Additional file 5 – Supplementary methods on PAML analyses of mammalian
Zp3 and Zp2 sequence data from GenBank database

Sequences Retrieved

The Zp3 and Zp2 full mRNA sequences of mammalian species were downloaded
from GenBank database. A total of 15 and 10 mammalian species were retrieved for
Zp3 and Zp2, respectively. The species with GenBank accession numbers are as
follows:

Zp3: *Mus musculus* (NM_011776), *Rattus rattus* (Y10823), *Rattus norvegicus*
(NM_053762), *Lagurus lagurus* (AF515621), *Lasiopodomys brandtii* (AF304487),
*Homo sapiens* (NM_001110354), *Macaca radiate* (X82639), *Callithrix sp* (S71825),
*Canis lupus familiaris* (NM_001003224), *Vulpes vulpes* (AY598032), *Felis catus*
(NM_001009330), *Sus scrofa* (NM_213893), *Bos grunniens* (GQ856646), *Bos taurus*
(NM_173974), *Oryctolagus cuniculus* (NM_001195720). The underlined species
were analyzed by Swanson et al. [1].

Zp2: *Mus musculus* (NM_011775), *Rattus norvegicus* (NM_031150), *Homo sapiens*
(NM_003460), *Macaca radiate* (Y10690), *Callithrix jacchus* (Y10767), *Felis catus*
(NM_001009875), *Canis lupus familiaris* (NM_001003304), *Vulpes vulpes*
(AY598031), *Sus scrofa* (NM_213848), *Bos taurus* (NM_173973). The underlined
species were analyzed by Swanson et al. [1].
**Figure S1 - Maximum likelihood trees of Zp3 and Zp2 of mammalian species**

analyzed in this study. (A) Zp3 of 15 mammalian species, under TPM2uf+G model $(\alpha = 0.777)$; (B) Zp3 of the 8 species analyzed in a previous study, under TIM2+G model $(\alpha = 0.662)$; (C) Zp2 of 10 species, under TVM+G model $(\alpha = 1.233)$; (D) Zp2 of the 8 species analyzed in a previous study, under TVM+G model $(\alpha = 1.305)$.

Numbers above nodes indicate bootstrap values with 1000 replications.
**Data Analyses**

Sequence alignment was initially generated for amino acid sequences by translating full coding sequences (CDSs) into protein sequences, and later translated back to DNA sequences. Amino acid sequences were aligned by PRANKSTER (http://www.ebi.ac.uk/goldman-srv/prank/prankster/) – a graphical interface for the multiple sequence alignment program PRANK [2, 3]. The “+F” option was enforced when making alignment. The program PRANK has been demonstrated to outperform a set of published alignment software programs [3, 4].

Phylogenetic trees were reconstructed by maximum likelihood (ML) method in PAUP*4.0b10 [5]. The best-fit nucleotide substitution models were selected by Akaike information criterion (AIC) with jModelTest version 0.1.1 [6]. ML tree search was heuristic, using 100 random addition analyses with TBR branch-swapping. Bootstrap values for nodes were obtained by 1000 replications using heuristic search and 2 random addition analyses with TBR branch-swapping.

Likelihood ratio tests (LRTs) were conducted to detect evidence for positive selection in mammalian Zp3 and Zp2 full coding sequences. The CODEML program in PAML package version 4.4 [7] was implemented to run the site models (M0, M1a, M2a, M7, M8 and M8a). The ML tree topology (Figure S1 as shown above) with branch lengths re-estimated from one-ratio (M0) model was used as user tree for running M1a, M2a, M7, M8 and M8a. The site models allow the $\omega$ ratio (i.e., $\omega = d_N/d_S$ – the nonsynonymous/synonymous rate ratio) to vary among codons [8, 9]. The $\omega$ ratio measures the direction and strength of selection on amino acid changes, with values of $\omega < 1$, = 1, and > 1 indicating purifying (or negative) selection, neutral
evolution, and positive selection, respectively. LRTs compare null models that do not allow for any codons with $\omega>1$ against alternative models that does. Three LRTs (M1a-M2a, M7-M8, and M8a-M8) were carried out to test evidence of positive selection [8-12]. When LRTs show evidence for positive selection, Bayes empirical Bayes (BEB) method can be applied to calculate the posterior probabilities that each codon is from the site class of positive selection under models M2a and M8 [12].

References


