectively; Table 2). This confirms for primate D domains what had previously been stated for zonadhesin in general (Swanson et al., 2003). CODEML M8 estimates a proportion of 1% of all codon sites to be positively selected ($\omega_p = 2.749$). CODEML M3, in turn, assumes a higher proportion of positively evolving sites (13.3%): However, many of the codon sites pinpointed by M3 are under weak positive selection ($\omega_p = 1.128$; Table 2). Consequently, results of CODEML M8 were taken as reference to minimize the number of false positives.

The BEB approach implemented in CODEML M8 reveals a set of 29 codon sites as candidates for positive selection that are all confirmed by CODEML M3 (see asterisks in Fig. 3). For these 29 codon sites, the support from posterior probability is generally lower from model M8 ($0.5 < p_{(\omega>1)} \leq 0.92$) than from model M3 ($p_{(\omega>1)} \geq 0.9$). About 50% of the 29 sites coincide with evidence for convergent evolution at the amino acid level (see shaded one-letter codes in Fig. 3). At position 1154, for example, an asparagin-coding triplet apparently evolved independently along the lineages leading to mouse (GAT) and *S. sciureus* (GAC; not shown). Moreover, 19 of the 29 candidate sites for positive selection get significant support from M3 ($p_{(\omega>1)} \geq 0.95$ up to $p_{(\omega>1)} > 0.99$), and, thus, can be regarded as very likely candidate sites (see shaded asterisks in Fig. 3). Whether referring to the total set of 29 candidate sites or to the smaller set of 19 very likely candidates for positive selection, a high proportion of the sites (76% and 68%, respectively) coincides with changes in predicted posttranslational modification. Subsequent HyPhy analysis confirms rate heterogeneity and positive evolution for zonadhesin D domains. Moreover, HyPhy analysis provides nearly the same support for the 29 candidates and the 19 very likely candidates for positive selection. Since CODEML and HyPhy employ different algorithms, we take the wide consistence of the data as evidence for the robustness of the present findings.

In a previous approach on zonadhesin’s evolution, 14 sites of domains D0–4 (all with $p_{(\omega>1)} < 0.9$) had been suggested to be under positive evolution using CODEML M8 (Swanson et al., 2003; see rhombs in Fig. 3). As can be seen in Fig. 3, nine of the previously pinpointed sites are confirmed by the present analysis. Probably, the differences in number and position of codon sites

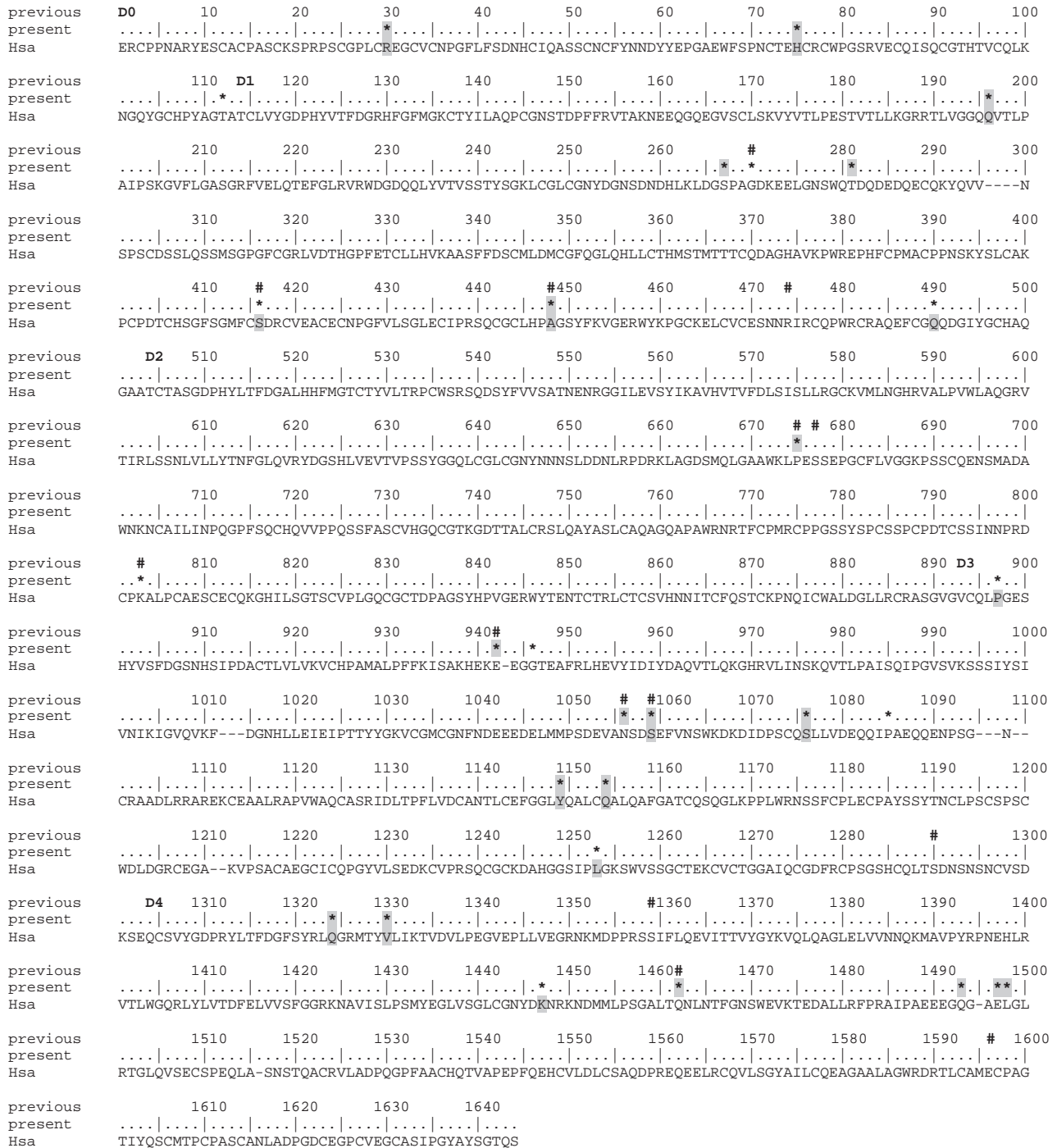


Fig. 3. One-letter code of the amino acid sequence of human zonadhesin D domains. Hyphens indicate gaps in the alignment present analysis is based on. “D0–4” above the ruler indicates the beginning of the respective domains. Asterisks on the ruler highlight candidate sites for positive selection, as inferred by the present analysis from cDNA data of eight primates, rabbit, mouse, and pig ($p_{(\omega > 1)} > 0.5$ from CODEML M3 and M8). Shaded asterisks indicate very likely candidates for positive selection ($p_{(\omega > 1)} \geq 0.95$ from CODEML M3). Shaded one-letter codes point to candidate sites that coincide with evidence for convergent evolution on the amino acid level. Rhombs characterize sites previously suggested to be under positive selection (Swanson et al., 2003).

suggested to be under positive selection are partly a consequence of different samplings. While the previous approach was based on orthologues from pig, rabbit, mouse, and human (Swanson et al., 2003), we here consider orthologues from 11 species, including 8 primates. Moreover, observed differences might be due to the

fact that the previous analysis was confined to the NEB approach of previous CODEML versions that ignores possible sampling errors. In contrast, due to the meanwhile advanced version of CODEML (implemented in PAML version 3.14), the present study uses the results of the BEB approach taking into account possible sampling errors.

Irrespective of the differences in detail, both studies indicate that some of the codon sites coding for zonadhesin D domains are positively selected. Similarly, positive selection has been shown for zonadhesin MAM domains (Swanson et al., 2003; Herlyn and Zischler, in press). Except for zonadhesin, accelerated and/or positive evolution has been shown for a series of proteins involved in mating (e.g., *Er1* in *Euplotes*), sex determination (e.g., *mid* in *Chlamydomonas*, *SRY* in mammals), fertilization (e.g., sperm-ligands bindin and lysin in sea urchins and in abalones, respectively; *SEM2* in primates), and spermatogenesis (e.g., protamines and *SPANX* in primates), thus covering representatives of many taxa spanning from algae via marine invertebrates to mammals (reviewed in Swanson and Vacquier, 2002; see also Wang et al., 2002; Kamei and Glabe, 2003; Swanson et al., 2003; Torgerson and Singh, 2003; Dorus et al., 2004; Kouprina et al., 2004). Inasmuch, the evolution of zonadhesin reflects the general pattern that sex-related proteins often evolve rapidly and positively.

3.2. Moving window analysis

Using the reconstructed primate ancestor as a reference, moving window analysis of primate zonadhesin D domains reflects great differences among species (Fig. 4). In contrast to, for instance, *SEMG2*, a main structural component of semen coagulum (Dorus et al., 2004), there are no obviously corresponding patterns of sequence evolution across species. For example, though the *M. murinus* curve shows a tendency for maxima at the 5' end of domains D1–4, the peaks in *S. sciureus* and *H. sapiens* (and the other Anthropeidea species not shown in Fig. 4) are more

irregularly distributed. Thus, no congruent differential selection signature of single domains can be drawn from present moving window analysis. However, the outlined high degree of interspecific differences in the selection patterns across D domains corresponds well with the lineage-specific emergence of posttranslationally relevant motifs outlined in Section 3.3, stressing the species-specific evolution of zonadhesin D domains. Additional differences occur between taxa regarding the general ω level of the peaks. While the *M. murinus* peaks reach rather high ω values (up to >6), the peaks of the Anthropeidea representatives reach maximal ω values of circa 2 (Platyrrhini) and 1 (Catarrhini). As lemurs in general, *M. murinus* is known for a high degree of sperm competition due to female promiscuity (Fietz, 1999), whereas the degree of sperm-competition is overall lower in Platyrrhini and Catarrhini. Seeing that, the present moving window analysis apparently reflects a positive correlation of positive selection and the degree of sperm-competition due to female promiscuity. Though to a lower extent than in lemurs, female promiscuity occurs also in platyrrhine and catarrhine species (e.g., Digby, 1999; Bercovitch et al., 2002). This might explain, why peaks exceeding $\omega=1$ also appear in Catarrhini species (Fig. 4).

A positive correlation of the degrees of sperm-competition and positive selection had first been postulated from evolutionary analysis of protamine data (Wyckoff et al., 2000), and was recently substantiated on the basis of *SEMG2* data (Dorus et al., 2004). Based on sequences of Hominidae (Wyckoff et al., 2000) and Anthropeidea (Dorus et al., 2004), respectively, both studies especially emphasized an enhancement of positive evolution on the lineage to

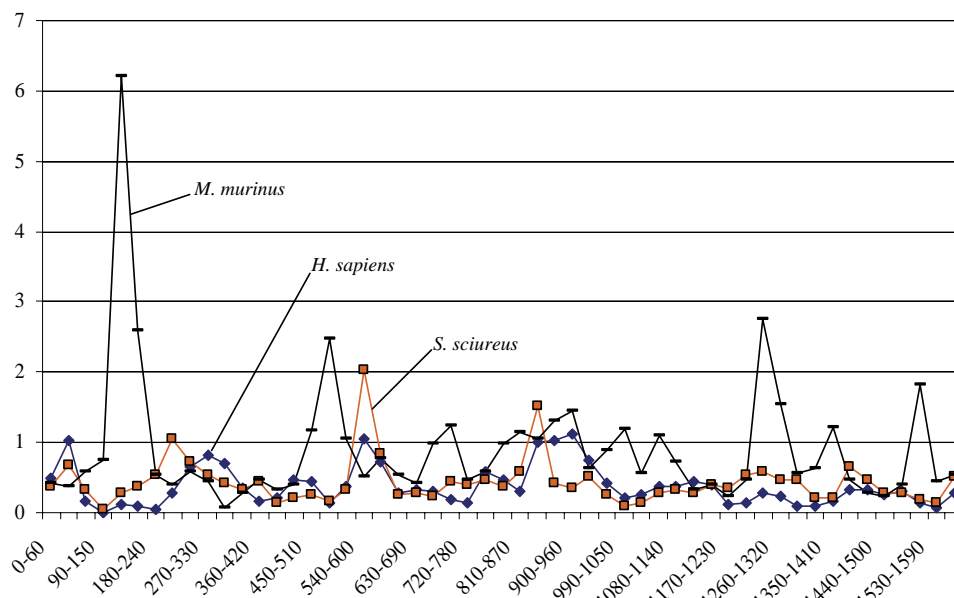


Fig. 4. Moving window analysis of primate cDNAs coding for zonadhesin D domains taking the reconstructed primate ancestor sequence as reference (abscissa: ω ; ordinate: successive windows). Window size and shift size are 60 and 30 codon positions, respectively. For a better overview only three of eight pairwise comparisons are shown (see Section 3.2.).

chimpanzees. The present findings thus extend the previous conclusions to the even more sperm-competitive lemurs. Inasmuch, positive correlation of positive evolution and female promiscuity seems to represent a general principle in primate, maybe even metazoan evolution.

3.3. Posttranslational modification and positive selection

Figs. 5 and 6 illustrate that the emergence of new posttranslational motifs (all numbers with positive signs in Figs. 5 and 6) clearly exceeds the number of subsequent losses (all numbers with negative signs in Figs. 5 and 6), regardless of surveying the evolution of *N*-myristylation, *N*-

glycosylation, amidation, or phosphorylation motifs across zonadhesin D domains. The long branches leading to pig, mouse, rabbit, and mouse lemur naturally accumulate many changes in predicted motif pattern, as they represent the entire lineages leading from the recent species to the stem species of Laurasiatheria, Rodentia, Lagomorpha, and Strepsirhini, respectively (Murphy et al., 2001; Figs. 5 and 6). Irrespective of these considerations, all supra-specific taxa and most of the species are characterized by a unique set of predicted motifs. As the two macaque species fully agree in the pattern of predicted posttranslational modification (Figs. 5 and 6), they represent the only exception from this general rule.

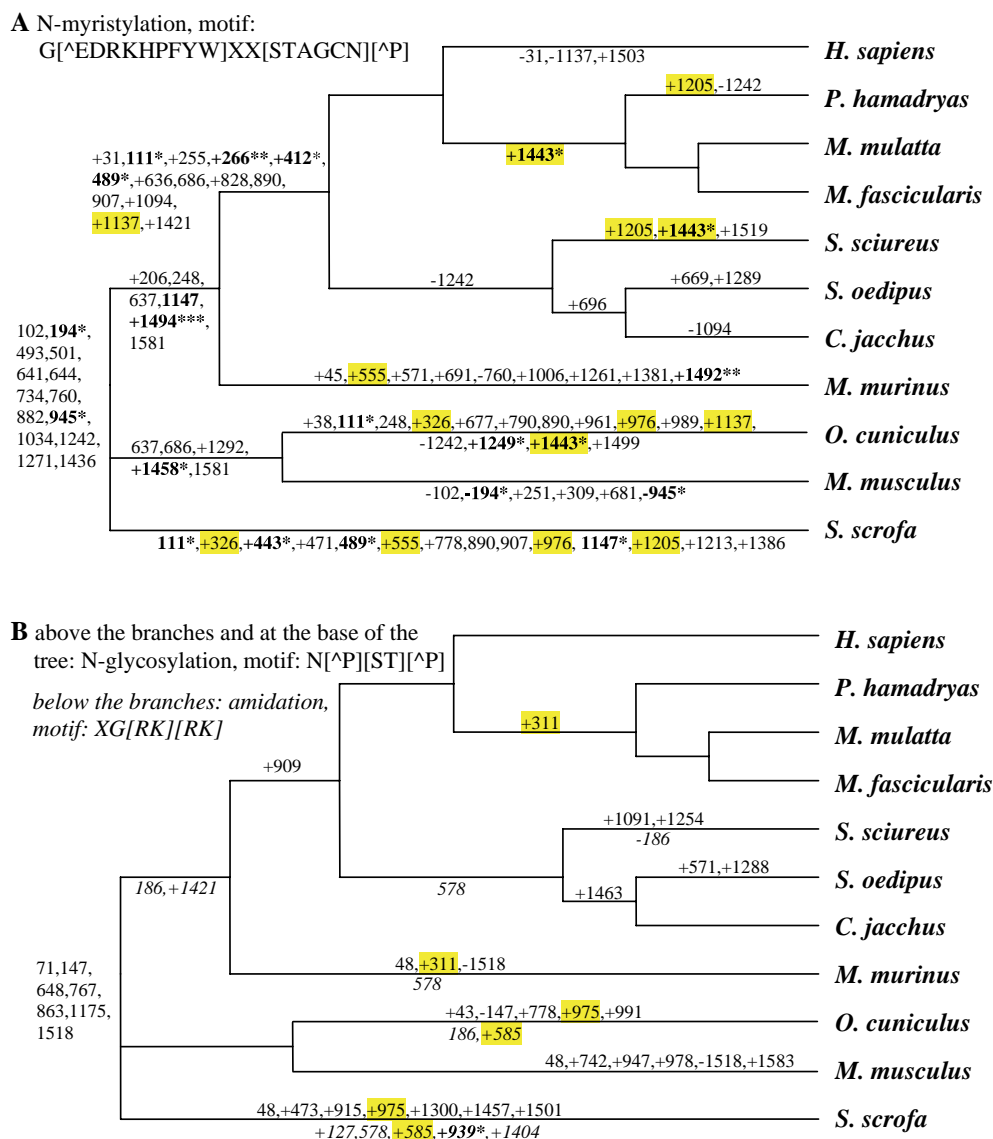


Fig. 5. Cladograms depicting the probable emergence (positive numbers) and loss (negative numbers) of (A) N-myristylation and (B) N-glycosylation and amidation motifs. At the tree basis and in the case of ambiguous character distributions across the sample, the simple presence of a motif is indicated (numbers without sign). The data are drawn from unweighted presence/absence analysis of motifs predicted on the basis of cDNA sequences. Shades highlight convergent changes in the predicted motif pattern. Bold numbers indicate changes in motif pattern due to overlaps of the corresponding motif with positively selected sites. Numbers of asterisks represent the number of candidate sites for positive selection involved. Note: All numbers refer to the first site of the respective motifs.

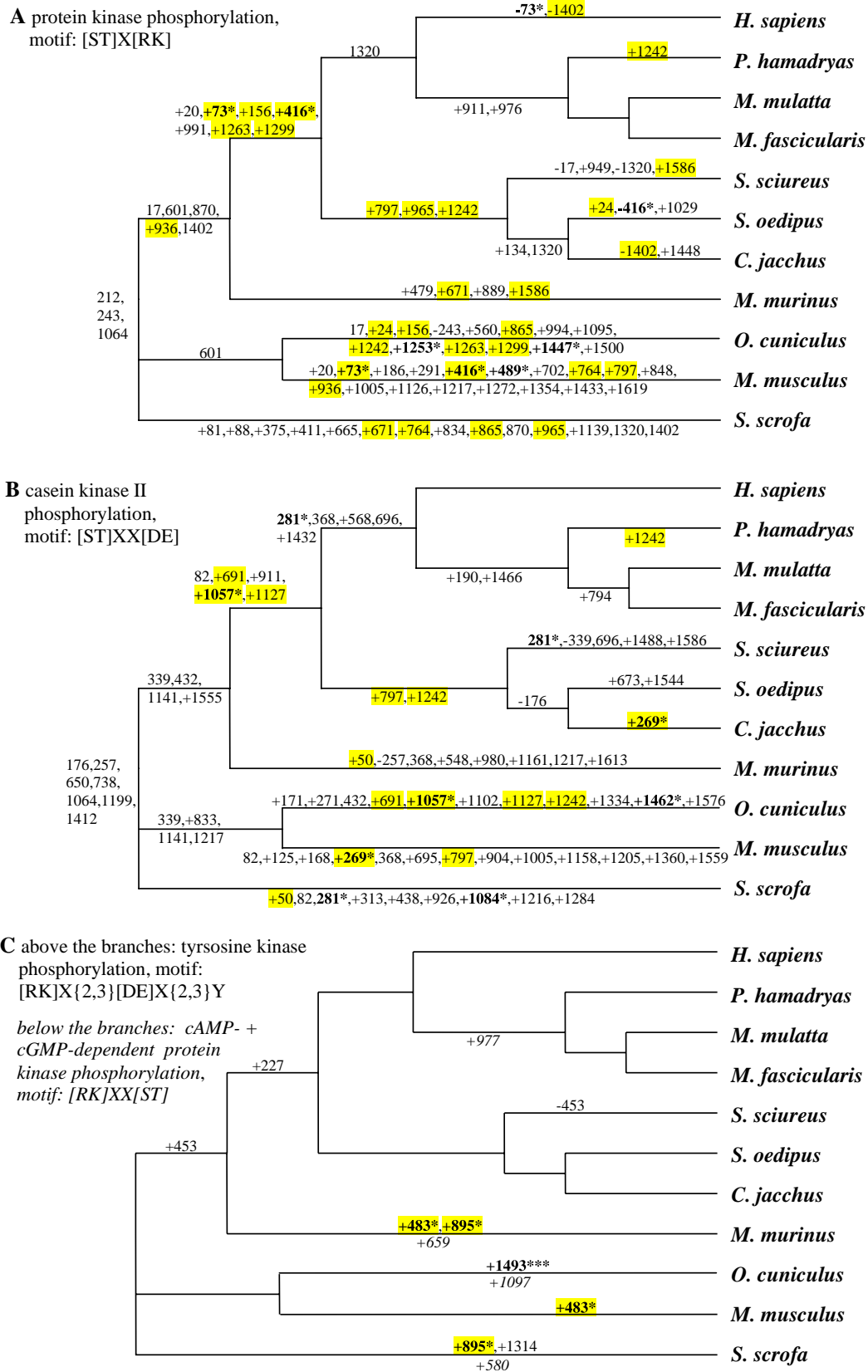


Fig. 6. Cladograms illustrating the probable emergence (positive numbers), loss (negative numbers), and simple occurrence (numbers without sign) of (A) protein kinase phosphorylation motifs, (B) casein kinase II phosphorylation motifs, and (C) tyrosine kinase phosphorylation, cAMP-, and cGMP-dependent protein kinase phosphorylation motifs. For further explanations, see legend of Fig. 5.

Underlying the principle of phylogenetic shadowing (Boffelli et al., 2003), widely conserved motifs that were realized already at the base of the tree (e.g., *N*-glycosylation motifs 71, 147, 648, etc. in Fig. 5B; see Figs. 5 and 6) are essential for the function of zonadhesin in all Boreoeutheria (Murphy et al., 2001). In analogy, motifs established or at least present (in the case of unclear evolutionary history) along the stem lineages of Glires, Anthropeidea, and primates in general can be regarded as essential for zonadhesin's function in the respective taxa (Figs. 5 and 6). As a severalfold independent evolution of motifs across homologue positions appears improbable without adaptive value (e.g., Briscoe, 2002), convergent evolution of posttranslational motifs also points to functional relevance. In the case of zonadhesin D domains, many of the posttranslational motifs evolved convergently along at least two lineages (see shaded numbers in Figs. 5 and 6). For example, the *N*-glycosylation site 311 most probably evolved in *M. murinus* and in the stem lineage of Cercopitheidae (see shaded +311 in Fig. 5B). According to unweighted presence/absence analysis, even threefold convergent events can be assumed. *N*-myristylation site 1443, for instance, most probably emerged in rabbit, *S. sciureus*, and in the stem lineage of Cercopitheidae (see shaded +1443 in Fig. 5A). Moreover, overlapping motifs have evolved. For example, a threefold independent evolution of protein kinase C and casein kinase II phosphorylation motifs can be assumed for position 1242 along the lineages leading to *P. hamadryas*, *Platyrrhini*, and rabbit (see shaded +1242 in Fig. 6A, B). Such overlaps of functionally corresponding motifs might point to redundancy due to special relevance of the respective modification, i.e. phosphorylation in the case of amino acid position 1242.

In general, positive evolution is thought to be of adaptive value (see, e.g., lysozyme and RNase evolution in the stem lineage of colobines; Messier and Stewart, 1997; Zhang, 2003). Thus, also overlaps of positively selected sites with posttranslational motifs can be regarded as hints for the functional relevance of the respective motifs. Based on the present dataset, 22 of the 29 candidate sites for positive selection (=76%; see Section 3.1.) overlap with at least one change in posttranslational modification (see bold numbers in Figs. 5 and 6). On the other hand, only 722 of the 1644 amino acid positions (=44%) of the present dataset overlap with posttranslational motifs that are variable across the sampling. Fisher's exact test rejects the null hypothesis of equal rates of positively selected codon sites across variable motifs (22/722) and background (7/922) with significance ($p=0.001$). Thus, the observed accumulation of candidate sites for positive selection with changes in posttranslational modification represents a significant deviation from the null hypothesis. Mostly, one positively selected site is associated with the presence or emergence of a new motif (see bold numbers with one asterisk in Figs. 5 and 6). However, also two (e.g., *N*-myristylation motif 266 along the stem lineage of Anthropeidea; see bold +266** in Fig. 5A) and three positively selected sites (e.g., tyrosin kinase phosphoryla-

tion site 1493 along the branch leading to rabbit; see bold +1493*** in Fig. 6C) can coincide with the emergence of one posttranslational motif. Inasmuch, positive selection seems not only to promote interspecific differences by exchanges of single amino acids, but also by changes in posttranslational modification. At the same time, it has to be considered that solely the N-terminus of proteins can undergo myristylation. Functionality of internal N-myristylation motifs as predicted here would thus require preceding cleavage. Presuming adequate cleavage, the resulting decomposition products might then be myristylated at the N-terminus.

3.4. Inference of processing and dimerization pattern

Not all amino acid exchanges lead to changes in motif pattern. For example, three epidermal growth factor (EGF) like motifs are conserved within primates, pig, and at least one of the two Glires representatives mouse and rabbit, although evolution is marked by several, partly convergent amino acid replacements (Fig. 7A, B, and D). As even more conserved within primates can be regarded an asparagin-prolin doublet (DP) 87–88 base pairs (bp) downstream of each EGF-like motif (Fig. 7A, B, and D). On the other hand, the corresponding EGF-like motif in domain D3 is present solely in Cercopitheidae and rabbit, and none of the sequences exhibits a DP doublet about 90 bp downstream (Fig. 7C; see also Fig. 3). As EGF-like domains are suggested to facilitate protease binding (see Gao and Garbers, 1998), and DP bonds represent possible hydrolyzation sites of precursor zonadhesin (Bi et al., 2003), the high conservation of EGF-like motifs and DP doublets along domains D1, D2, and D4 points to an analogous processing of primate and pig zonadhesin into four subunits (see Hardy and Garbers, 1995; Hickox et al., 2001; Bi et al., 2003). To simplify matters, an abstract terminology is subsequently employed, which describes procession products of precursor zonadhesin unrelated to weight (see subunit I–IV in Fig. 1). Processing of rabbit and mouse zonadhesin, on the other hand, might deviate from this general pattern, as both species have lost either the EGF-like motif or the DP cleavage site in domain D1 and D3, respectively (Fig. 7A and D).

The present dataset also sheds light on co-evolutionary forces acting on the level of single motifs. Given the repetitive character of domains D1–4 pointing towards a common origin, the high conservation of EGF-like motifs and DP doublets in domains D1, D2, and D4 suggests that the absence of the respective motifs in domain D3 represents a derived character state. Whilst all species lack the DP doublet in domain D3, but only some the corresponding EGF-like motif, the following evolutionary scenario seems probable (Fig. 7C): According to this, the DP motif and thus the probable hydrolyzation site (Bi et al., 2003) was lost in domain D3 in the stem line of Boreoeutheria (see Murphy et al., 2001). This loss apparently entailed a relaxation of

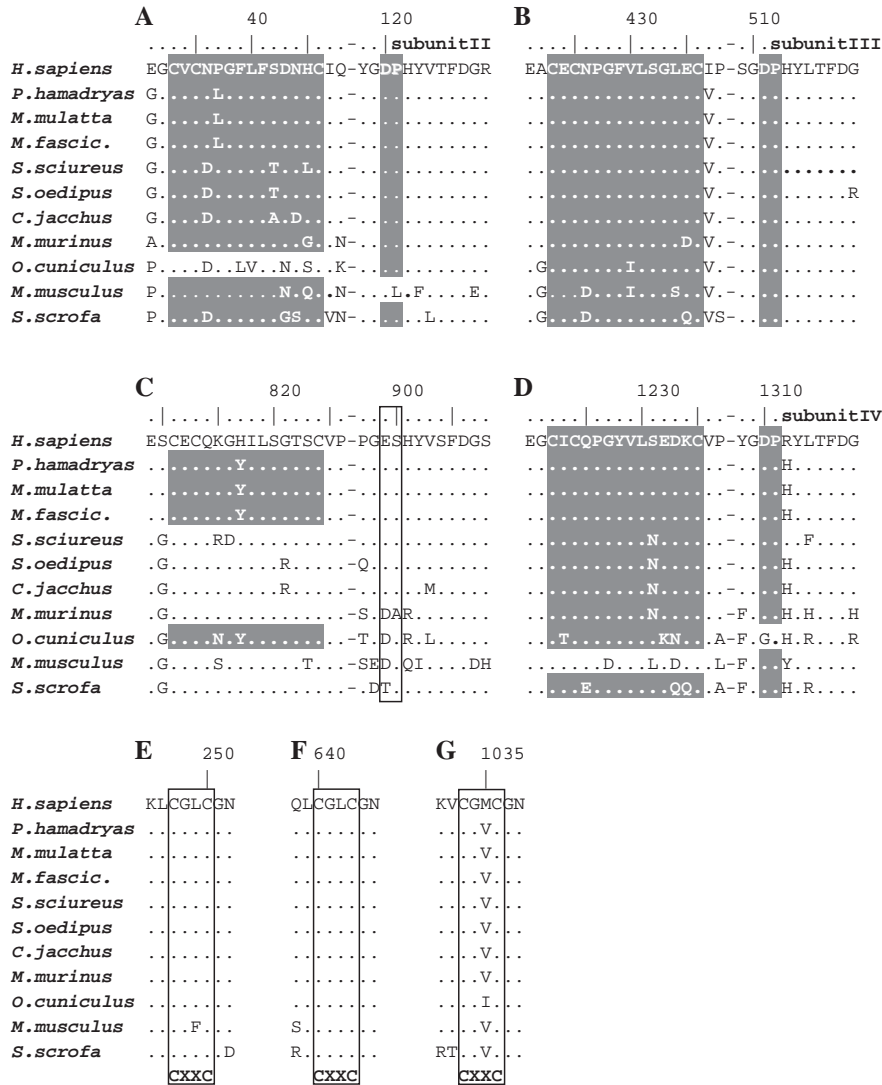


Fig. 7. Subsections of the amino acid alignment of zonadhesin domains D1-4. Dots indicate identities with the human reference sequence. A–D The degree of conservation of EGF-like motifs and corresponding DP doublets (shaded motifs in A–D). Given the total loss of the DP motif in domain D3, the homologue positions have been highlighted by a box (C). Hyphens in the alignments represent about 85 excised amino acids. Bold type “subunit II, III, and IV” in A, B, and D, respectively, highlight the first amino acid of the corresponding zonadhesin subunits. E–F Conservation of CXXC motifs across domains D1–D3.

functional constraints acting on the corresponding EGF-like motif that until then was involved in peptidase attachment (Gao and Garbers, 1998). The freshly emerged functional dispensibility then resulted in a convergent reduction of the EGF-like motif in the lineages leading to pig, mouse, mouse lemur, Platyrrhini, and humans.

In prepro von Willebrand factor and mucin 2, CXXC motifs are responsible for self-oligomerization via disulfide bridges (see Gao and Garbers, 1998; Lea et al., 2001). The overall high phylogenetic conservation of three CXXC motifs along domains D1–D3 throughout the entire sampling (Fig. 7E–F) thus suggests that in all primates considered, subunits II and III form a dimer via disulfide bonds (for subunits I–IV see Fig. 1). This is similar to what has previously been shown for pig (Hardy and Garbers, 1995; Bi et al., 2003). The CXXC motifs are thus apparently

negatively selected due to high functional constraints. This corroborates that the dimerization of subunits II and III is highly significant for zonadhesin functionality, a characteristic expressed also by the high zona-pellucida avidity of the resulting dimer (Bi et al., 2003).

3.5. Driving forces behind the observed evolutionary patterns

For marine molluscs and sea urchins, the sperm-ligands lysin and bindin have been shown to rapidly evolve in adaptation to corresponding zona pellucida receptors (reviewed in Swanson and Vacquier, 2002; see also Kamei and Glabe, 2003). As mentioned in Section 1., recent immunovisualization and results of zona pellucida binding assays revealed that zonadhesin domains D1-3 and subunits

II–III, respectively, bind to the zona pellucida (Gao and Garbers, 1998; Hickox et al., 2001; Lea et al., 2001; Bi et al., 2003). Moreover, subunit IV shows affinity for the zona pellucida, at least in rabbit (Lea et al., 2001). Zona pellucida glycoproteins 2 and 3, in turn, have been shown to evolve rapidly in mouse and human (Swanson et al., 2001b). The occurrence of positive selection and of taxon-specific differences in posttranslational modification along domains D1–4 might thus represent adaptations to a constantly changing receptor on the zona pellucida. As sperm-competition accelerates the evolution of sperm-associated proteins (Wyckoff et al., 2000; Dorus et al., 2004; present analysis), female cryptic choice might represent the deeper driving force behind the observed patterns.

The present findings moreover suggest that positive evolution contributes to species-specificity in sperm-egg interaction not only by amino acid replacements and sequence rearrangements (Metz and Palumbi, 1996; Hellberg et al., 2000; Swanson et al., 2001a), but also by changes in posttranslational modification. Considering similar patterns in immune relevant proteins of HIV-1 (De Oliveira et al., 2004), an association of positive evolution and changes in posttranslational modification seems to represent a general principle in protein evolution. In the case of zonadhesin (and other proteins involved in sperm-egg interaction), differences in posttranslational modification might contribute to changes in activity, conformation, and/or binding properties of zonadhesin D domains. This might consequently lead to the establishment of reproductive barriers between populations and thus to speciation. On the other hand, the high conservation of processing pattern, dimerization, and part of the posttranslational modification indicates that zonadhesin's quaternary structure needs to meet certain basic requirements to retain its function in zona pellucida binding.

4. Conclusions

Based on new zonadhesin D domain encoding sequences of non-human primates (1 Strepsirhini, 3 Platyrrhini, 3 Cercopithecoidea) plus GenBank orthologues of human, mouse, rabbit, and pig, the present site-specific sequence analysis (CODEML and HyPhy) indicates that zonadhesin domains D0–4 are subject to positive evolution. Usage of the beta ω M8 leads to the identification of 29 codon sites ($p_{(\omega>1)} > 0.5$, M3 and M8), for which positive selection can be assumed (mean $\omega \sim 2.7$). About 50% of these candidate sites coincide with evidence for convergent evolution on the amino acid level. Nineteen of the 29 codon sites pinpointed are very likely candidates for positive selection ($p_{(\omega>1)} \geq 0.95$ from M3). Differences in position and number of identified amino acid sites between the present study and a previous approach (Swanson et al., 2003) are probably due to different samples and different analytical tools.

Moving window underscores the lineage-specific evolution of zonadhesin D domains. The level of positive selection is highest in *M. murinus* (peaks up to $\omega > 6$), followed by Platyrrhini (ω peaks around 2), and Catarrhini including humans (ω maxima at circa 1). As lemurs in general and mouse lemur in detail are more promiscuous than Platyrrhini and Catarrhini, the present analysis reflects a positive correlation of positive evolution and sperm-competition due to promiscuity. The present study thus extends previous conclusions drawn from the evolution of protamine and *SEMG2* within Anthroidea (Wyckoff et al., 2000; Dorus et al., 2004) to the even more promiscuous lemurs.

Applying the principle of phylogenetic shadowing (Boffelli et al., 2003) on zonadhesin D domains, a set of phylogenetically old and widely conserved *N*-myristylation, *N*-glycosylation, amidation, and phosphorylation motifs has been identified that can be regarded as essential for the function of zonadhesin within Boreoeutheria, Glires, primates, and Anthroidea. Moreover, convergent evolution of motifs along different lineages has been shown to be a common feature in the evolution of predicted motifs. Finally, Fisher's exact test indicates that the observed accumulation of candidate sites for positive selection across variable posttranslational motifs is significantly different ($p = 0.001$) from the general distribution of candidate sites for positive selection across the entire alignment. From this we conclude that positive selection promotes differences between lineages not only via exchanges of amino acids but also via changes in posttranslational modification.

The high conservation of possible protease attachment sites, i.e. EGF-like motifs, and of possible hydrolyzation sites, i.e. DP doublets, in domains D1, 2, and 3 suggests identical processing of zonadhesin in primates and pig into four subunits (Bi et al., 2003), termed subunits I–IV herein. The high conservation of three CXXC motifs further points to dimerization of zonadhesin subunits II and III in human via disulfide bonds (see e.g., Gao and Garbers, 1998). Presence/absence analysis of the EGF-like motifs and DP doublets across domains D1–4 suggests co-evolutionary forces for both motifs.

Given zonadhesin's affinity for zona pellucida (Hardy and Garbers, 1995; Lea et al., 2001; Bi et al., 2003), positive evolution of D domains might be driven by a constantly changing receptor on the egg site. The resulting changes in primary structure and posttranslational modification of zonadhesin D domains might contribute to the species-specificity of zonadhesin's binding to the zona pellucida *via* changes in activation, conformation, and binding properties. Possibly, they also contribute to the formation of prezygotic barriers between populations and thus to speciation. On the other hand, the high conservation of processing, dimerization, and part of the posttranslational modification highlights the fact that zonadhesin's backbone needs to meet certain basic requirements in order to retain function.

Acknowledgements

We are indebted to Dr. C. Roos, Primate Genetics, and to the members of the Departments of Reproductive Biology, Neurobiology, and Veterinary Medicine and Primate Husbandry (all German Primate Center Göttingen) for providing most of the material. Special thanks go to Dr. M. Perret (Laboratoire d' Ecologie, Brunoy/France) and Prof. B. Brenig and his group (Veterinary Medicine, University of Göttingen) for the mouse lemur and pig material, respectively. Moreover, we would like to thank two anonymous reviewers for helpful comments. The laboratory work was carried out at the German Primate Center Göttingen. Financial support was granted by the German Research Foundation (HE 3487/1-1).

References

- Bercovitch, F.B., et al., 2002. Multiple sirehood in free-ranging twin rhesus macaques (*Macaca mulatta*). *Am. J. Primatol.* 57, 31–34.
- Bi, M., Hickox, J.R., Winfrey, V.P., Olson, G.E., Hardy, D.M., 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.
- Boffelli, D., et al., 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 299, 1391–1394.
- Briscoe, A.D., 2002. Homology modeling suggests a functional role for parallel amino acid substitutions between bee and butterfly red- and green-sensitive opsins. *Mol. Biol. Evol.* 19, 983–986.
- Creevey, C., McInerney, J.O., 2002. An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene* 300, 43–51.
- De Oliveira, T., et al., 2004. Mapping sites of positive selection and amino acid diversification in the HIV genome: an alternative approach to vaccine design? *Genetics* 167, 1047–1058.
- Digby, L.J., 1999. Sexual behavior and extragroup copulations in a wild population of common marmosets (*Callithrix jacchus*). *Folia Primatol.* 70, 136–145.
- Dorus, S., Evans, P.D., Wyckoff, G.J., Choi, S.S., Lahn, B.T., 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics*, Advance online publication.
- Fietz, J., 1999. Mating system of *Microcebus murinus*. *Am. J. Primatol.* 48, 127–133.
- Filip, L.C., Mundy, N.I., 2004. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. *Mol. Biol. Evol.* 21, 1504–1511.
- Gao, Z., Garbers, D.L., 1998. Species diversity in the structure of zonadhesin, a sperm-specific membrane protein containing multiple cell adhesion molecule-like domains. *J. Biol. Chem.* 273, 2421–2415.
- Hardy, D.M., Garbers, D.L., 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.
- Hellberg, M.E., Moy, G.W., Vacquier, V.D., 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. *Mol. Biol. Evol.* 17, 458–466.
- Herlyn, H., Zischler, H., in press. Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. *Mol. Phylogenet. Evol.*
- Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215–219.
- Hickox, J.R., Bi, M., Hardy, D.M., 2001. Heterogeneous processing and zona pellucida binding activity of pig zonadhesin. *J. Biol. Chem.* 276, 41502–41509.
- Kamei, N., Glabe, C.G., 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes Dev.* 17, 2502–2507.
- Kosakovsky, S.L., Frost, S.D.W., Muse, S.V., 2004. HyPhy: hypothesis testing using phylogenies. *Bioinformatics Advance Access*.
- Kouprina, N., et al., 2004. The SPANX gene family of cancer/testis-specific antigens: rapid evolution and amplification in African great apes and hominoids. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3077–3082.
- Lea, I.A., Sivashanmugam, P., O'Rand, M.G., 2001. Zonadhesin: characterization, localization, and zona pellucida binding. *Biol. Reprod.* 65, 1691–1700.
- Li, W.H., 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36, 96–99.
- Messier, W., Stewart, C.-B., 1997. Episodic adaptive evolution of primate lysozymes. *Nature* 385, 151–154.
- Metz, E.C., Palumbi, S.R., 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* 13, 397–406.
- Murphy, W.J., et al., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294, 2348–2351.
- Olson, G.E., Winfrey, V.P., Bi, M., Hardy, D.M., NagDas, S.K., 2004. Zonadhesin assembly into the hamster sperm acrosomal matrix occurs by distinct targeting strategies during spermiogenesis and maturation in the epididymis. *Biol. Reprod.* 71, 1128–1134.
- Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and application to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Smith, R.J., Cheverud, J.M., 2002. Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's Rule in Primates. *Int. J. Primatol.* 23, 1095–1135.
- Sumiyama, K., Saitou, N., Ueda, S., 2002. Adaptive evolution of the IgA hinge region in primates. *Mol. Biol. Evol.* 19, 1093–1099.
- Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* 16, 1315–1328.
- Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. *Nat. Rev., Genet.* 3, 137–144.
- Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F., Aquadro, C.F., 2001a. Evolutionary *EST* analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7375–7379.
- Swanson, W.J., Yang, Z., Wolfner, M.F., Aquadro, C.F., 2001b. Positive selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2509–2514.
- Swanson, W.J., Nielsen, R., Yang, Q., 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* 20, 18–20.
- Torgerson, D.G., Singh, R.S., 2003. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. *Mol. Biol. Evol.* 20, 1705–1709.
- Wang, Y., Su, B., 2004. Molecular evolution of *microcephalin*, a gene determining human brain size. *Hum. Mol. Genet.* 13, 1131–1137.
- Wang, X., Zhang, J., Zhang, Y.-P., 2002. Erratic evolution of SRY in higher primates. *Mol. Biol. Evol.* 19, 582–584.
- Wyckoff, G.J., Wang, W., Wu, C.-I., 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* 403, 304–309.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556 (<http://abacus.gene.ucl.ac.uk/software/paml.html>).
- Yang, Z., Nielsen, R., Goldman, N., Krabbe Pedersen, A.-M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Zhang, J., 2003. Parallel functional changes in the digestive RNases of ruminants and colobines by divergent amino acid substitutions. *Mol. Biol. Evol.* 20, 1310–1317.