

Sexual Selection and the Molecular Evolution of ADAM Proteins

Scott Finn · Alberto Civetta

Received: 14 May 2010 / Accepted: 9 August 2010 / Published online: 22 August 2010
© Springer Science+Business Media, LLC 2010

Abstract Rapid evolution has been identified for many reproductive genes and recent studies have combined phylogenetic tests and information on species mating systems to test sexual selection. Here we examined the molecular evolution of the ADAM gene family, a diverse group of 35 proteins capable of adhesion to and cleavage of other proteins, using sequence data from 25 mammalian genes. Out of the 25 genes analyzed, all those expressed in male reproductive tissue showed evidence of positive selection. Positively selected amino acids within the protein adhesion domain were only found in sperm surface ADAM proteins (ADAMs 1, 2, 3, 4, and 32) suggesting selection driven by male × female interactions. We tested heterogeneity in rates of evolution of the adhesion domain of ADAM proteins by using sequence data from Hominiidae and macaques. The use of the branch and branch-site models (PAML) showed evidence of higher d_N/d_S and/or positive selection linked to branches experiencing high postmating selective pressures (chimpanzee and macaque) for *Adams* 2, 18, and 23. Moreover, we found consistent higher proportion of nonsynonymous relative to synonymous and noncoding sequence substitutions in chimpanzee and/or macaque only for *Adams* 2, 18, and 23. Our results suggest that lineage-specific sexual selection bouts might have driven the evolution of the adhesion sperm protein surface domains of ADAMs 2 and 18 in primates. *Adams* 2 and 18 are localized in chromosome 8 of primates and

adjacent to each other, so their evolution might have also been influenced by their common genome localization.

Keywords ADAM gene family · Sperm · Sexual selection

Introduction

Several studies in a wide variety of taxa have demonstrated that genes with a role in reproduction are often rapidly evolving and in many cases rapid evolution is driven by adaptive selection (Civetta and Singh 1998; Swanson and Vacquier 2002; Civetta 2003a; Clark et al. 2006; Panhuis et al. 2006). In mammals, fertilization is internal and complex in the sense that it requires the sperm to cross a series of barriers that involve complex interactions with egg proteins (Primakoff and Myles 2007). There are many examples of mammalian sperm surface proteins that show evidence of rapid interspecific sequence divergence driven by positive selection (Torgerson et al. 2002; Swanson et al. 2003; Glassey and Civetta 2004; Dorus et al. 2010). However, few studies have directly tried to link episodes of positive selection with changes that can be used as a measure of sexual selection. One of the best examples comes from studies of genes coding for components of the semen coagulum. Population genetics analysis of Seme-nogelin 1 (*Semg1*) has shown evidence of low polymorphism but high rate of evolution leading to species with promiscuous mating systems (Kingan et al. 2003) but studies using a phylogenetic based approach have found no evidence of long-term selection for *Semg1* along more distant lineages leading to primates for whom sperm competition is common (Ramm et al. 2008; O'Connor and Mundy 2009). The discrepancy in results simply highlights

Electronic supplementary material The online version of this article (doi:10.1007/s00239-010-9382-7) contains supplementary material, which is available to authorized users.

S. Finn · A. Civetta (✉)
Department of Biology, University of Winnipeg, 515 Portage
Ave., Winnipeg, MB R3B 2E9, Canada
e-mail: a.civetta@uwinnipeg.ca

the fact that the two studies have measured selection at different time points in the gene's evolutionary history. Associations between bouts of positive selection and differences in mating types have been found consistently for *Semg2* in both primates and rodents but the evidence is less clear or even lacking for other genes expressed in reproductive tissue (Dorus et al. 2004; Hurlle et al. 2007; Ramm et al. 2008; O'Connor and Mundy 2009).

While adaptive evolution of reproductive genes is widespread it is not always clear whether sexual selection, species-specific adaptations to fertilization or other factors such as species diversification in immune-related function have triggered rapid evolution (Good and Nachman 2005; Dean et al. 2008). The ADAM gene family is an ideal group of genes to address whether positive selection at reproductive genes is likely driven by postmating selective pressures imposed by differences in species mating habits. The family is comprised of 35 characterized genes with a variety of functions, both tissue-specific as well as ubiquitous patterns of expression, and a common evolutionary history. More importantly, the ADAMs are transmembrane proteins involved in cell adhesion, proteolysis, and signaling (Primakoff and Myles 2000; Edwards et al. 2008) and so well characterized proteolytic and adhesion motifs can be found within their amino acid sequences. Therefore, evolutionary studies can be aimed not only at attempting to associate evolutionary rate heterogeneity and positive selection episodes with differences in mating systems but also at mapping the distribution of positively selected sites across domains with different cellular functions. In mammals, ADAM1, ADAM2, and ADAM3 are the best characterized in terms of their role during fertilization. They play a role during sperm zona pellucida binding as well as egg membrane recognition and fusion (Blobel et al. 1992; Wolfsberg et al. 1995; Primakoff and Myles 2000; Evans 2002). ADAMs 4, 5, 6, 18, 24, and 32 are also sperm surface proteins while other ADAMs have been identified only as testes expressed (Frayne et al. 2002; Kim et al. 2006; Zhu et al. 2009). Molecular evolution studies have suggested that the evolution of *Adam1*, *Adam2*, and *Adam32* have been driven by positive selection and that selection is limited to ADAMs with a role in fertilization (Civetta 2003b; Glassey and Civetta 2004). However, formal tests of positive selection of many ADAM genes were limited by the low number of species for which sequences were available (Glassey and Civetta 2004). A recent sperm proteome study has found 23 sperm surface genes to be under positive selection when comparing mouse sperm genes with orthologs in rat, human, chimpanzee or macaque, and dog or cow. Of the 23, five were ADAM genes including *Adam2* (Dorus et al. 2010).

Here we revisited the molecular evolution of mammalian ADAM gene family taking advantage of a larger

number of sequences currently available in public databases. We also analyzed the distribution of positively selected sites across protein domains. We found that only sperm surface ADAM proteins show evidence of positive selection localized to their adhesion domain. We further tested lineage-specific rates of evolution within the adhesion domain of ADAM proteins in primates and found evidence of lineage-specific rate heterogeneity and higher proportion of nonsynonymous relative to both synonymous and noncoding sequence substitutions for the adhesion domains of *Adams 2* and 18 in chimpanzee and *Adams 2* and 23 in macaques.

Methods

Mammalian Sequence Data Collection and Analysis

We retrieved nucleotide and amino acid sequence data from GenBank for 25 *Adam* genes (for accession numbers see Supplementary Table 1). Amino acid sequences were aligned using the global alignment algorithm ClustalX (Thompson et al. 1997) and the local alignment algorithm DiAlign2 (Morgenstern et al. 1996), and the generated alignments were used as reference for nucleotide sequence alignments using Pal2Nal (Suyama et al. 2006). To confirm that the gene sequences labeled as member of a specific ADAM gene family were members of a monophyletic group, the entire ADAM gene phylogeny was reconstructed using a Neighbor-Joining Distance method (MEGA 4.0) and a Maximum Parsimony (MP) approach (PAUP 4.0) (Swofford 2003; Tamura et al. 2007). For the MP tree reconstruction, gaps were treated as a fifth character state and all characters were weighted equally. A heuristic search was conducted using the Tree Bisection-Reconnection (TBR) branch-swapping algorithm and the search was allowed to swap to completion. The strict consensus tree was chosen. The reliability of the inferred distance and MP trees was assessed by bootstrapping with 1,000 replicates (Felsenstein 1985).

Tests of selection were performed for each ADAM gene using the codeml package within Phylogenetic Analysis by Maximum Likelihood (PAML; v4.2) (Yang 1997). We tested for evidence of positive selection by comparing the likelihood of a pair of nested models. Model M7 uses a beta distribution to fit ω values to classes between 0 and 1 while model M8 adds a class of ω ratio higher than 1. Twice the log-likelihood difference between the two models was compared to a chi-square threshold value with the appropriate degrees of freedom (Yang et al. 2000). To validate that any signal of positive selection resulting from an M8 vs. M7 model comparison were not the result of relaxed selection, the likelihood of model M8 was also

compared to the likelihood of null model M8a, where ω is fixed at one (Swanson et al. 2003; Wong et al. 2004). We adjusted all P value thresholds to reduce the number of false positives to less than one. In cases where we detected evidence of codons under positive selection, the Bayes Empirical Bayes (BEB) method was used to identify specific codon sites under selection and to assign a posterior probability (Yang et al. 2005). Positively selected sites were mapped within protein domains identified by scanning amino acid sequences against the PROSITE patterns and profiles database (<http://www.expasy.ch/prosite/>).

Primate Sequence Data Collection and Analysis

We classified species into two groups based on their mating system classification (Dixon 1998). Species with a multi-male–multifemale or a disperse mating system and hence more likely experiencing postmating sexual selection competition included chimpanzee (*Pan troglodytes*), and macaques (*Macaca mulatta* or *Macaca fascicularis*). Species with a polygynous or monogamous mating system and less likely to experience postmating sexual selection included orangutan (*Pongo abelii*) and gorilla (*Gorilla gorilla*). Humans (*Homo sapiens*) were classified as a group experiencing less intense postcopulatory sexual selection than chimpanzee and macaques, and grouped with the monogamous/polygynous species (Dixon 1998, p. 219).

DNA sequences from the adhesion domain of gorilla and orangutan ADAM genes were retrieved from GenBank by using the amino acid sequence data available from humans and chimpanzee to tblastn the gorilla and orangutan Whole-Genome Shotgun reads (WGS) database. The hits were selected on the basis of alignment similarities and the sources used for assemblies are provided in appendix section (Supplementary Table 2). We calculated the ω ratio, d_N and d_S using the free ratio model within PAML (Yang 1997). For ADAM genes showing evidence of ω higher than one in any lineage, we also tested rates of nucleotide divergence within the coding regions of the adhesion domain relative to noncoding introns or flanking sequence (i.e. neutral control). The position of introns within adhesion domains of gorilla and orangutan sequences were delineated by gaps in between tblastn hits listed in supplementary material (Supplementary Table 2). Human and chimpanzee intron sequence data was retrieved by aligning the nucleotide sequence coding for each ADAM adhesion domain to the chromosome genome reference sequence assembly within GenBank (*Adams* 2, 18, and 32 NC_000008.10, *Adam12* NC_000010.10, *Adam23* NC_000002.11 for *Homo sapiens* and *Adams* 2, 18, and 32, NC_006475.2, *Adam12* NC_006477.2, *Adam23* NC_006470.2 for *Pan troglodytes*). For macaques, we retrieved intron sequence using the adhesion domain amino acid sequence to tblastn the *Macaca*

mulatta WGS database (Supplementary Table 2). We found no intron sequence for *Adam30* and so noncoding sequence flanking the gene was retrieved as outlined above using tblastn searches against WGS databases or retrieval from the chromosome genome reference sequence assembly within Genbank (NC_000001.10 for *Homo sapiens* and NC_006468.2 for *Pan troglodytes*).

We used the baseml program within PAML to reconstruct ancestral sequences under models that assume rate variation over nucleotide sites and different patterns of nucleotide substitutions. We used model HKY85 which assumes different frequencies for the four nucleotides (π) and different transition/transversion rate ratios. Simpler models assume no differences in transition/transversion rate ratios (F81) equal frequencies for all nucleotides (K80) and both equal nucleotide frequencies and no difference in transition/transversion rate ratios (JC69) (Yang et al. 1994). The log-likelihood of the species tree topology for each ADAM gene was calculated under the different models and twice the log-likelihood difference between models were compared to a chi-square distribution with degrees of freedom given by the difference in number of parameters estimated for each model. The model that better explained the data was used to reconstruct a sequence state at each internal node of the tree. The ancestral sequence was used to estimate noncoding sequence divergence along each branch using MEGA (v4.0).

We tested for evidence of heterogeneity in evolutionary rate by using the branch model within PAML. We compared the likelihood of a two-ratio model (M2) with foreground branches being those likely experiencing postmating selective pressures to the likelihood of the one-ratio model which assumes equal rates of evolution across branches. We also tested for evidence of positive selection along the chimpanzee and macaque foreground branches using the mixed branch-site model (model = 2; NSsites = 2) within codeml (Yang and Nielsen 2002). The log-likelihood of the branch-site model was compared to the same model but fixing the ω value of the foreground branches to 1, so that any significant variation in ω between foreground and background branches could be attributed to positive selection as opposed to differences in selective constraints (Zhang et al. 2005). All tests were also conducted using sequence data from Hominidae and macaques available for SEMG2 (Supplementary Table 2) to serve as a positive control.

Results

ADAM Genes Phylogeny

We used the reconstructed phylogeny of the whole ADAM gene family as a tool to identify true monophyletic groups

for analysis of selection. We found consistent grouping of genes regardless of the approach used. The reconstruction of the ADAM genes phylogeny using both distance and MP methods (not shown) confirmed orthology for most ADAM genes with the exception of *Adam20* and *Adam29*. *Bos taurus* and *Canis familiaris Adam20* genes did not group with the other three *Adam20* genes and thus *Adam20* was not included in phylogenetic tests of rate heterogeneity or positive selection. *Bos taurus Adam29* did not cluster with other *Adam29* genes and was excluded from the PAML analysis of *Adam29* (Fig. 1). The distance method tree identified all testes expressed ADAM genes as having a common monophyletic origin and subdivided into two clades. One such clade (*Adams* 1, 4, 6, 20, 21, 29, and 30) is in agreement with one of the three major clades resolved in a previous study that have phylogenetically grouped ADAM genes, and so is the grouping of *Adam17* and *Adam10* (see clades A and C in Huxley-Jones et al. 2007). Both the distance and MP trees placed one of the testes expressed ADAMs group (*Adams* 2, 3, 5, 18, and 32) as a common monophyletic cluster with high bootstrap support (100 and 94, respectively) (Fig. 1).

Positive Selection Within the Adhesion Domain of ADAM Sperm Surface Genes in Mammals

We tested for evidence of positive selection among mammals for 24 genes of the ADAM family for which we could find sequence data from four or more species. Fifteen genes showed evidence of positive selection (significant likelihood ratio test: M8 vs. M7 and M8 vs. M8a) (Table 1). Seven (*Adams* 1, 2, 3, 4, 5, 6, and 32) of the 15 are known to be sperm surface ADAM expressed genes, while *Adams* 21, 28, 29, and 30 are testes expressed genes and *Adam7* is an epididymis expressed gene possibly involved in sperm maturation (Oh et al. 2005). Only three of the fifteen genes detected to be under positive selection are not reproductive-related genes (*Adams* 8, 11, and 15) and nine other genes showed no evidence of positive selection (Table 1). In summary, all reproductive ADAM genes (both testes expressed and epididymis) showed some evidence of positive selection but positive selection was not restricted to ADAM genes with a role in reproduction (Table 1).

ADAM genes are characterized by the presence of identifiable metalloprotease and adhesion (disintegrin and cysteine-rich) domains within their protein sequences. The family is also known as the MDC family, the name highlights the three protein domains found in all members of the gene family (metalloprotease, disintegrin, and cysteine-rich). An analysis of the distribution of positively selected sites across domains revealed that only sperm surface ADAMs 1, 2, 3, 4, and 32 showed evidence of positive

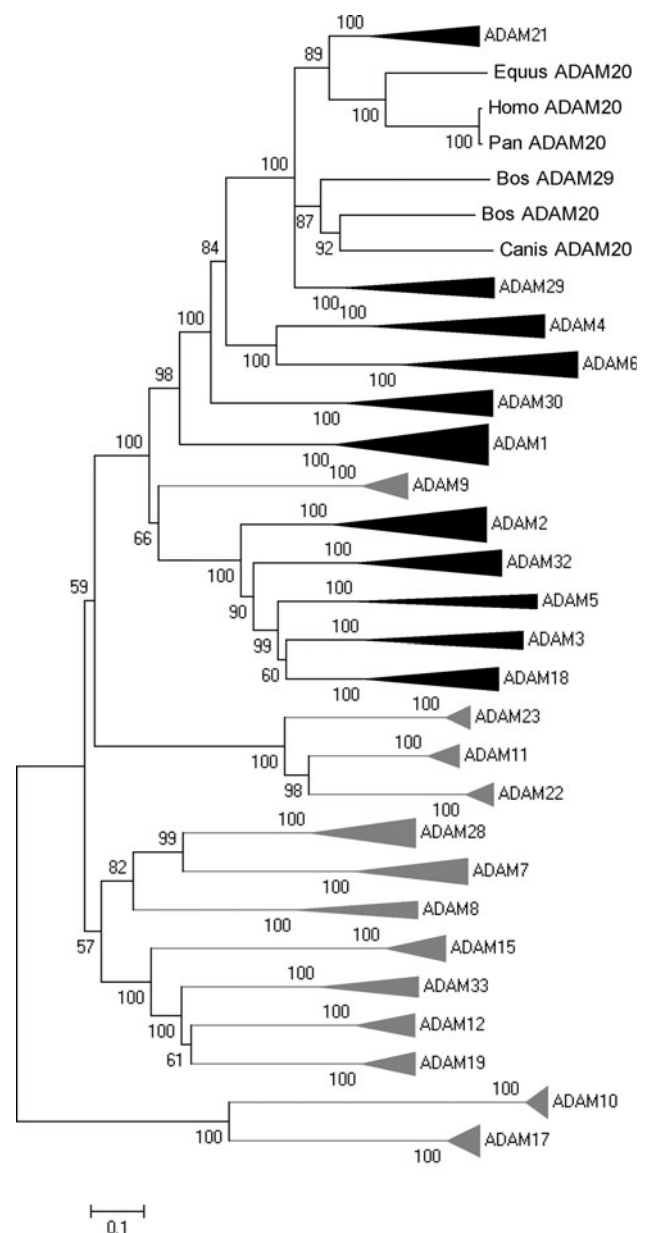


Fig. 1 Neighbor-Joining Poisson corrected phylogenetic tree. All ADAM genes, with the exception of ADAM20, form monophyletic groups and the gene terminal nodes are collapsed. Black nodes are testes expressed ADAM genes. Bootstrap values are shown

selection within the adhesion domain of the protein (Fig. 2).

Changes in Evolutionary Rate and Selection Within the Adhesion Domain of Primate ADAM Genes

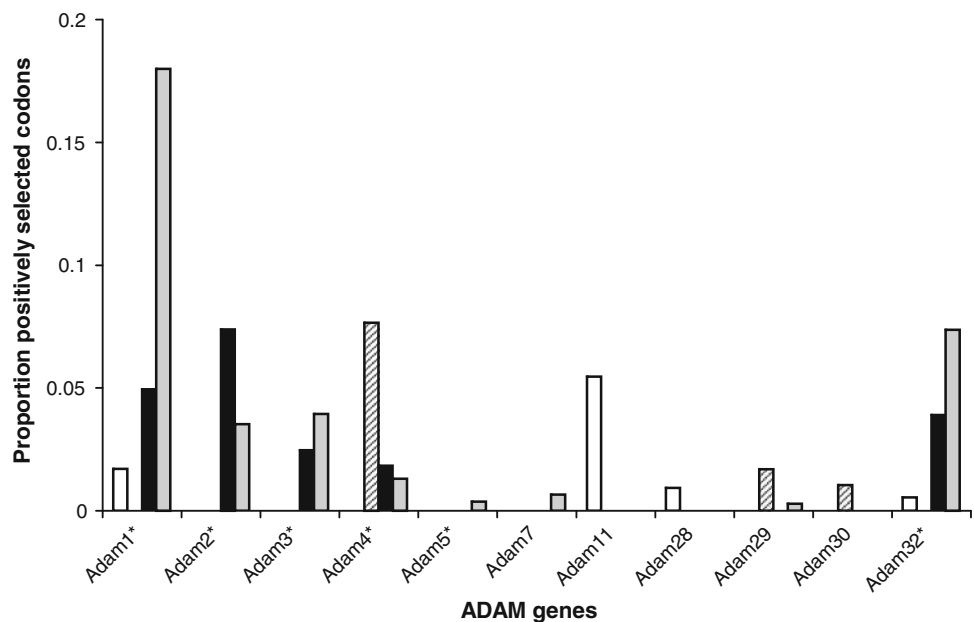
The fact that only some mammalian sperm surface ADAMs showed evidence of positively selected sites within the protein adhesion domain led us to hypothesize that such domain might be critical in terms of male × male and/or male × female interactions for fertilization. Therefore, it is

Table 1 Results from codon tests of positive selection using mammalian ADAM genes

Gene	<i>N</i>	Tl	$\ell_{(M7)}$	$\ell_{(M8)}$	$2\Delta\ell_{M7-M8}$	$\omega_{(M8)}$	$\ell_{(M8a)}$	$2\Delta\ell_{M8-M8a}$
<i>Adam1</i>	14	5.62	-15764.7	-15664.1	201.1***	3.0	-15764.5	200.8***
<i>Adam2</i>	12	5.41	-14002.1	-13931.3	141.4***	3.4	-13983.4	104.3***
<i>Adam3</i>	6	3.21	-9029.2	-8998.0	62.5***	3.4	-9020.2	44.5***
<i>Adam4</i>	8	3.96	-11459.5	-11393.1	132.8***	4.3	-11447.2	108.3***
<i>Adam 5</i>	5	3.75	-9095.4	-9083.6	23.6***	11.4	-9092.0	16.7***
<i>Adam6</i>	9	5.59	-13933.3	-13913.4	39.8***	2.2	-13930.0	33.2***
<i>Adam7</i>	9	2.75	-9747.8	-9735.0	25.7***	2.7	-9743.9	17.8***
<i>Adam8</i>	6	3.43	-7935.3	-7929.1	12.3***	8.7	-7934.1	10.0**
<i>Adam9</i>	9	1.65	-7829.1	-7829.0	0.1		-7829.0	0.0
<i>Adam10</i>	11	1.02	-2075.9	-2075.5	0.8		-2075.5	0.0
<i>Adam11</i>	8	1.29	-6346.1	-6312.0	68.2***	8.9	6340.3	56.6***
<i>Adam12</i>	8	1.90	-8837.8	-8837.7	0.3		-8837.7	0.0
<i>Adam15</i>	9	2.06	-8139.2	-8134.1	10.2**	2.6	-8137.7	7.2**
<i>Adam17</i>	11	2.38	-144.6	-144.6	0.0		-144.6	0.0
<i>Adam18</i>	8	2.45	-8122.2	-8120.2	3.9		-8119.7	1.0
<i>Adam19</i>	8	1.57	-4713.4	-4713.3	0.1		-4713.4	0.2
<i>Adam21</i>	7	2.07	-7742.1	-7738.0	8.1**	3.3	-7741.1	6.2*
<i>Adam22</i>	8	1.08	-6712.5	-6712.5	0.0		-6712.4	0.2
<i>Adam23</i>	8	1.26	-6768.9	-6768.7	0.3		-6768.9	0.3
<i>Adam28</i>	10	2.48	-9025.0	-9014.5	21.0***	2.5	-9023.3	17.6***
<i>Adam29</i>	7	3.05	-9054.0	-9030.5	47.0***	3.3	-9051.5	42.0***
<i>Adam30</i>	9	3.6	-7556.0	-7544.1	23.7***	2.6	-7554.8	21.4***
<i>Adam32</i>	9	3.36	-8381.9	-8334.7	94.5***	3.2	-8369.9	70.4***
<i>Adam33</i>	7	2.28	-6201.9	-6199.4	4.9		-6199.7	0.6

Note: *N* is the number of terminal branches in the tree topology analyzed. ℓ is the likelihood of the different models tested (M7, M8, and M8a; see methods). Tl is the tree length. $2\Delta\ell$ is the likelihood ratio test statistic (M8 vs. M7 and M8 vs. M8a). We use conservative thresholds of * $P < 0.025$, ** $P < 0.01$, and *** $P < 0.001$ so that we reduced the expected number of false positives to less than 1, 0.5, and 0.05, respectively. ω (d_N/d_S per codon) estimates are given for genes showing evidence of positive selection

Fig. 2 Distribution of codon sites for 11 out of 15 *Adam* genes showing evidence of positive selection. *Adam* genes with a BEB posterior probability lower than 0.95 of being under positive selection are not shown (*Adams* 6, 8, 15, and 21). The proportion of positively selected sites is shown for the amino end (white), metalloprotease domain (dashed), adhesion domain (black), and carboxy-tail end (gray) in that order for each gene. The name of sperm surface proteins is marked with an asterisk



likely that the evolution of the sperm surface protein adhesion domains is driven by sexual selection at the molecular level. To test this hypothesis we decided to focus our analysis of selection on the adhesion domain of a group of species with different mating types. Moreover, in order to avoid false positives due to ancient bouts of selection unrelated to reproductive differences and to loose true positives due to the inclusion of highly diverged species, we restricted our analysis to relatively closely related species of primates.

We used the branch model within PAML to test for variation in rates of evolution by comparing all species of the Hominidae family as well as macaques and looking for associations between acceleration in rates of evolution and differences in mating system types. We found that a model of evolution assuming a ω_A ratio for terminal branches leading to species with multimale–multifemale or dispersed mating systems (macaques and chimpanzee) and a different ω_B ratio for other branches as opposed to a single-ratio (ω_0) model, fit the data better for *Adam2* ($2\Delta\ell = 6.06$; $P = 0.014$; $\omega_A = \infty$ vs. $\omega_B = 0.30$) and *Adam23* ($2\Delta\ell = 6.38$; $P = 0.012$; $\omega_A = 1.94$ vs. $\omega_B = 0.07$) (Table 2). We further used a mixed branch-site model to test for positive selection linked to the two terminal branches more likely experiencing high postmating selective pressures and found significant results for *Adam18* ($2\Delta\ell = 5.60$; $P = 0.018$), *Adam23* ($2\Delta\ell = 50.44$; $P < 0.001$), and the positive control *Semg2* ($2\Delta\ell = 8.66$; $P = 0.003$).

Estimates of ω ratios for each branch in the tree showed *Adams 2, 12, 18, 23, 28, 30, and 32* having at least one branch with a ratio higher than one (Table 3). Since this ratio can be inflated due to low number of substitutions with no synonymous changes, we also compared the ratio of nonsynonymous substitutions relative to the proportion of substitutions in noncoding regions for genes showing ω higher than 1 along any given branch. We reconstructed ancestral sequences using models that assume different patterns of nucleotide substitutions. Log-likelihood estimates under different models of nucleotide substitution consistently showed model K80 fitted the data better than Model JC69 and model HKY85 was significantly better than model F81 (Supplementary Table 3). Because all ADAM genes tested had higher frequencies of AT compared to CG, we used model HKY85 for inferring nucleotide ancestral sequences (notice that log-likelihood estimates for HKY85 is lower than F81) (Supplementary Table 3). A consistent pattern of higher proportion of nonsynonymous substitutions relative to both synonymous and noncoding substitutions were found along the chimpanzee branch for *Adam2* and *Adam18* and along the macaque branch for *Adam2* and *Adam23* (Table 4).

Table 2 Likelihood ratio tests for the adhesion domain of primate ADAM proteins using the branch and branch-site models within PAML

Gene	<i>N</i>	Model	ℓ	$2\Delta\ell$	<i>P</i>
<i>Semg2</i>	6	M0	−2394.46		
		M2	−2393.63	1.65	0.200
		M2N2 (ω fix)	−2394.74		
		M2N2 (ω free)	−2390.41	8.66	0.003**
<i>Adam2</i>	6	M0	−667.944		
		M2	−664.914	6.06	0.014*
		M2N2 (ω fix)	−664.991		
		M2N2 (ω free)	−663.89	2.20	0.138
<i>Adam7</i>	5	M0	−541.54		
		M2	−541.37	0.355	0.552
		M2N2 (ω fix)	−539.64		
		M2N2 (ω free)	−539.21	0.854	0.355
<i>Adam9</i>	5	M0	−469.54		
		M2	−469.539	0.000	1.000
		M2N2 (ω fix)	−469.536		
		M2N2 (ω free)	−469.536	0.000	1.000
<i>Adam10</i>	5	M0	−577.17		
		M2	−577.17	0.000	1.000
		M2N2 (ω fix)	−577.18		
		M2N2 (ω free)	−577.17	0.035	0.852
<i>Adam12</i>	5	M0	−423.13		
		M2	−423.08	0.116	0.733
		M2N2 (ω fix)	−422.77		
		M2N2 (ω free)	−422.77	0.000	1.000
<i>Adam15</i>	5	M0	−676.91		
		M2	−675.15	3.504	0.061
		M2N2 (ω fix)	−669.15		
		M2N2 (ω free)	−668.23	1.826	0.177
<i>Adam18</i>	5	M0	−454.74		
		M2	−453.47	2.538	0.111
		M2N2 (ω fix)	−453.26		
		M2N2 (ω free)	−450.46	5.603	0.018*
<i>Adam19</i>	5	M0	−361.35		
		M2	−361.35	0.000	1.000
		M2N2 (ω fix)	−361.35		
		M2N2 (ω free)	−361.35	0.000	1.000
<i>Adam21</i>	5	M0	−544.63		
		M2	−544.27	0.721	0.396
		M2N2 (ω fix)	−544.63		
		M2N2 (ω free)	−544.63	0.000	1.000
<i>Adam22</i>	5	M0	−492.84		
		M2	−492.84	0.000	1.000
		M2N2 (ω fix)	−492.84		
		M2N2 (ω free)	−492.84	0.000	1.000

Table 2 continued

Gene	<i>N</i>	Model	ℓ	$2\Delta\ell$	<i>P</i>
<i>Adam23</i>	5	M0	−663.93		
		M2	−660.74	6.384	0.012*
		M2N2 (ω fix)	−641.854		
		M2N2 (ω free)	−616.634	50.440	<0.001***
<i>Adam28</i>	6	M0	−310.45		
		M2	−310.43	0.043	0.835
		M2N2 (ω fix)	−309.11		
		M2N2 (ω free)	−309.11	0.000	1.000
<i>Adam30</i>	6	M0	−515.89		
		M2	−515.04	1.706	0.191
		M2N2 (ω fix)	−515.07		
		M2N2 (ω free)	−515.04	0.056	0.813
<i>Adam32</i>	6	M0	−460.35		
		M2	−459.55	1.613	0.204
		M2N2 (ω fix)	−460.29		
		M2N2 (ω free)	−460.29	0.000	1.000

Note: *N* is the number of terminal branches in the tree topology analyzed. ℓ is the likelihood of the different models tested (M2 vs. M0 and M2N2 vs. M2N2 ω fix; see methods). $2\Delta\ell$ is the likelihood ratio test statistic. Significance thresholds are corrected as in Table 2 to reduce the expected number of false positives to less than one

Discussion

Our results expand on previous findings on the molecular evolution of ADAM family genes (Glasse and Civetta 2004) to show that all ADAM genes expressed in male reproductive tissue show evidence of positive selection. In fact, it has been recently shown that cell-surface sperm

genes have a significantly higher d_N than d_S in mouse and positive selection was detected for *Adams* 1, 2, 4, 6, and 24 using phylogenetic comparisons among five different species of mammals (Dorus et al. 2010). These results can be either driven by a strong selective signal across the entire or most of the evolutionary history of the gene, which might be likely for a gene like *Adam2*, or some lineages experiencing bouts of selection. We also confirmed our prior findings of positive selection of codon sites within the disintegrin/cysteine-rich adhesion domains of *Adams* 1, 2, and 32 (Civetta 2003b; Glasse and Civetta 2004) and expanded the result to establish that the signal of selection is also found in the adhesion domain of other sperm surface ADAM proteins (ADAMs 3 and 4). We know that gene disruption of *Adams* 1, 2, and 3 impairs the ability of sperm to move from the uterus to the oviduct and that knockouts for these genes are defective in sperm binding to the egg zona pellucida (Cho et al. 1998; Shamsadin et al. 1999; Kim et al. 2006; Yamaguchi et al. 2009). It is also known that ADAMs 1, 2, and 3 form protein complexes in both testicular cells and sperm (Cho et al. 2000; Nishimura et al. 2004, 2007). ADAMs 4, 5, 6, 30, and 32 are expressed in the sperm head and the use of mice knockouts has revealed that both ADAMs 4 and 6 associate with ADAMs 1 and 2 in the formation of sperm protein complexes (Kim et al. 2006; Han et al. 2009). Our results provide evidence that only some mammalian sperm ADAM genes, many of which are known to associate to form protein complexes with known functions in sperm movement within the female reproductive tract and in sperm–egg binding, have evolved by adaptive diversification of their protein adhesion domain. This result raises the possibility that such

Table 3 Results for lineage-specific ω ratio as well as proportions of nonsynonymous (d_N) to synonymous (d_S) substitutions per site

Gene	<i>l</i>	<i>Ptr</i>	<i>Hsa</i>	<i>Ggo</i>	<i>Pab</i>	<i>Mac</i>
<i>Adam2</i>	339	∞ (0.008; 0)	0 (0; 0)	∞ (0.004; 0)	0.475 (0.026; 0.054)	1.458 (0.061; 0.042)
<i>Adam7</i>	330	0 (0; 0.024)	0 (0; 0)	0.331 (0.004; 0.012)	0 (0; 0.024)	0 (0; 0.094)
<i>Adam9</i>	333	0 (0; 0)	0 (0; 0)	0 (0; 0.012)	0 (0; 0.038)	0 (0; 0)
<i>Adam10</i>	426	0 (0; 0.008)	0 (0; 0)	0 (0; 0)	0 (0; 0.016)	0 (0; 0)
<i>Adam12</i>	261	0 (0; 0)	0 (0; 0.015)	0 (0; 0.062)	∞ (0.005; 0)	0.053 (0.005; 0.098)
<i>Adam15</i>	246	0 (0; 0.011)	0.570 (0.006; 0.011)	0 (0; 0)	0 (0; 0.022)	0.192 (0.005; 0.026)
<i>Adam18</i>	276	∞ (0.015; 0)	0 (0; 0)	∞ (0.005; 0)	0.318 (0.015; 0.047)	∞ (0.014; 0)
<i>Adam19</i>	243	0 (0; 0.02)	0 (0; 0.02)	0 (0; 0)	0 (0; 0.08)	0 (0; 0.087)
<i>Adam21</i>	327	0.138 (0.004; 0.032)	0 (0; 0)	0 (0; 0)	0.846 (0.018; 0.021)	0 (0; 0.044)
<i>Adam22</i>	333	0 (0; 0)	0 (0; 0)	0 (0; 0.011)	0 (0; 0.034)	0 (0; 0)
<i>Adam23</i>	330	0 (0; 0)	0 (0; 0)	0 (0; 0.014)	∞ (0.004; 0)	∞ (0.106; 0)
<i>Adam28</i>	186	0 (0; 0)	0 (0; 0.026)	0 (0; 0)	∞ (0.007; 0)	∞ (0.021; 0)
<i>Adam30</i>	327	0 (0; 0)	0 (0; 0.019)	0 (0; 0)	∞ (0.004; 0)	19.181 (0.028; 0.001)
<i>Adam32</i>	312	0 (0; 0.179)	0 (0; 0)	∞ (0.005; 0)	2.925 (0.013; 0.004)	0 (0; 0)

d_N and d_S estimates are shown in parenthesis in that order

Note: *Ptr* = *Pan troglodytes*, *Hsa* = *Homo sapiens*, *Ggo* = *Gorilla gorilla*, *Pab* = *Pongo abelii*, *Mac* = *Macaca mulatta*. *l* = number of nucleotide sites analyzed

Table 4 Proportion of nonsynonymous to synonymous, as well as nonsynonymous to noncoding substitutions, per site across lineages

Gene	<i>l</i>	<i>Ptr</i>	<i>Hsa</i>	<i>Ggo</i>	<i>Pab</i>	<i>Mac</i>
<i>Adam2</i>	2224	∞ ; 2.000 (0.008; 0.005)	0; 0 (0; 0.008)	∞ ; 0.444 (0.004; 0.008)	0.475; 1.529 (0.026; 0.017)	1.458 ; ∞ (0.061; 0)
<i>Adam12</i>	1715	0; 0 (0; 0.010)	0; 0 (0; 0.005)	0; 0 (0; 0.015)	∞ ; 0.625 (0.005; 0.008)	0.053; ∞ (0.005; 0)
<i>Adam18</i>	3299	∞ ; 1.875 (0.015; 0.008)	0; 0 (0; 0.005)	∞ ; 0.833 (0.005; 0.006)	0.318; 0.938 (0.015; 0.016)	∞ ; 0.368 (0.014; 0.038)
<i>Adam23</i>	1230	0; 0 (0; 0.003)	0; 0 (0; 0.003)	0; 0 (0; 0.003)	∞ ; 0.250 (0.004; 0.016)	∞ ; ∞ (0.106; 0)
<i>Adam28</i>	2636	0; 0 (0; 0.011)	0; 0 (0; 0.05)	0; 0 (0; 0.012)	∞ ; 0.333 (0.007; 0.021)	∞ ; 0.389 (0.021; 0.054)
<i>Adam30</i>	973	0; 0 (0; 0.001)	0; 0 (0; 0.004)	0; 0 (0; 0.010)	∞ ; 0.333 (0.004; 0.011)	19.181; 0.538 (0.028; 0.055)
<i>Adam32</i>	1327	0; 0 (0; 0.002)	0; 0 (0; 0.004)	∞ ; 0.833 (0.005; 0.006)	2.925; 0.929 (0.013; 0.014)	0; 0 (0; 0)

Estimates are bolded when both are higher than one. The proportions of nonsynonymous and noncoding substitutions per site are shown in parenthesis in that order. *l* is the number of noncoding nucleotide sites analyzed

signal is driven by male \times female postcopulatory interactions prior to fertilization. Alternatively, positive selection at adhesion domains in mammals could be driven by ancient bouts of selection associated to other aspects of an adhesion protein domain.

The results obtained from the analysis of molecular evolution of adhesion ADAM domains in primates using both the branch model and the branch-site model allowed us to identify signals of selection (*Adam18* and *Adam23*) and accelerated d_N/d_S ratios (*Adam2* and *Adam23*) along the two terminal branches leading to promiscuous species. The detection of positive selection for our positive control *Semg2* in the foreground branches is reassuring. However, the use of both the branch and branch-site models required grouping together different tree branches that might have experienced different rates and patterns of evolution. This can lead to a single branch or few branches producing a positive result of selection or rate heterogeneity, particularly if a small number of species are used in the analysis. It is also worth noticing that if only some lineages have experienced positive selection, like in the case of *Semg1* (Kingan et al. 2003), then the signal might not be detected using a phylogenetic approach. Clearly, lineage-specific sexual selection bouts occur in sperm genes of primates and caution should be exercised in extrapolating conclusions about the role of sexual selection along specific lineages from broad phylogenetic tests of selection. In fact the estimates of d_N/d_S ratios along each branch show accelerated evolution for *Adam2* and *Adam18* along both chimpanzee and macaque, while restricted to macaques in *Adam23*. It is unclear why *Adam23* shows accelerated evolution in macaque and such a strong signal of positive selection when using the branch-site model. ADAM23 has an active adhesive motif and the gene is expressed in the brain (Evans 2001), so given that *Adam23* has no known function or expression in sperm, the signal along the macaque branch is more likely unrelated to species mating habits. Particularly interesting is that contrary to *Adams 2* and 18, *Adam23* shows no nonsynonymous substitutions along the promiscuous chimpanzee branch, further suggesting that the signal picked for *Adam23* is not related to a dispersed or promiscuous mating system.

Our analysis rules out the sperm surface ADAM32 protein as being driven by selection linked to species differences in primates mating system. The finding of accelerated evolution and higher proportion of nonsynonymous relative to both synonymous and noncoding substitutions in *Adam2* and *Adam18* in primates more likely experiencing postcopulatory reproductive challenges is quite interesting in light of what we know about these two proteins. Contrary to other sperm proteins that have become non-functional in some primates (e.g., ADAM1, ADAM3, and ADAM5) (Jury et al. 1997, 1998; Frayne and Hall 1998;

Frayne et al. 1999) both ADAMs 2 and 18 are functional and localized in the sperm of primates as diverged as humans and macaques (Frayne et al. 2002). Besides being expressed in the sperm surface, primate ADAM2 and ADAM18 also show the same developmental expression and contain an identical putative integrin-binding motif (Frayne et al. 2002). Moreover, both genes are located in adjacent genome positions in primates. In humans, *Adams* 2 and 18 are adjacent to each other (13.7 kb apart) in chromosome 8p11.22. The two genes have also an adjacent location in chromosome 8 in chimpanzee (13.8 kb apart), and the rhesus monkey (20.9 kb apart). It is possible that, in primates, *Adam18* might perform complementary or similar postcopulatory roles to *Adam2* and that not only their likely functional similarities but also their close chromosome position might explain why both genes show similar patterns of evolution.

Acknowledgment This work was supported by funds from an individual Discovery grant (NSERC) to Alberto Civetta.

References

- Blobel CP, Wolfsberg TG, Turck CW, Myles DG, Primakoff P, White JM (1992) A potential fusion peptide and an integrin ligand domain in a protein active in sperm–egg fusion. *Nature* 356:248–252
- Cho C, Bunch DO, Faure JE, Goulding EH, Eddy EM, Primakoff P, Myles DG (1998) Fertilization defects in sperm from mice lacking fertilin beta. *Science* 281:1857–1859
- Cho C, Ge H, Branciforte D, Primakoff P, Myles DG (2000) Analysis of mouse fertilin in wild-type and fertilin beta(–/–) sperm: evidence for C-terminal modification, alpha/beta dimerization, and lack of essential role of fertilin alpha in sperm–egg fusion. *Dev Biol* 222:289–295
- Civetta A (2003a) Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. *Genome* 46:925–929
- Civetta A (2003b) Positive selection within sperm–egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. *Mol Biol Evol* 20:21–29
- Civetta A, Singh RS (1998) Sex-related genes, directional sexual selection, and speciation. *Mol Biol Evol* 15:901–909
- Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11–22
- Dean MD, Good JM, Nachman MW (2008) Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. *Mol Biol Evol* 25:383–392
- Dixson AF (1998) Primate sexuality. Oxford University Press, Oxford
- Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet* 36:1326–1329
- Dorus S, Wasbrough ER, Busby J, Wilkins EC, Karr TL (2010) Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. *Mol Biol Evol* 27:1235–1246
- Edwards DR, Handsley MM, Pennington CJ (2008) The ADAM metalloproteinases. *Mol Aspects Med* 29:258–289
- Evans JP (2001) Fertilin β and other ADAMs as integrin ligands: insights into cell adhesion and fertilization. *Bioessays* 23:628–639
- Evans JP (2002) The molecular basis of sperm–oocyte membrane interactions during mammalian fertilization. *Hum Reprod Update* 8:297–311
- Felsenstein J (1985) Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* 39:783–791
- Frayne J, Hall L (1998) The gene for the human tMDC I sperm surface protein is non-functional: implications for its proposed role in mammalian sperm–egg recognition. *Biochem J* 334:171–176
- Frayne J, Dimsey EA, Jury JA, Hall L (1999) Transcripts encoding the sperm surface protein tMDC II are non-functional in the human. *Biochem J* 341:771–775
- Frayne J, Hurd EAC, Hall L (2002) Human tMDC III: a sperm protein with a potential role in oocyte recognition. *Mol Hum Reprod* 8:817–822
- Glassey B, Civetta A (2004) Positive selection at reproductive ADAM genes with potential intercellular binding activity. *Mol Biol Evol* 21:851–859
- Good JM, Nachman MW (2005) Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol Biol Evol* 22:1044–1052
- Han C, Choi E, Park I, Lee B, Jin S, do Kim H, Nishimura H, Cho C (2009) Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex required for fertilization in mice. *Biol Reprod* 80:1001–1008
- Hurle B, Swanson W, NISC Comparative Sequencing Program, Green ED (2007) Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Res* 17:276–286
- Huxley-Jones J, Clarke TK, Beck C, Toubaris G, Robertson DL, Boot-Handford RP (2007) The evolution of the vertebrate metzincins; insights from *Ciona intestinalis* and *Danio rerio*. *BMC Evol Biol* 7:63
- Jury JA, Frayne J, Hall L (1997) The human *fertilin alpha* gene is non-functional: implications for its proposed role in fertilization. *Biochem J* 321:577–581
- Jury JA, Frayne J, Hall L (1998) Sequence analysis of a variety of primate *fertilin alpha* genes: evidence for non-functional genes in the gorilla and man. *Mol Reprod Dev* 51:92–97
- Kim T, Oh J, Woo JM, Choi E, Im SH, Yoo YJ, Kim DH, Nishimura H, Cho C (2006) Expression and relationship of male reproductive ADAMs in mouse. *Biol Reprod* 74:744–750
- Kingan SB, Tatar M, Rand DM (2003) Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J Mol Evol* 57:159–169
- Morgenstern B, Dress A, Werner T (1996) Multiple DNA and protein sequence alignment based on segment-to-segment comparison. *Proc Natl Acad Sci USA* 93:12098–12103
- Nishimura H, Kim E, Nakanishi T, Baba T (2004) Possible function of the ADAM1a/ADAM2 fertilin complex in the appearance of ADAM3 on the sperm surface. *J Biol Chem* 279:34957–34962
- Nishimura H, Myles DG, Primakoff P (2007) Identification of an ADAM2–ADAM3 complex on the surface of mouse testicular germ cells and cauda epididymal sperm. *J Biol Chem* 282:17900–17907
- O’Connor TD, Mundy NI (2009) Genotype–phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics* 25:i94–i100
- Oh J, Woo JM, Choi E, Kim T, Cho BN, Park ZY, Kim YC, Kim DH, Cho C (2005) Molecular, biochemical, and cellular characterization of epididymal ADAMs, ADAM7 and ADAM28. *Biochem Biophys Res Commun* 331:1374–1383

- Panhuis TM, Clark NL, Swanson WJ (2006) Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philos Trans R Soc Lond B Biol Sci* 361:261–268
- Primakoff P, Myles DG (2000) The ADAM gene family: surface proteins with adhesion and protease activity. *Trends Genet* 16:83–87
- Primakoff P, Myles DG (2007) Cell–cell membrane fusion during mammalian fertilization. *FEBS Lett* 581:2174–2180
- Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD (2008) Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol Biol Evol* 25:207–219
- Shamsadin R, Adham IM, Nayernia K, Heinlein UAO, Oberwinkler H, Engel W (1999) Male mice deficient for germ–cell cyritestin are infertile. *Biol Reprod* 61:1445–1451
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612
- 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
- Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Torgerson DG, Kulathinal RJ, Singh RS (2002) Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol Biol Evol* 19:1973–1980
- Wolfsberg TG, Primakoff P, Myles DG, White JM (1995) ADAM, a novel family of membrane proteins containing a disintegrin and metalloprotease domain: multipotential functions in cell–cell and cell–matrix interactions. *J Cell Biol* 131:275–278
- Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051
- Yamaguchi R, Muro Y, Isotani A, Tokuhiko K, Takumi K, Adham I, Ikawa M, Okabe M (2009) Disruption of ADAM3 impairs the migration of sperm into oviduct in mouse. *Biol Reprod* 81:142–146
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917
- Yang Z, Goldman N, Friday A (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol Biol Evol* 11:316–324
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479
- Zhu GZ, Gupta S, Myles DG, Primakoff P (2009) Testase 1 (ADAM 24) a sperm surface metalloprotease is required for normal fertility in mice. *Mol Reprod Dev* 76:1106–1114