RESEARCH REPORTS
Biological

J.-Y. Sire[1]*, Y. Huang[2,3], W. Li[2], S. Delgado[1], M. Goldberg[4], and P.K. Den Besten[2]

[1]Evolution & Développement du squelette, UMR 7138, Université Pierre et Marie Curie, 7 Quai Saint-Bernard, Bat A2, Case 5, 75005 Paris, France; [2]Department of Orofacial Sciences, University of California at San Francisco, San Francisco, CA, USA; [3]Guanghua School of Stomatology, Sun Yat-sen University, Guangzhou, Guangdong, P.R. China; and [4]Laboratoire Différenciation de Cellules Souches et Prions, U747, Université Paris Descartes, Paris, France; *corresponding author, jean-yves.sire@upmc.fr

Evolutionary Story of Mammalian-specific Amelogenin Exons 4, "4b", 8, and 9

J Dent Res DOI: 10.1177/0022034511423399

APPENDICES

APPENDIX 1

Common names, Latin names, reference to genome assembly, and accession numbers of *AMEL* sequences of the 39 tetrapod species used in this study.

MAMMALIA (37 SPECIES)

Primates (11)

Human, *Homo sapiens*, GRC h37, NM_182680 (*AMELX*), NM_001143 (*AMELY*) - Chimpanzee, *Pan troglodytes*, CHIMP2.1, EF537869 - Gorilla, *Gorilla gorilla*, gorGOR3, present study - Orangutan, *Pongo pygmaeus*, PPYG2, EF537870 - Gibbon, *Nomascus leucogenys,* present study - Rhesus monkey, *Macaca mulatta,* MMUL1.0, EF537871 - Baboon, *Papio hamadryas,* Pham, present study - Marmoset, *Callithrix jacchus,* Callithrix_jacchus-3.2, EF537872 - Bushbaby, *Otolemur garnettii,* otoGar1, AB091787 - Mouse lemur, *Microcebus murinus,* micMur1, present study - Tarsier, *Tarsius syrichta,* tarSyr1, EF537873.

Scandentia (1)

Tree shrew, *Tupaia belangeri,* tupBel1, EU168855.

Rodentia (6)

Mouse, *Mus musculus*, NCB1 m37, NM_009666 - Rat, *Rattus norvegicus,* RGSC 3.4, U51195 - Deer mouse, *Peromyscus maniculatus,* present study - Kangaroo rat, *Dipodomys ordii*, dipOrd1, present study - Guinea pig, *Cavia porcellus,* cavPor3, AJ012200 - Squirrel, *Spermophilus tridecemlineatus,* speTri1, EU168861.

Leurasiatheria (11)

Cow, *Bos taurus*, Btau_4.0, EU168863 - Alpaca, *Vicugna pacos,* vicPac1, present study - Pig, *Sus scrofa,* Sscrofa9, U43405 - Dolphin, *Tursiops truncatus, turTru1,* AH014446 - Horse, *Equus caballus,* EquCab2, AB032193 - Dog, *Canis lupus familiaris,* CanFam 2.0, EU168873 - Cat, *Felis catus,* CAT, EU168880 - Macrobat, *Pteropus vampyrus,* pteVam1, EU168886 - Microbat, *Myotis lucifugus,* myoLuc1, EU168887 - Hedgehog, *Erinaceus europaeus,* eriEur1, EU168884 - Shrew, *Sorex araneus*, sorAra1, EU168888.

Afrotheria (8)

Armadillo, *Dasypus novemcinctus*, dasNov2, present study - Sloth, *Choloepus hoffmanni,* choHof1, present study - Elephant, *Loxodonta africana,* Loxafr 3.0, AY788990 - Tenrec, *Echinops telfairi,* TENREC, EU168892 - Hyrax, *Procavia capensis,* proCap1, EU168895 - Opossum, *Monodelphis domestica,* monDom5, U43407 - Wallaby, *Macropus eugenii,* Meug_1.0, present study - Platypus, *Ornithorhynchus anatinus,* OANA5, AF095566.

SAUROPSIDA

Anole, *Anolis carolinensis,* AnoCar1, present study.

AMPHIBIA

Western clawed toad, *Xenopus tropicalis,* JGI 4.1, NM_0011 13681.

APPENDIX 2

Description of the Hyphy Method

In the analyses, four values were computed for every variable site: observed and normalized expected numbers of synonymous (N_S and E_S) and non-synonymous (N_N and E_N) substitutions.

Hyphy (for HYpothesis testing using PHYlogenies) software (http://hyphy.org; Pond *et al*., 2005), or its improved online version, SLAC (for Single Likelihood Ancestor Counting; http://www.datamonkey.org), estimates $d_N = N_N/E_N$ and $d_S = N_S/E_S$. When $d_N > d_S$, a codon is considered positively selected. When $d_N < d_S$, a codon is considered negatively selected.

A p-value derived from a two-tailed extended binomial distribution was used to assess significance. It is worth noting

that the extended binomial distribution is an approximation of the true distribution of non-synonymous and synonymous under the hypothesis of neutrality (Pond *et al*., 2005). The model assumes that, under neutrality, a random substitution will be synonymous with probability $P = E_S/(E_S + E_N)$, and computes how likely P is when N_S, out of $N_N + N_S$ substitutions, are synonymous. SLAC uses a p-value of the extended binomial distribution that is different from the p-values that derive from a simulation of the null distribution (*i.e*., $d_N = d_S$). The parameter chosen for significance level was 0.2. Indeed, given our dataset of sequences, such a p-value is considered appropriate to detect true positives in datasets containing more than 30 sequences, and SLAC is one of the only methods that allow a high p-value (Pond *et al.,* 2005).

Hyphy Results

Codon	dN-dS	Normalized dN-dS	p-value
4	0.671437	0.160971	0.152221
8	1.14821	0.275273	0.118766
9	0.717396	0.171989	0.177939
14	0.837361	0.200749	0.131821
15	0.772594	0.185222	0.158807

APPENDIX REFERENCE

Pond SLK, Frost SD, Muse SV (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676-679.

APPENDIX 3

	Forward Primers		Reversed Primers		PCR Products (bp)	
Primer Pairs	Sequences	Locations	Sequences	Locations	gDNA	cDNA
M1	CCGAAGTGGATACTTTGGTTG	exon 7	ATGCCTTGTCATGCCCTGGTA	exon 8	2132	not tested
M2	ATCATCCCTGTGCTGTCTC	exon 6			3993	413
M3	CTCTGCCTCCACTGTTCTC	exon 6	TTGATAGCCTGAGAATGTGAC	exon "4b"	3704	124
M4	ACCACTCCATGACTCCAAC	exon 6			3869	289
H1	CCATGCTTCCTGATCTGACTC	exon 6	TGTAGGAAATTGGTTGAAGTCG	putative exon 9	4212	114
H2	AAATCATCCCCGTGCTGTC	exon 6	GTAGGAAATTGGTTGAAGTCGT	putative exon 9	4526	428

Sequences and locations of the primers used for PCR amplifications. The sizes of PCR products for gDNA and cDNA templates are predicted from gDNA and putative cDNA sequences. All primers amplified gDNA templates by PCR and showed correct PCR product sizes.