Cell Calcium 53 (2013) 170–179

Contents lists available at SciVerse ScienceDirect

Cell Calcium

journal homepage: www.elsevier.com/locate/ceca

Evolution of the S100 family of calcium sensor proteins

Danna B. Zimmer[a,*], Jeannine O. Eubanks[b], Dhivya Ramakrishnan[b,1], Michael F. Criscitiello[b]

[a] Center for Biomolecular Therapeutics and Department of Biochemistry & Molecular Biology, University of Maryland School of Medicine, 108 North Greene Street, Baltimore, MD 20102, United States

[b] Comparative Immunogenetics Laboratory, Department of Veterinary Pathobiology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, TX 77843-4467, United States

ARTICLE INFO

Article history:
Received 4 October 2012
Received in revised form 1 November 2012
Accepted 3 November 2012
Available online 14 December 2012

Keywords:
Mammals
Phylogenetic analyses
S100 proteins
Evolution
Calcium sensors

ABSTRACT

The S100s are a large group of Ca^{2+} sensors found exclusively in vertebrates. Transcriptomic and genomic data from the major radiations of mammals were used to derive the evolution of the mammalian S100s genes. In human and mouse, S100s and S100 fused-type proteins are in a separate clade from other Ca^{2+} sensor proteins, indicating that an ancient bifurcation between these two gene lineages has occurred. Furthermore, the five genomic loci containing S100 genes have remained largely intact during the past 165 million years since the shared ancestor of egg-laying and placental mammals. Nonetheless, interesting births and deaths of S100 genes have occurred during mammalian evolution. The S100A7 loci exhibited the most plasticity and phylogenetic analyses clarified relationships between the S100A7 proteins encoded in the various mammalian genomes. Phylogenetic analyses also identified four conserved subgroups of S100s that predate the rise of warm-blooded vertebrates: A2/A3/A4/A5/A6, A1/A10/A11/B/P/Z, A13/A14/A16, and A7s/A8/A9/A12/G. The similarity between genomic location and phylogenetic clades suggest that these subfamilies arose by a series of tandem gene duplication events. Examination of annotated S100s in lower vertebrates suggests that the ancestral S100 was a member of the A1/A10/A11/B/P/Z subgroup and arose near the emergence of vertebrates approximately 500 million years ago.

1. Introduction

Ca^{2+} is a ubiquitous secondary messenger that regulates diverse cellular processes in the plant and animal kingdoms. Each cell assembles a unique Ca^{2+} signaling system, which consists of cell surface receptors, channels, pumps/exchangers, and buffering proteins, that tightly controls $[Ca^{2+}]_i$ (intracellular Ca^{2+} levels) [1,2]. Spatiotemporal Ca^{2+} signals are transduced into biological responses by the reversible binding of Ca^{2+} to proteins containing EF-hand Ca^{2+} binding domains [3]. STIM proteins contain two EF-hand domains and fine-tune Ca^{2+} levels in the endoplasmic reticulum lumen [2]. Members of the calmolulin/troponin/S100 superfamily undergo a large conformational change ("Ca^{2+} switch") in response to Ca^{2+} binding that exposes a hydrophobic cleft required for interaction with their cytoplasmic target proteins and subsequent exertion of their biological effects [4]. Calmodulin (CaM), troponin C (TnC), neuronal calcium sensor (NCS) family members, and Ca^{2+} binding protein (CaBP)/calneuron family members contain four EF-hand domains. CaM is the most widely distributed Ca^{2+} sensor protein and is found in fungi, plants, invertebrates and vertebrates. It is also the most highly conserved Ca^{2+} sensor and the CaM amino acid sequence is invariant in vertebrates from bony fish to mammals [5]. The mammalian CaM family consists of three *CALM* genes, which encode identical proteins but differ in their respective 3′ and 5′ untranslated regions. TnC is found in all striated muscles and is encoded by two genes in vertebrate genomes: *TNNC1* is expressed in cardiac and slow-twitch muscle fibers and *TNNC2* is expressed in fast-twitch skeletal muscle fibers [6]. Expression of the NCS gene family is restricted to neurons and retinal photoreceptors. This family arose from a single *Freq* gene with a common ancestor in fungi and has expanded to 14 genes in mammals—a single *NCS1* (*Freq*) gene, five visinin-like-protein (*VILP*) genes, a single recoverin (*RCVN*) gene, three guanylate cyclase activator protein (*GCAP*) genes, and four potassium channel interacting protein (*KChIP*) genes [7–12]. Expression of the CaBP/calneuron family (9 genes) is restricted to vertebrate neurons and photoreceptor cells [11,12]. The CaBP/calneuron genes arose as a group in teleosts with an increased number of splice variants in mammals [11].

Members of the S100 family contain two EF-hand helix-loop-helix Ca^{2+} binding domains. The C-terminal EF-hand contains a

* Corresponding author at: Center for Biomolecular Therapeutics, Institute for Bioscience & Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850, United States. Tel.: +1 240 314 6514.
E-mail address: dzimmer@som.umaryland.edu (D.B. Zimmer).
[1] Current address: Caris Life Sciences, Phoenix, AZ 85040, United States.

0143-4160/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ceca.2012.11.006

12 amino acid Ca^{2+} binding loop and is indistinguishable from EF-hands found in other Ca^{2+} sensors. The N-terminal EF-hand, also referred to as a pseudo or non-canonical EF-hand, contains a 14 amino acid Ca^{2+} binding loop that is unique to S100s [3]. The pseudo/non-canonical EF-hand is postulated to have arisen through gene duplication or exon recombination from a CaM gene with subsequent loss of two EF-hands [13]. The term S100 refers to the solubility of the two founding family members, S100A1 and S100B, in 100% saturated ammonium sulfate and was used in the early literature to denote a mixture of S100A1 and S100B [14]. As new family members were discovered, the S100 nomenclature evolved, giving rise to numerous aliases (http://www.genenames.org/genefamilies/S100).

S100 family members exhibit a high degree of structural similarity, but are not functionally interchangeable. With the exception of S100G, which is monomeric, S100 proteins are typically symmetric dimers [4]. Individual family members exhibit variable affinities for divalent metal ions (Ca^{2+}, Zn^{2+}, Cu^{2+}), oligomerization properties, post-translational modifications, and unique spatial/temporal expression patterns. S100s bind to and regulate a large number of target proteins, some of which are regulated by a single family member and others by multiple family members. Individual family members also have different binding orientations for target proteins that are due in part to differences in surface charge density [15]. The diversity among S100s, the protein targets that they interact with, and their cellular distribution allows cells to transduce a universal Ca^{2+} signal into a unique biological response.

The human genome encodes 21 S100 proteins, four of which are singletons dispersed throughout the genome: *S100B* on chromosome 21, *S100G* on the X chromosome, *S100P* on chromosome 4, and *S100Z* on chromosome 5. Genes for the remaining 17 S100 family members are located in the epidermal differentiation complex (EDC) on human chromosome 1. The EDC consists of over 50 genes within a 2 Mb region that are expressed predominantly in the skin. The EDC is conserved among humans, rodents, marsupials, and birds, but not fishes [16]. This region also encodes the seven S100-fusion type proteins (SFTP), small proline-rich proteins, cornified envelope proteins, involucrin, and loricrin [17]. The SFTPs, trichohyalin (*TCHH*), trichohyalin-like 1 (*TCHHL1*), repetin (*REPN*), hornerin (*HRNR*), fillagrin (*FGL*), fillagrin-2 (*FGL2*), and cornerin (*CRRN*), share the same structural organization at the protein level: a full-length S100 protein domain that is fused in-frame to differing repeat domains [18]. The S100 protein domain reversibly binds Ca^{2+} and the different repeat domains are believed to function as intermediate-filament associated proteins or cornified epithelium proteins. It has been proposed that SFTP genes arose from the fusion of an S100 gene with an epidermal structural gene in the 1q21 region [19,20]. S100s have been identified in mammals, birds, reptiles, amphibians, and fish, but not outside of the vertebrates. They are thought to have arisen 460 million years ago during the Ordovician period long before vertebrates appeared on land [13,21]. Several S100s are unique to fishes and likely arose from gene duplication events that occurred after the ray-finned fish lineage branched from tetrapod lineages [22–25].

This study uses recent whole genome sequencing data from diverse species to examine the molecular evolution of the S100 protein family in mammals. S100s and SFTPs were in a separate clade from other Ca^{2+} sensor proteins indicating an ancient bifurcation between these two gene lineages. Furthermore, the five genomic loci containing S100 genes have remained largely intact in the 165 million years since the shared ancestor of egg-laying and placental mammals. Nonetheless, interesting births and deaths of S100 genes have occurred during mammalian evolution. The S100A7 loci exhibited the most plasticity and phylogenetic analyses clarified relationships between the S100A7 proteins encoded in the various mammalian genomes. Phylogenetic analyses also identified four conserved subgroups that predate the rise of warm-blooded vertebrates: A2/A3/A4/A5/A6, A1/A10/A11/B/P/Z, A13/A14/A16, and A7s/A8/A9/A12/G. The similarity between genomic location and phylogenetic clades suggests that the subgroups arose by tandem gene duplication events. Examination of annotated S100s in lower mammals suggests that the ancestral S100 was a member of the A1/A10/A11/B/P/Z subgroup and arose near the emergence of vertebrates approximately 500 million years ago. Additional genomic and transcriptomic resources will be required to clarify the earliest natural history of this extraordinary family of signaling molecules.

2. Methods

2.1. Identification of S100s

A stepwise approach was used to identify coding sequences for the mammalian S100 genes. Keyword searches collecting all annotated S100s in the National Center for Biotechnological Information (NCBI) and Ensembl databases were followed by BLAST searches of the GenBank nonredundant protein and expressed sequence tag (EST) databases to retrieve other S100 family members using mouse and human sequences as query. For opossum, platypus, and lizard, the Ensembl database was used extensively. Other sequences in these species were obtained from EST searches in NCBI. Out of the four lamprey S100 sequences, three were published [21] but not available in any database, and the fourth sequence was found in the EST searches. The *Xenopus* sequences were obtained by BLAST searches using mouse, human and frog sequences as query as has been described [26].

2.2. Evolutionary distances

Open reading frame and amino acid alignments of S100 homologs were initially made in Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) with ClustalW employing gap opening penalties of 10 and gap extension penalties of 0.1 for pairwise alignments, then 0.2 for multiple alignments and the protein-weighting matrix of Gonnett and Blossum [27,28]. These alignments were then further modified after visual inspection. MEGA5 [29] was used to infer the phylogenetic relationships of S100 and related Ca^{2+} sensors. Distance matrices were computed for nucleotide alignments using the Maximum Composite Likelihood method and for amino acids using a Dayhoff matrix based method [30]. Accession numbers of sequences used in these analyses are provided in Supplemental Tables 1 and 2.

2.3. Phylogenetic analyses

MEGA5 was used to construct phylogenetic trees. All three codon positions of nucleotide data were included in the analyses: 1290 nucleotide positions total for the human Ca^{2+} sensor analysis and 1056 for the mouse Ca^{2+} sensor analysis. Representative genes were included in the mouse and human trees to represent major families and produce the most meaningful phylogeny. For the pan-mammalian S100 analysis, 135 amino acids columns were used. An outgroup was not specified in the pan-mammalian analysis due to the unclear orthology of potential lower vertebrate sequences. The evolutionary histories were inferred using the Neighbor-Joining method as has been previously described [31]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [32]. Trees were viewed using the Treeview software [33].

2.4. Synteny analysis and annotation

Eight species were chosen for exhaustive S100 synteny analysis based on balanced phylogenetic representation of the major mammalian radiations and state of genome projects (GRCh37 of human, GRCm38 of mouse, CanFam3.1 of dog, UMD3.1of cow, dasNov2 of armadillo, loxAfr3 of elephant, BROAD05 of opossum, and OANA5 of platypus). Additional work was performed in Tasmanian devil (DEVIL7.0) and wallaby (Meug_1.0) genomes for S100B. Analysis of genomic synteny was obtained from either the Gene Page or Map Viewer at the NCBI web site, with similar use of Ensembl when necessary. If an S100 gene was not found in one of the eight model species, the genes flanking it in other species were used as search queries in an attempt to find the orthologous locus. Intergenic sequences without an S100 gene present in other species were manually searched.

3. Results and discussion

3.1. Phylogenetic analysis of human and mouse S100s

We began our analysis with the two species for which the most extensive genomic information and characterization was available, mouse and human. While excellent phylogenetic analyses of S100s have been produced based on amino acid sequence [13,23,34,35], nucleotide sequences provide additional data that has proven useful for other families [31]. However, positional homology is best established at the amino acid level due to the evolution of coding DNA as triplets of nucleotides, the degeneracy of the genetic code, and the larger alphabet of proteins that slow sequence similarity degradation and saturation phenomena [36]. Therefore, individual phylogenies for these species were created from amino acid alignments reverted to nucleotides (Fig. 1). Even though the nomenclature has been updated and additional Ca^{2+} sensor families have been included, the resulting trees were very similar to previously published S100 trees [37,38]. Humans have 21 S100 proteins, four of which are not present in the mouse, S100P, S100A7, S100A7L2, and S100A12. In both species, there were two essential groups, one group consisted of S100s and SFTPs and the other group consisted of non-S100 Ca^{2+} sensors. A clade with parvalbumin and oncomodulin was the only inclusive difference in the major groupings between the two species' trees. Additional analyses will be needed to determine if this difference is indicative of a common precursor for S100s/parvalbumin or a reflection of the resolution of the dendograms. Nonetheless, these results indicate that the S100s and SFTPs are more closely related to each other than any other Ca^{2+} sensor

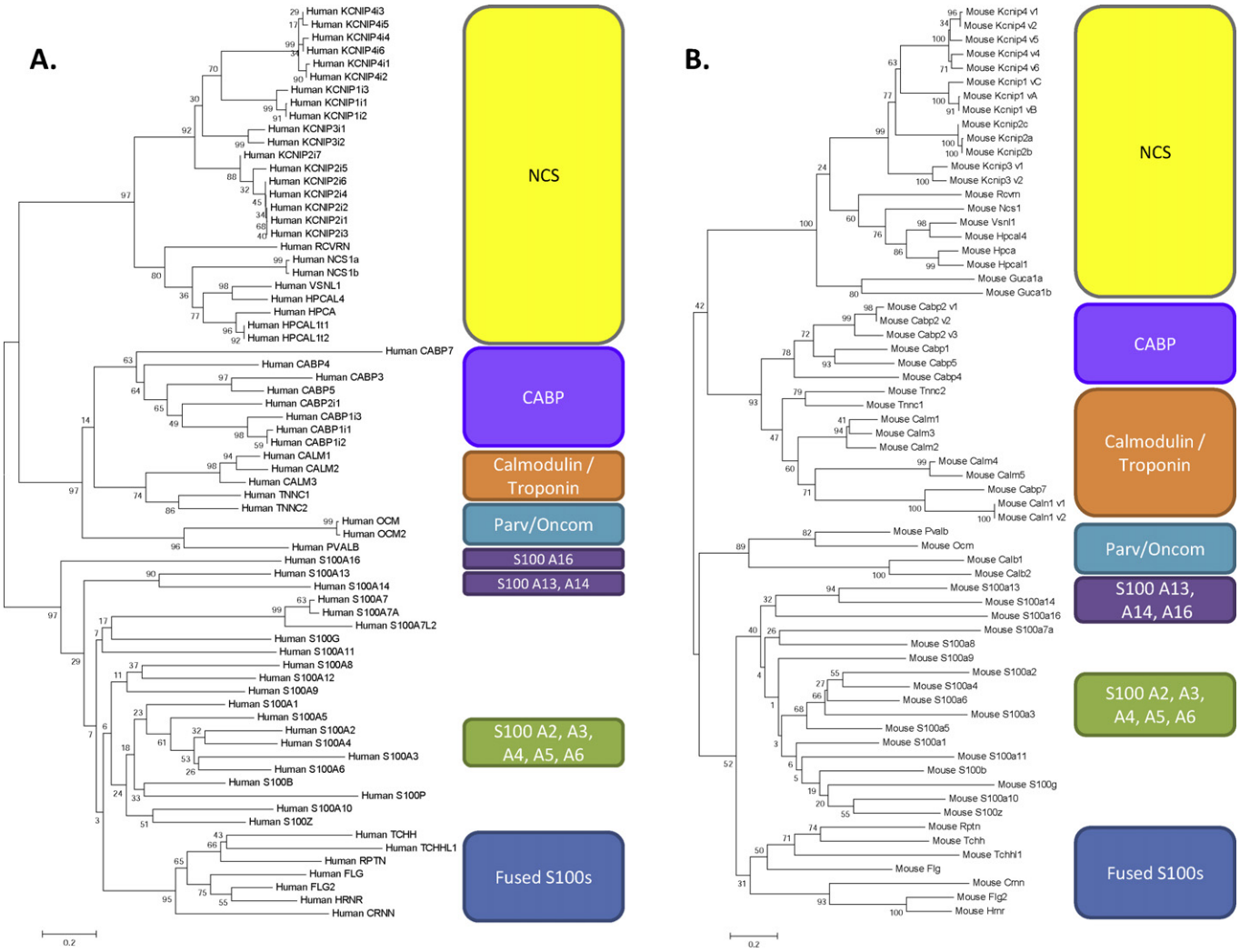


Fig. 1. Ca^{2+} sensor protein phylogenetic trees for human and mouse. Neighbor-joining dendrograms of human (Panel A) and mouse (Panel B) Ca^{2+} sensor nucleotide sequences. Trees are drawn to scale with the branch lengths in the units of base substitutions per site. Colored blocks to the right of the trees highlight groups of sequences that cluster with congruent relationships in the analysis of multiple species. Accession numbers can be found in Supplemental Table 1. Numbers at nodes represent percentage of support from 1000 bootstrap replications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

family in the analysis and are consistent with the hypothesis that the S100 pseudo EF hand evolved after the canonical EF-hand [13].

The individual Ca^{2+} sensor protein families (NSLC, CABP, CaM/TnC and parvalbumin) each fell into their own clade. Furthermore, subgroups/subfamilies emerged within the S100 family: A2/A3/A4/A5/A6 was strongly supported by both trees, and subgroups of A13/A14/A16, A7s/A8/A9/A12/G, and A1/A11/B/P/Z showed progressively weaker conservation in these two species. Members within these subgroups were largely located in juxtaposition to one another in the genome, suggesting that a series of tandem gene duplications contributed to the expansion of the S100 family. The older natural history between these groups was difficult to determine because, although each of the subfamilies forms clades, there was little statistical support from the bootstrap iterations for deeper relationships between the groups.

3.2. Genomic organization

As a first step in establishing the molecular evolution of the S100s, we expanded our analysis to include additional mammalian species. Since S100 expression is regulated at the epigenetic level [39], genomic rather than EST data was used to collect the complement of S100 family members encoded in the following select mammalian genomes (Table 1). The platypus was selected to represent the oldest egg-laying mammals (the monotremes), and the opossum was chosen as the most complete marsupial genome. Molecular phylogenetics suggests two major branches of extant placental mammals, and species from the two major bifurcations of each branch were included [40]. The Atlantagenata yielded Xentharthra (represented here by the armadillo) and Afrotheria (represented by the elephant). The Boreoeutheria included Laurasiatheria (represented by cow and dog) and the Euarchontoglires (represented by the mouse and human). Like the human and mouse, S100 family members in dog, cow, elephant, and opossum are encoded at six regions: S100As in two different regions on a single chromosome and S100B, S100P, S100G, and S100Z on four other chromosomes (Supplemental Figs. 1–6). With only minor exceptions, the genomic organization of S100 family members was highly conserved in mammals. These data are in agreement with previous studies, the only exception being earlier reports that the canine S100A cluster is located on a different chromosome from that of other mammals [41]. While a detailed analysis of the organization of individual S100 genes was beyond the scope of this study, it should be noted that the position of intron/exon boundaries relative to 5′UTRs, initiation codons and coding sequences was quite variable and not as highly conserved as was indicated prior to the completion of the mouse and human genomes [42].

In all mammalian species, S100A11 and S100A10 were located in tandem adjacent to the SFTP genes (Supplemental Fig. 1). Adjacent genes on both sides were conserved as well. However, there were gene insertions between the SFPTs and neighboring genes in both the mouse (7 genes) and human (13 genes) genomes. The mouse genome also contained a four-gene insertion between S100A10 and its conserved neighbor *THEM4*. The opossum genome encoded an additional S100A10 on chromosome 4, referred to as S100A10(2), that was not found in any other species including the Australian marsupials (wallaby or Tasmanian devil). S100A10(2) was not detected in any "higher" mammals or the "lower" platypus by nearest neighbor analysis. The proteins encoded by these loci differed by 17 (out of 97) amino acids. Their pairwise matrix amino acid identity score (0.254) was much lower than scores for other opossum S100s (0.601–3.201), suggesting that S100A10(2) arose from an S100A10 gene duplication/translation event (Fig. 4). Finally, opossum ESTs were not detected for either loci; therefore we could not conclude that these genes were expressed.

The S100A1-A100A16 loci were located on the same chromosome in all species and exhibited several interesting changes (Supplemental Fig. 2). The A1/A13/A14/A16 cluster was intact in all mammalian species for which complete contigs were available. The only change in the A8/A9/A12 cluster was the loss of S100A12 in the mouse. Changes in the A2/A3/A4/A5/A6 cluster included a translocation of S100A5 in the dog, loss of S100A6 in the cow, and loss of S100A5 and S100A2 in the opossum. The largest number of changes occurred in the S100A7 loci. Mouse, dog, and opossum genomes contained a single S100A7 gene (S100A7A) while the human, cow, and elephant genomes contained three S100A7 loci. Subsequent phylogenetic analyses of all mammalian S100s (Section 3.3) revealed that the human locus annotated as S100A7A actually encodes an S100A7-like protein, referred to as S100AL73, suggesting another duplication in this cluster.

The remaining four S100 genomic loci encoding singletons were also highly conserved. The S100B locus was present in all species except the opossum (Supplemental Fig. 3). The platypus represented a more ancient clade but contained S100B, although there has been a block inversion of a set of genes flanking *prmt2* that neighbor S100B. These results suggest that the loss of S100B is confined to the marsupial lineage. The S100B gene was found in two divergent Australian marsupials, although on an orphan contig without the commonly syntenic *prmt2* and *dip2a*. Thus, the loss of S100B may be confined to the American marsupials or even a smaller evolutionary group including the short-tailed opossum. S100G was not present in platypus, but was found in all other species with minor changes in nearest neighbors (Supplementary Fig. 4). The lack of platypus S100G sequences and the strong conservation of syntenic genes on this well assembled portion of the platypus genome suggest that S100G emerged in the marsupials and little has changed since then. The S100P locus had more checkered presentation in the assayed mammalian genomes (Supplementary Fig. 5). It was lost in mouse, cow, and elephant where nearest neighbors were conserved. In contrast to the plasticity of S100P, S100Z is maintained with high fidelity in mammalian genomes. All mammals examined contained an S100Z locus and the location within the genome did not change (Supplementary Fig. 6).

3.3. Mammalian S100 tree

Next we performed phylogenetic analysis of mammalian S100s, excluding other Ca^{2+} sensors, to determine the relationship among various S100 family members (Fig. 2). This analysis confirmed the two subgroups observed in the human and mouse trees (Fig. 1): A2/A3/A4/A5/A6 and A13/A14/A16. In addition, two other subgroups emerged: A7s/A8/A9/A12/G and A1/A10/A11/Z/B/P. Three of these subgroups are clustered within the genome, suggesting that these subfamilies arose by tandem gene duplication events before the rise of mammals. These subgroups may also have structural and functional significance (Fig. 3). Members of the A7/A8/A9/A12 subgroup have demonstrated antimicrobial activity and play a major role in innate immunity [43]. Members of the A2/A3/A4/A5/A6 and A1/A10/A11/Z/B/P subgroups have aromatic residues at the carboxyl termini that participate in target protein interactions [44,45]. In addition, four members of the A2/A3/A4/A5/A6 subgroup have amino acids with large side chains at position 85 and a charged side-chain at position 49 (using S1004 numbering) that impact 3D structure [46]. However, other attributes that contribute to S100 diversity such as affinity for divalent metals, oligomerization properties, post-translational modification(s), surface charge density, target protein profile and spatial/temporal expression patterns do not appear to be subgroup specific. For example, S100B is negatively charged and S100A10 is neutral, but both are in the same subgroup [15]. S100A4 and S100P are in different subgroups, but both regulate myosin filament

assembly [47,48]. Additional comparative studies will be needed to determine if these subgroups have functional as well as evolutional significance.

These analyses also clarified the relationship of the S100A7 genes in mammals. There was support for the relationship between the S100A7A genes, which were adjacent to S100A8, in opossum, elephant, cow, dog, and mouse. However, some members of both major lineages of extant placentals (as evidenced by human, elephant, and cow) maintained a very different cluster with two additional genes, S100A7 and S100A7L, proximal to the A2/A3/A4/A5/A6 cluster. Humans have lost the more ancient S100A7A gene. All S100A7A genes form a clade with the only exception being the human S100A7A gene. Examination of amino acid sequences indicates that human S100A7A locus actually encoded an S100A7-like protein, designated S100A7L3. S100A7L3 had truncated N- and C-termini when compared to S100A7As. In addition, amino acid identity scores from pairwise identity matrices indicate that S100A7L3 was more homologous to S100A7 and S100A7L genes from elephant and cow than S100A7A genes (Supplemental Figs. 7 and 8). The distance between these two lineages of A7 family members was clear by 99% bootstrap analysis support in phylogenetic analysis (Figs. 1 and 2).

Pairwise comparisons of S100 genes from human and opossum, the two most divergent species in the analysis with complete whole genome sequence data for S100 encoding regions, are shown in Fig. 4. The amino acid identity score for S100A12s from human and opossum (1.218) was in the same range as scores for S100A12 comparisons to other human and opossum S100s (0.842–2.642), indicating that these loci are highly divergent. In contrast, S100A6, S100A8 and S100A9 exhibit greater conservation (amino acid identity scores 0.500–0.800). S100A1, S100A3, S100A4, S1005, S100A10, S100A11, S100A16, S100G, S100P, and S100Z were the most highly conserved proteins (amino acid identity scores <0.500).

3.4. Model for evolution of mammalian S100s

The recently concentrated genomic resources in phylogenetically diverse mammalian species allowed us to probe the modern history of the S100 family (Fig. 5). The platypus genome shows that all four groups of S100s were inherited from a reptilian therapsid ancestor of the mammals, although S100G did not translocate to what would become the X allosome until mammals gave live birth (as evidenced in the opossum). Although the S100 genomic cohort was inherited by the earliest mammals, S100A5 did not emerge until the evolution of the placental mammals, and S100A2 was lost at least twice (exemplified by the lineages leading to opossum and cow). The opossum lineage dispensed with S100B, yet evolved an additional S100A10. S100P was found to exhibit the most plasticity in mammals, having been lost at least three times in the lineages leading to elephant, cow, and mouse. The S1007 subgroup birthed two additional S100A7 family members in early placental mammals, only to be lost in dog and mouse. These more recent A7s have expanded in humans to three, and the older S100A7A has been lost in humans. Gaps in current genomic scaffold data left plausible alternative hypotheses that while not favored, cannot currently be negated. These include the possibilities that A1/A7A/A12/A14/A16/P actually arose after the Monotremes in a Therian ancestor and that A2/A3/A5/A7/A7L2/A8/A16 were lost in the Xenartha lineage. But the most parsimonious explanation is that these genes were missing in the developing genome projects in platypus and armadillo, respectively. Physiological and functional studies will be needed to ascertain the impact of S100 gene births and deaths on Ca^{2+} signal transduction. Nonetheless, these findings are consistent with observations that under physiological conditions, the phenotypic effects of deleting or adding S100 genes in mammals are unremarkable, and the current view that S100s provide cell-type specificity and diversity to Ca^{2+} signaling pathways [43,44,49].

3.5. Data mining from non-mammalian vertebrates

To develop a working hypothesis for S100 gene and protein evolution, we identified annotated S100s in lower species using genome information at Ensembl and NCBI as well as the literature. Immunoreactivity was not used since S100 family member antibodies can exhibit cross-reactivity with other family members. Data were grouped by taxa and are presented according to the groups/subfamilies identified by phylogenetic analyses (Table 2). Searches of the comprehensive yeast database, FLYbase, and wormbase confirmed previous reports that the *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* genomes do not encode S100 proteins. In fact, the lowest species containing

Table 1
S100 family members in mammals.

	Human	Mouse	Dog	Cow	Armadillo	Elephant	Opossum	Platypus
S100A1	Y	Y	X	Y	Y	Y	Y	
S100A2	Y	Y	Y	Y		Y	–	Y
S100A3	Y	Y	Y	Y		Y	Y	Y
S100A4	Y	Y	Y	Y	Y	Y	Y	Y
S100A5	Y	Y	Y	Y		Y	–	–
S100A6	Y	Y	Y	–	Y	Y	Y	Y
S100A7A	–	Y	Y	Y		Y	Y	
S100A7	Y	–	–	Y		Y	–	
S100A7L2	Y	–	–	Y		Y	–	
S100A7L3 (S100A7A)	Y	–	–	–		–	–	
S100A8	Y	Y	Y	Y		Y	Y	Y
S100A9	Y	Y	Y	Y	Y	Y	Y	Y
S100A10	Y	Y	Y	Y	Y	Y	Y	
S100A10(2)	–	–	–	–		–	Y	
S100A11	Y	Y	Y	Y	Y	Y	Y	Y
S100A12	Y	–	Y	Y	Y	Y	Y	
S10013	Y	Y	Y	Y	Y	Y	Y	Y
S100A14	Y	Y	Y	Y	Y	Y	Y	
S100A16	Y	Y	Y	Y		Y	Y	
S100B	Y	Y	Y	Y	Y	Y	–	Y
S100G	Y	Y	Y	Y	Y	Y	Y	–
S100P	Y	–	Y	–	Y	–	Y	
S100Z	Y	Y	Y	Y	Y	Y	Y	Y

Y, present in genomic contigs; –, no gene by nearest neighbor and confirmed by EST BLAST; X, loci annotated in this study.

an annotated/published S100 was the lamprey *Petromyzon marinus*. The single lamprey S100 locus in Ensembl was annotated as S100A12 but BLAST searches indicated that this locus encoded S100B. However, it was not possible to confirm that this was S100B by nearest neighbor gene analysis. In fact, lamprey scaffolds were too short to confirm the presence or absence of any other S100s. Of the 10 lamprey S100 cDNAs reported in the literature, three (GENSCAN00000118637, GENESCAN00000111444 and GENESCAN00000006715) were no longer in the databases and one encoded a heat shock protein (EC384389.1) [21,23]. Translated sequences for the remaining six contained EF-hand motifs but none matched existing lamprey scaffolds. Thus, additional genome annotation will be needed before the complement of S100 family members encoded in the lamprey genome can be determined. Nonetheless, the presence of an S100B gene in the lamprey genome supports our hypothesis that the A1/A10/A11/Z/B/P subgroup is the oldest. Only members of the A1/A10/A11/Z/B/P subgroup were annotated in Ensembl for the genomes of bony fishes, coelacanth and amphibians.

We also used these analyses in lower species to determine if the absence of S100G in platypus was a species-specific deletion or site for the emergence of a gene. For example, S100G was not present in the opossum, but there was a single report of an S100G cDNA isolated from chicken (*Gallus gallus*) [50]. The absence of an S100G in birds (*G. gallus* and zebra finch *Taeniopygia guttata*) and reptiles (*Anolis caroleninsis*) was confirmed by nearest neighbor analysis and BLAST searches of translated ESTs using the opossum amino acid sequence as a query. In the case of *Xenopus*, bony fishes (zebrafish *Danio rerio*), coelacanth (*Latimeria chalumnae*), and lamprey, the contigs were too short to confirm the absence of an S100G gene at the genome level. However, no significant matches were obtained with BLAST searches of translated ESTs using the opossum

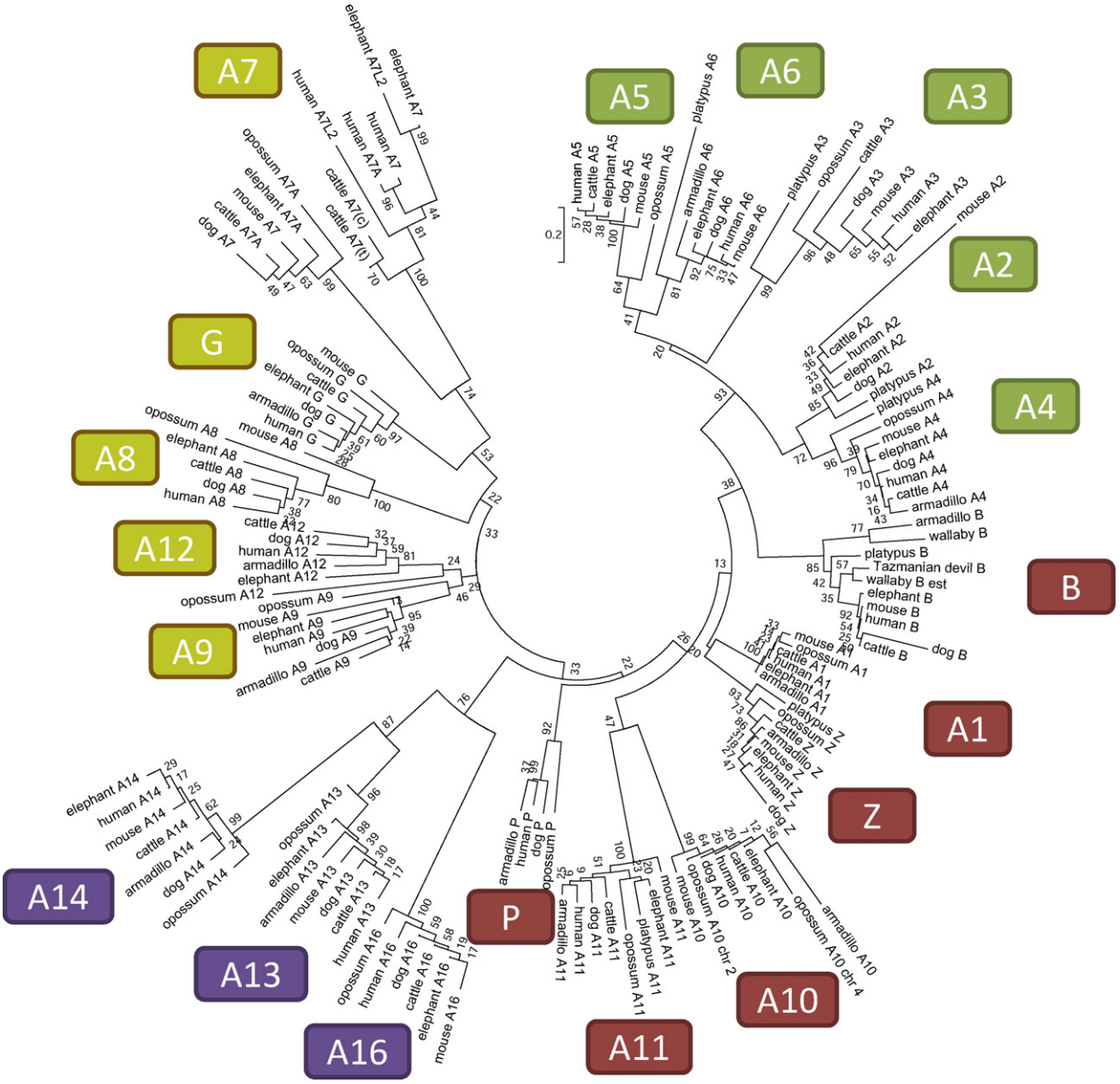


Fig. 2. Evolutionary tree of putative mammalian S100s. Neighbor-joining dendrogram of vertebrate S100s. Using pairwise deletion, 182 alignment positions were used in the analysis. Tree is drawn to scale with the branch lengths in the units of amino acid substitutions per site. Colored blocks to the right of the tree highlight groups of sequences that cluster as monophyletic clades with congruent relationships in the analysis here and in Fig. 1. Accession numbers can be found in Supplemental Table 2. Numbers at nodes represent percentage of support from 1000 bootstrap replications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Amino acid alignment of human S100 proteins. Amino acid sequences of human S100 family members (NCBI) organized according to subgroups identified by genomic organization and phylogenetic analyses. Dashes indicate gaps inserted for aligning EF-hands. The S100 (pseudo) EF-hand Ca^{2+} binding loop is in dark blue, the C-terminal EF-hand Ca^{2+} binding loop in light blue, the linker region connecting the EF-hands in green, alpha-helices in black, non-helical regions in gray and amino acids that may correlate with function domains in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

amino acid sequence as a query. Collectively, these data indicate that the absence of an S100G gene in platypus is not the result of a species-specific deletion.

We also investigated whether S100A1 was present in reptiles, amphibians, and coelacanth. Scaffolds containing the neighboring *chtop* were present in the *Xenopus* and anole genomes, but these scaffolds contained no other annotated genes and appeared too short to encode S100A1. Therefore, we searched EST libraries using the turkey (*Meleagris gallopavo*) S100A1 amino acid sequence as query. When limited to reptiles, three ESTs with *E* values $<1 \times 10^{-26}$ were identified. When limited to *Xenopus*, 12 ESTs with *E* values $<8 \times 10^{-20}$ were obtained. Amino acid alignments of the highest matching *Xenopus* ESTs (JK841104.1) mapped to a scaffold containing *chtop*. The reptilian anole match (DV560051) does not match an existing scaffold. Final verification of this locus will require further development of the genome project. Nonetheless, these data suggest that early vertebrate genomes encoded an S100A1 gene.

3.6. Model for S100 protein family gene evolution

Based on findings from this study, our current working hypothesis for the general evolution of S100 genes in vertebrates is depicted in Fig. 6. We suggest that the S100 A1/A10/A11/B/P/Z subgroup arose near the emergence of vertebrates approximately 500 million years ago. As members of this cluster of S100s are spread among four loci on four different chromosomes in mammals, it is possible that the initial divergence of the A1/A10/A11/B/P/Z subgroup is a product of the two rounds of whole genome duplication that occurred early in vertebrate history [51]. This would likely make S100B, S100P and S100Z paralogs of an S100A gene, perhaps S100A1. The additional full genome duplication of teleosts

Fig. 4. Pairwise identity matrix for human and opossum S100 family members. Number of predicted amino acid substitutions per site using a Dayhoff matrix based model analyzing 146 positions between 38 sequences. Darker pink in the heatmap indicates greater divergence between the two sequences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

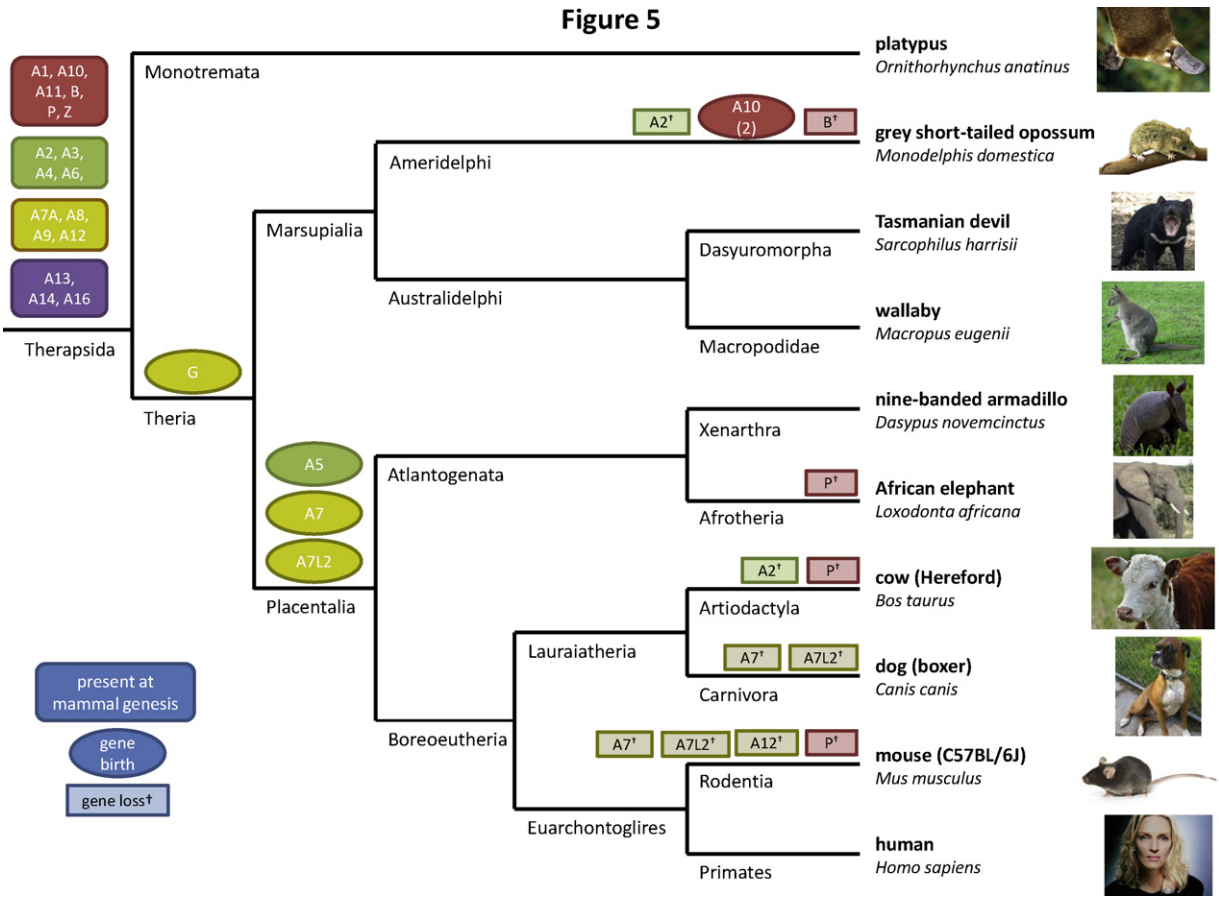


Fig. 5. Model of S100 mammalian evolution. Colored rectangles indicate genes present at mammal genesis, circles denote gene emergence and light rectangles gene loss. The topology of the tree is correct but the distances are not to scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
S100 family members in lower vertebrates.

	Mammals	Birds Chicken Turkey Zebrafinch	Reptiles Lizard Turtle	Amphibians *Xenopus*	Coelacanth	Boney fishes *Tetraodon* Cod Fugu	Lamprey
S100A1	E	E E	X	X	X	E E	
S100A10	E	P E	E E	P X	E		
S100A11	E	E E	E			E E	
S100B	E	E E E	E	–	E		E
S100P	E	E	E	E	E	E	
S100Z	E	E E	E E	E	E	E E P	
S100A13	E	E E					
S100A14	E	E					
S100A16	E	E					
S100A2	E		E				
S100A3	E						
S100A4	E	E E					
S100A5	E		E				
S100A6	E	P E					
S100A7	E						
S100A8	E						
S100A9	E	E E					
S100A12	E						
S100G	E	–	–	–	–	–	–
P26olf				P E			
Dicalcin				P E			

E, Ensembl annotation; P, published; X, annotated in this study; –, no gene or EST identified.

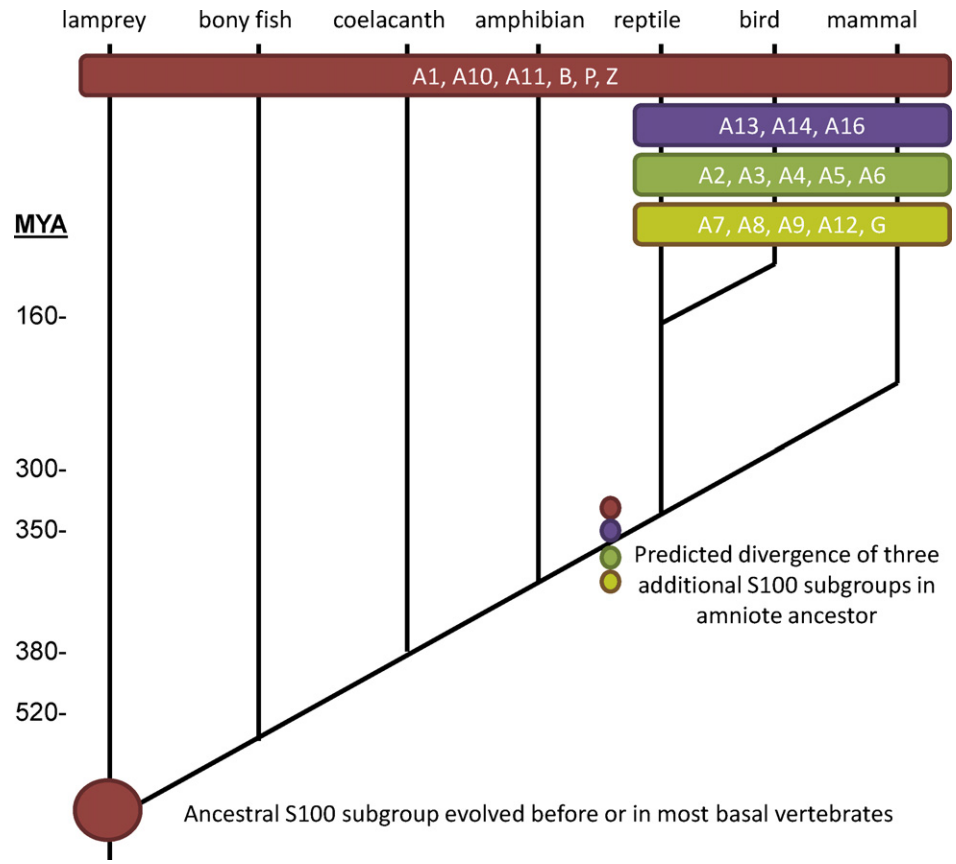


Fig. 6. Model of S100 vertebrate natural history. Hypothesized emergence of the S100 A1/A10/A11/B/P/Z group early in vertebrate evolution, with the relatively recent emergence of the remaining three S100 groups identified in this work (tentatively placed near the common ancestor of the Synapsida that led to mammals and the Sauropsida that led to reptiles and birds). MYA, millions of years ago.

complicates their use in this analysis, and the *Xenopus tropicalis* genome does not suggest the emergence of the three other S100 groups in the earliest tetrapods (amphibians). Resources in reptiles and birds are developing, and these must be used to determine where the large duplicative radiation of the chromosome one locus occurred, birthing the three other structural subgroups (A13/A14/A16, A2/A3/A4/A5/A6, and A7/A8/A9/A12/G) with subsequent translocation of S100G to a fifth S100-encoding chromosome. As genomic and transcriptomic resources continue to develop in cartilaginous fish, non-teleost bony fishes and reptiles, the evolutionary relationship between the S100s and SFTPs can be clarified. Structural and functional analyses of the ancestral S100 gene, once identified, will provide new insights regarding the evolution of Ca^{2+} signaling.

4. Conclusions

To summarize, we have identified four major subgroups of S100 genes based on phylogenetic relationships conserved across diverse species of mammals. One of these four subgroups, A1/A10/A11/B/P/Z, is evolutionarily as old as the vertebrates. The other three are much younger, and likely arose after amphibians, but before the first mammals. The phylogenetic relationships between these four subgroups often concur with genomic synteny, suggesting that each subgroup evolved by tandem duplication. Although there is great conservation of S100 genes among mammals, we observed marked plasticity in the S1007 subfamily and the occurrence of S100P. Understanding the earlier emergence of this fundamental group of Ca^{2+} sensor proteins will require extension of these analyses into lower vertebrates as genome/transcriptome resources become available. Structural and functional analyses of the ancestral S100 gene, once identified, when coupled with physiological studies on the impact of S100 gene births and deaths will provide new insights regarding the mechanisms by which cells transduce a universal Ca^{2+} signal into a cell-specific response that can adapt to environmental changes.

Conflict of interest statement

The authors declare that they have no competing interests.

Author's contributions

D.Z. and M.C. conceived the study, supervised the data collection/analyses and drafted the manuscript. J.E. collected data, prepared alignments, and organized figures. D.R. assisted with data collection/analysis as partial fulfillment of the requirements for a MS degree in the professional Program in Biotechnology. All authors read and approved the final manuscript.

Acknowledgments

We thank S. Dindot and D. Weinstein for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ceca.2012.11.006.

References

[1] M.J. Berridge, Calcium signalling remodelling and disease, Biochemical Society Transactions 40 (2012) 297–309.
[2] J. Soboloff, B.S. Rothberg, M. Madesh, D.L. Gill, STIM proteins: dynamic calcium signal transducers, Nature Reviews Molecular Cell Biology 13 (2012) 549–565.
[3] H. Kawasaki, S. Nakayama, R.H. Kretsinger, Classification and evolution of EF-hand proteins, Biometals 11 (1998) 277–295.
[4] K.L. Yap, J.B. Ames, M.B. Swindells, M. Ikura, Diversity of conformational states and changes within the EF-hand protein superfamily, Proteins 37 (1999) 499–507.
[5] F. Friedberg, A.R. Rhoads, Evolutionary aspects of calmodulin, IUBMB Life 51 (2001) 215–221.
[6] T.E. Gillis, C.R. Marshall, G.F. Tibbits, Functional and evolutionary relationships of troponin C, Physiological Genomics 32 (2007) 16–27.
[7] R.D. Burgoyne, Neuronal calcium sensor proteins: generating diversity in neuronal Ca^{2+} signalling, Nature Reviews Neuroscience 8 (2007) 182–193.
[8] R.D. Burgoyne, J.L. Weiss, The neuronal calcium sensor family of Ca^{2+}-binding proteins, Biochemical Journal 353 (2001) 1–12.
[9] K.-H. Braunewell, The darker side of Ca^{2+} signaling by neuronal Ca^{2+}-sensor proteins: from Alzheimer's disease to cancer, Trends in Pharmacological Sciences 26 (2005) 345–351.
[10] K.-H. Braunewell, A. Szanto, Visinin-like proteins (VSNLs): interaction partners and emerging functions in signal transduction of a subfamily of neuronal Ca^{2+}-sensor proteins, Cell and Tissue Research 335 (2009) 301–316.
[11] H. McCue, L. Haynes, R. Burgoyne, Bioinformatic analysis of CaBP/calneuron proteins reveals a family of highly conserved vertebrate Ca^{2+}-binding proteins, BMC Research Notes 3 (2010) 118.
[12] H.V. McCue, L.P. Haynes, R.D. Burgoyne, The diversity of calcium sensor proteins in the regulation of neuronal function, Cold Spring Harbor Perspectives in Biology 2 (2010) a004085.
[13] T. Ravasi, K. Hsu, J. Goyette, K. Schroder, Z. Yang, F. Rahimi, L.P. Miranda, P.F. Alewood, D.A. Hume, C. Geczy, Probing the S100 protein family through genomic and functional analysis, Genomics 84 (2004) 10–22.
[14] B.W. Moore, A soluble protein characteristic of the nervous system, Biochemical and Biophysical Research Communications 19 (1965) 739–744.
[15] M. Ikura, J.B. Ames, Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality, Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 1159–1164.
[16] K. Magdalini, H. Marcel, H. Daniel, The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the 'fused genes' family, Experimental Dermatology 21 (2012) 643–649.
[17] J. Henry, E. Toulza, C.Y. Hsu, L. Pellerin, S. Balica, J. Mazereeuw-Hautier, C. Paul, G. Serre, N. Jonca, M. Simon, Update on the epidermal differentiation complex, Frontiers in Bioscience 17 (2012) 1517–1532.
[18] S.J. Brown, W.H.I. McLean, One remarkable molecule: filaggrin, Journal of Investigative Dermatology 132 (2012) 751–762.
[19] P. Krieg, M. Schuppler, R. Koesters, A. Mincheva, P. Lichter, F. Marks, Repetin (Rptn), a new member of the "fused gene" subgroup within the S100 gene family encoding a murine epidermal differentiation protein, Genomics 43 (1997) 339–348.
[20] R. Contzler, B. Favre, M. Huber, D. Hohl, Cornulin, a new member of the "fused gene" family, is expressed during epidermal differentiation, Journal of Investigative Dermatology 124 (2005) 990–997.
[21] R.O. Morgan, S. Martin-Almedina, M. Garcia, J. Jhoncon-Kooyip, M.-P. Fernandez, Deciphering function and mechanism of calcium-binding proteins from their evolutionary imprints, Biochimica et Biophysica Acta – Molecular Cell Research 1763 (2006) 1238–1249.
[22] V.G. Fonseca, J. Rosa, V. Laize, P.J. Gavaia, M.L. Cancela, Identification of a new cartilage-specific S100-like protein up-regulated during endo/perichondral mineralization in gilthead seabream, Gene Expression Patterns 11 (2011) 448–455.
[23] A.M. Kraemer, L.R. Saraiva, S.I. Korsching, Structural and functional diversification in the teleost S100 family of calcium-binding proteins, BMC Evolutionary Biology 8 (2008) 48.
[24] C.D. Hsiao, M. Ekker, H.J. Tsai, Skin-specific expression of ictacalcin, a homolog of the S100 genes, during zebrafish embryogenesis, Developmental Dynamics 228 (2003) 745–750.
[25] J. Bobe, F.W. Goetz, A S100 homologue mRNA isolated by differential display PCR is down-regulated in the brook trout (Salvelinus fontinalis) post-ovulatory ovary, Gene 257 (2000) 187–194.
[26] Z.E. Parra, Y. Ohta, M.F. Criscitiello, M.F. Flajnik, R.D. Miller, The dynamic TCRdelta: TCRdelta chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes, European Journal of Immunology 40 (2010) 2319–2329.
[27] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symposium Series 41 (1999) 95–98.
[28] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, Molecular Biology and Evolution 24 (2007) 1596–1599.
[29] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Molecular Biology and Evolution 28 (2011) 2731–2739.
[30] R. Schwarz, M. Dayhoff, Matrices for detecting distant relationships, in: M. Dayhoff (Ed.), Atlas of Protein Sequences, National Biomedical Research Foundation, 1979, pp. 353–358.
[31] M.F. Criscitiello, Y. Ohta, M. Saltis, E.C. McKinney, M.F. Flajnik, Evolutionarily conserved TCR binding sites, identification of T cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark, Journal of Immunology 184 (2010) 6950–6960.
[32] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791.
[33] R.D. Page, Visualizing phylogenetic trees using TreeView, in: A.D. Baxevanis, et al. (Eds.), Current Protocols in Bioinformatics/Editoral Board, John Wiley & Sons, Hoboken, NJ, 2002 (Chapter 6, Unit 6.2).
[34] R.O. Morgan, S. Martin-Almedina, M. Garcia, J. Jhoncon-Kooyip, M.P. Fernandez, Deciphering function and mechanism of calcium-binding proteins from their evolutionary imprints, Biochimica et Biophysica Acta 1763 (2006) 1238–1249.
[35] X. Shang, H. Cheng, R. Zhou, Chromosomal mapping, differential origin and evolution of the S100 gene family, Genetics Selection Evolution 40 (2008) 449–464.
[36] F. Abascal, R. Zardoya, M.J. Telford, TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations, Nucleic Acids Research 38 (2010) W7–W13.
[37] J.K. Kulski, C.P. Lim, D.S. Dunn, M. Bellgard, Genomic and phylogenetic analysis of the S100A7 (Psoriasin) gene duplications within the region of the S100 gene cluster on human chromosome 1q21, Journal of Molecular Evolution 56 (2003) 397–406.
[38] Y. Zhou, W. Yang, M. Kirberger, H.W. Lee, G. Ayalasomayajula, J.J. Yang, Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins, Proteins 65 (2006) 643–655.
[39] W. Lesniak, Epigenetic regulation of S100 protein expression, Clinical Epigenetics 2 (2011) 77–83.
[40] W.J. Murphy, E. Eizirik, S.J. O'Brien, O. Madsen, M. Scally, C.J. Douady, E. Teeling, O.A. Ryder, M.J. Stanhope, W.W. de Jong, M.S. Springer, Resolution of the early placental mammal radiation using Bayesian phylogenetics, Science 294 (2001) 2348–2351.
[41] C. de Guzman Strong, S. Conlan, C.B. Deming, J. Cheng, K.E. Sears, J.A. Segre, A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis, Human Molecular Genetics 19 (2010) 1453–1460.
[42] D.B. Zimmer, J. Chessher, W. Song, Nucleotide homologies in genes encoding members of the S100 protein family, Biochimica et Biophysica Acta 1313 (1996) 229–238.
[43] R. Donato, B.R. Cannon, G. Sorci, F. Riuzzi, K. Hsu, D.J. Weber, C.L. Geczy, Functions of S100 proteins, Current Molecular Medicine (2012) [Epub ahead of print].
[44] D.B. Zimmer, D.J. Weber, The calcium-dependent interaction of S100B with its protein targets, Cardiovascular Psychiatry and Neurology (2010) [2010 Epub].
[45] A. Landar, R.R. Rustandi, D.J. Weber, D.B. Zimmer, S100A1 utilizes different mechanisms for interacting with calcium-dependent and calcium-independent target proteins, Biochemistry 37 (1998) 17429–17438.
[46] K.M. Vallely, R.R. Rustandi, K.C. Ellis, O. Varlamova, A.R. Bresnick, D.J. Weber, Solution structure of human Mts1 (S100A4) as determined by NMR spectroscopy, Biochemistry 41 (2002) 12670–12680.
[47] R.P. House, S.C. Garrett, A.R. Bresnick, Moving aggressively: S100A4 and tumor invasion, in: A. Fatatis (Ed.), Signaling Pathways and Molecular Mediators in Mestastasis, Springer, New York, NY, 2012, pp. 91–114.
[48] M. Du, G. Wang, T.M. Ismail, S. Gross, D.G. Fernig, R. Barraclough, P.S. Rudland, S100p dissociates myosin IIA filaments and focal adhesion sites to reduce cell adhesion and enhance cell migration, Journal of Biological Chemistry 287 (2012) 15330–15344.
[49] D.B. Zimmer, J. Chaplin, A. Baldwin, M. Rast, S100-mediated signal transduction in the nervous system and neurological diseases, Cellular and Molecular Biology (Noisy-le-grand) 51 (2005) 201–214.
[50] S.B. Zanello, R.L. Boland, A.W. Norman, cDNA sequence identity of a vitamin D-dependent calcium-binding protein in the chick to calbindin D-9K, Endocrinology 136 (1995) 2784–2787.
[51] S. Ohno, Evolution by Gene Duplication, Springer-Verlag, Berlin, New York, 1970.

9. SUPPLEMENTAL DATA

Supplemental Table 1. Calcium binding protein gene accession numbers

Species	Family	Member	Accession No.	Method	Database
Mouse	Calmodulin	Calm1	NP_33920	Keyword search	NCBI
Mouse		Calm2	NP_031615	Keyword search	NCBI
Mouse		Calm3	NP_031616	Keyword search	NCBI
Mouse		Calm4	NP_064420	Keyword search	NCBI
Mouse		Cam5	NP_001008706	Keyword search	NCBI
Mouse	Troponin C	Tnnc1	NP_033419	Keyword search	NCBI
Mouse		Tnnc2	NP_033420	Keyword search	NCBI
Mouse	Parvalbumin	Pvalb	NP_038673	Keyword search	NCBI
Mouse		Ocm	NP_149028	Keyword search	NCBI
Mouse	NSLC	Ncs1	NP_062655	Keyword search	NCBI
Mouse		Rcvrn	NP_033064	Keyword search	NCBI
Mouse		Vsnl1	NP_036168	Keyword search	NCBI
Mouse		Hpca	NP_034601	Keyword search	NCBI
Mouse		Hpcal1	NP_057886	Keyword search	NCBI
Mouse		Hpcal4	NP_778163	Keyword search	NCBI
Mouse		Kcnip1	NP_081674	Keyword search	NCBI
Mouse		Kcnip2 isoform a	NP_663749	Keyword search	NCBI
Mouse		Kcnip2 isoform b	NP_109641	Keyword search	NCBI
Mouse		Kcnip2 isoform c	NP_663750	Keyword search	NCBI
Mouse		Kcnip3	NP_062763	Keyword search	NCBI
Mouse		GUCA1A	NP_032215	Keyword search	NCBI
Mouse		GUAC1B	NP_666191	Keyword search	NCBI
Mouse		GUAC2B	NP_032217	Keyword search	NCBI
Mouse		Kcnip4	NP_084541	Keyword search	NCBI
Mouse	Calbindin D28	Calb1	NP_033918	Keyword search	NCBI
Mouse		Calb2	NP_031612	Keyword search	NCBI
Mouse	Fused S100s	TCHH	NP_001156570	Keyword search	NCBI
Mouse		TCHHL1	NP_082038	Keyword search	NCBI
Mouse		Rptn	NP_033126	Keyword search	NCBI
Mouse		FGL	XP_485270	Keyword search	NCBI
Mouse		FGL2	NP_001013826	Keyword search	NCBI
Mouse		HRNR	NP_598456	Keyword search	NCBI
Mouse		Crnn	NP_001074669	Keyword search	NCBI
Mouse	CaBP	Caln1	NP_067346	Keyword search	NCBI
Mouse		Cabp7	NP_620398	Keyword search	NCBI

Mouse		Cabp1	NP_038907	Keyword search	NCBI
Mouse		Cabp2	NP_038906	Keyword search	NCBI
Mouse		Cabp4	NP_653115	Keyword search	NCBI
Mouse		Cabp5	NP_038905	Keyword search	NCBI
Mouse	S100	S100B	ENSMUST00000036387	Keyword Search	Ensembl
Mouse		S100G	ENSMUST00000038769	Keyword Search	Ensembl
Mouse		S100P	ENSMUST00000071949	Keyword Search	Ensembl
Mouse		S100Z	ENSMUST00000022186	Keyword Search	Ensembl
Mouse		S100A1	ENSMUST00000060738	Keyword Search	Ensembl
Mouse		S100A2	ENSMUST00000179550	Keyword Search	Ensembl
Mouse		S100A3	ENSMUST00000001047	Keyword Search	Ensembl
Mouse		S100A4	ENSMUST00000107330	Keyword Search	Ensembl
Mouse		S100A5	ENSMUST00000107329	Keyword Search	Ensembl
Mouse		S100A6	ENSMUST00000001051	Keyword Search	Ensembl
Mouse		S100A7	ENSMUST00000079286	Keyword Search	Ensembl
Mouse		S100A8	ENSMUST00000069927	Keyword Search	Ensembl
Mouse		S100A9	ENST00000368738	Keyword Search	Ensembl
Mouse		S100A10	ENSMUST00000045756	Keyword Search	Ensembl
Mouse		S100A11	ENSMUST00000029515	Keyword Search	Ensembl
Mouse		S100A13	ENSMUST00000048138	Keyword Search	Ensembl
Mouse		S100A14	ENSMUST00000164481	Keyword Search	Ensembl
Mouse		S100A16	ENSMUST00000098911	Keyword Search	Ensembl
Human	Calmodulin	CALM1	NP_008819	Keyword search	NCBI
Human		CALM2	NP_001734	Keyword search	NCBI
Human		CALM3	NP_005175	Keyword search	NCBI
Human	Troponin C	TNNC1	NP_003271	Keyword search	NCBI
Human		TNNC2	NP_003270	Keyword search	NCBI
Human	Parvalbumin	PVALB	NP_002845	Keyword search	NCBI
Human		OCM	NP_001091091	Keyword search	NCBI
Human		OCM2	NP_0016179	Keyword search	NCBI
Human	NSLC	NCS1 isoform 1	NP_055101	Keyword search	NCBI
Human		NCS1 isoform 2	NP_001122298	Keyword search	NCBI
Human		RCVRN	NP_002894	Keyword search	NCBI
Human		VSNL1	NP_003376	Keyword search	NCBI
Human		HPCA	NP_002134	Keyword search	NCBI
Human		HPCAL1 transcript 1	NP_002140	Keyword search	NCBI
Human		HPCAL1 transcript 2	NP_602293	Keyword search	NCBI
Human		HPCAL4	NP_057341	Keyword search	NCBI
Human		KCNIP1 isoform 1	NP_001030009	Keyword search	NCBI
Human		KCNIP1 isoform 2	NP_055407	Keyword search	NCBI

Human		KCNIP1 isoform 3	NP_001030010	Keyword search	NCBI
Human		KCNIP2 isoform 1	NP_055406	Keyword search	NCBI
Human		KCNIP2 isoform 2	NP_775283	Keyword search	NCBI
Human		KCNIP2 isoform 3	NP_775284	Keyword search	NCBI
Human		KCNIP2 isoform 4	NP_775285	Keyword search	NCBI
Human		KCNIP2 isoform 5	NP_775286	Keyword search	NCBI
Human		KCNIP2 isoform 6	NP_775287	Keyword search	NCBI
Human		KCNIP2 isoform 7	NP_775289	Keyword search	NCBI
Human		KNCIP3 isoform 1	NP_038462	Keyword search	NCBI
Human		KNCIP3 isoform 2	NP_001030086	Keyword search	NCBI
Human		KNCIP4 isoform 1	NP_079497	Keyword search	NCBI
Human		KNCIP4 isoform 2	NP_671710	Keyword search	NCBI
Human		KNCIP4 isoform 3	NP_001030176	Keyword search	NCBI
Human		KNCIP4 isoform 3	NP_671711	Keyword search	NCBI
Human		KNCIP4 isoform 4	NP_671712	Keyword search	NCBI
Human		GUCA1A	NP_000400	Keyword search	NCBI
Human		GUAC1B	NP_002089	Keyword search	NCBI
Human		GUAC2B	NP_009033	Keyword search	NCBI
Human		KNCIP4 isoform 5	NP_001030175	Keyword search	NCBI
Human	Calbindin D28	CALB1	NP_004920	Keyword search	NCBI
Human		CALB2 isoform 1	NP_001731	Keyword search	NCBI
Human		CALB2 isoform 20K	NP_009018	Keyword search	NCBI
Human		CALB2 isoform 22K	NP_009019	Keyword search	NCBI
Human	Fused S100s	TCHH	NP_009044	Keyword search	NCBI
Human		TCHHL1	NP_001008536	Keyword search	NCBI
Human		REPN	NP_001116437	Keyword search	NCBI
Human		FGL	NP_002007	Keyword search	NCBI
Human		FGL2	NP_001014364	Keyword search	NCBI
Human		HRNR	NP_001009931	Keyword search	NCBI
Human		CRNN	NP_057274	Keyword search	NCBI
Human		CALN1	NP_113656	Keyword search	NCBI
Human			NP_001017440	Keyword search	NCBI
Human		CABP7	NP_872333	Keyword search	NCBI
Human		CABP1	NP_001028849	Keyword search	NCBI
Human			NP_112482	Keyword search	NCBI
Human			NP_004267	Keyword search	NCBI
Human		CABP2	NP_112481	Keyword search	NCBI
Human			NP_057450	Keyword search	NCBI
Human		CABP3	AAF25798	Keyword search	NCBI

Human		CABP4	NP_660201	Keyword search	NCBI
Human		CABP5	NP_062829	Keyword search	NCBI
Human	S100	S100B	ENST00000291700	Keyword Search	Ensembl
Human		S100G	ENST00000380200	Keyword Search	Ensembl
Human		S100P	ENST00000296370	Keyword Search	Ensembl
Human		S100Z	ENST00000513010	Keyword Search	Ensembl
Human		S100A1	ENST00000292169	Keyword Search	Ensembl
Human		S100A2	ENST00000368708	Keyword Search	Ensembl
Human		S100A3	ENST00000368713	Keyword Search	Ensembl
Human		S100A4	ENST00000368716	Keyword Search	Ensembl
Human		S100A5	ENST00000368718	Keyword Search	Ensembl
Human		S100A6	ENST00000368720	Keyword Search	Ensembl
Human		S100A7A	ENST00000368729	Keyword Search	Ensembl
Human		S1007L2	ENST00000368725	Keyword Search	Ensembl
Human		S100A7	ENST00000368723	Keyword Search	Ensembl
Human		S100A8	ENST00000368733	Keyword Search	Ensembl
Human		S100A9	ENSMUST00000069960	Keyword Search	Ensembl
Human		S100A10	ENST00000368811	Keyword Search	Ensembl
Human		S100A11	ENST00000271638	Keyword Search	Ensembl
Human		S100A12	ENST00000368737	Keyword Search	Ensembl
Human		S100A13	ENST00000339556	Keyword Search	Ensembl
Human		S100A14	ENST00000476873	Keyword Search	Ensembl
Human		S100A16	ENST00000368703	Keyword Search	Ensembl

Supplemental Table 2: Accession Numbers for Mammalian S100s

S100	Species	locus (S100 coordinates)	build	Annoted As	Gene accession
S100B	human	chr 21q22.3 48,018,875-48,025,121	GRCh37		ENSP00000291700
	mouse	chr 10c1 75,716,598-75,723,904	NCBIM37		ENSMUSP00000047968
	dog	chr 31 42,152,420-42,155,916	BROADD2		ENSCAFG00000012228
	cattle	chr 1 148,009,651-148,016,981	UMD3.1		ENSBTAG00000004777
	armadillo	scaf 6436 320,705-323,226	dasNOV2	not	ENSDNOP00000013481
	elephant	scaf 110 3,539,000-3,542,551	loxAfr3		ENSLAFG00000014935
	opossum	chr 2 523,811,717-523,856,003	BROAD05		ENSMODG00000008324
	Tazmanian devil	scaf GL 857102	DEVIL7.0		ENSSHAG00000005156
	wallaby	est			FY475974
	wallaby	scaf 58856 7,785-9,224	Meug_1.0		ENSMEUG00000013166
	wallaby	scaf 6612	Meug_1.0		ENSMEUG00000000150
	platypus	contig 273 280,484-285,865	OANA5	F7DN18	ENSOANG00000013550
S100G	human	chr x 16,668,281-16,672,793	GRCh37		ENSG00000169906
	mouse	chr x 159,399,924-159,402,531	NCBIM37		ENSMUSG00000040808
	dog	chr x 12,764,087-12,768,341	BROADD2		ENSCAFG00000012583
	cattle	chr X: 134101039-134104560	UMD3.1		ENSBTAG00000017020
	armadillo	scaf 307: 156,749-159,701	dasNOV2		ENSDNOG00000009022
	elephant	scaf 39 15,069,030-15,072,124	loxAfr3		ENSLAFG00000016495
	opossum	chr 7 23,039,919-23,043,569	BROAD05		ENSMODG00000017180
	platypus	contig 462 8,635,600-8,673,252	OANA5	F7FDW4	ENSOANG00000002569
S100P	human	chr 4 6,694,796-6,698,897	GRCh37		ENSG00000163993
	mouse	chr 5 37,138,613-37,139,918	NCBIM37		ENSMUSG00000060708
	dog	chr 3 61,624,720-61,627,128	BROADD2		ENSCAFG00000014333
	cattle	chr 6 118,023,637-118,025,125	UMD3.1	QOVCC3	ENSBTAG00000003766
	armadillo	scaf 102319 2,464-5,563	dasNOV2		ENSDNOG00000015334
	elephant	scaf 167 93,494-94,126	loxAfr3		ENSLAFG00000005182
	opossum	chr 5 224,088,517-224,093,675	BROAD05		ENSMODG00000002897
	platypus	contig 29525 9,396-11,190	OANA5	F6SVG7	ENSOANG00000022358
	platypus	contig 4744 9,396-11,190	OANA5	F7BP21	ENSOANG00000003982
S100Z	human	chr 5 76,145,826-76,217,475	GRCh37		ENSG00000171643
	mouse	chr 13 96,247,256-96,248,610	NCBIM37		ENSMUSG00000021679
	dog	chr 3 32,660,581-32,663,688	BROADD2		ENSCAFG00000023143
	cattle	chr 10 7,992,986-7,995,677	UMD3.1		ENSBTAG00000020201
	armadillo	scaf 5094 53,186-56,209	dasNOV2		ENSDNOG00000002472
	elephant	scaf 7 79,997,061-79,999,118	loxAfr3		ENSLAFG00000013035
	opossum	chr 3 40,003,360-40,006,269	BROAD05		ENSMODG00000019747
	platypus	chr 1 15,688,842-15,691,393	OANA5	F7G3H7	ENSOANG00000015763

S100A10	human	chr 1 151,955,391-151,966,866	GRCh37		ENSG00000197747
	mouse	chr 3 93,359,002-93,368,565	NCBIM37		ENSMUSG00000041959
	dog	chr 17 64,061,414-64,072,120	BROADD2		ENSCAFG00000023111
	cattle	chr 3 18,799,612-18,810,545	UMD3.1		ENSBTAG00000015147
	armadillo	scaf 63281 5,204-5,496	dasNOV2		ENSDNOG00000011265
	elephant	scaf 11 40,275,914-40,279,559	loxAfr3		ENSLAFG00000011183
	opossum	chr 2 497,258,786-497,265,212	BROAD05		ENSMODG00000018919
	opossum(2)	chr 4 26,381,268-26,381,561	BROAD05		ENSMODG00000025292
S100A11	human	chr 1 152,004,982-152,020,383	GRCh37		ENSG00000163191
	mouse	chr 3 93,324,410-93,330,209	NCBIM37		ENSMUSG00000027907
	dog	chr 17 64,097,517-64,102,698	BROADD2		ENSCAFG00000012916
	cattle	chr 3 18,768,796-18,770,416	UMD3.1	Q862H7	ENSBTAG00000015145
	armadillo	scaf 205856 486-1,079	dasNOV2		ENSDNOG00000005442
	elephant	scaf 11 40,230,602-40,231,806	loxAfr3		ENSLAFG00000026763
	opossum	chr 2 497,322,873-497,325,570	BROAD05		ENSMODG00000018920
	platypus	contig 12765 26,772-28,669	OANA5	F6S011	ENSOANG00000009693
S100A1	human	chr 1 153,600,402-153,604,513	GRCh37		ENSG00000160678
	mouse	chr 3 90,314,956-90,318,314	NCBIM37		ENSMUSG00000044080
	cattle	chr 3 16,812,602-16,816,588	UMD3.1	S10A1	ENSBTAG00000005163
	armadillo	scaf 6778 51,476-55,790	dasNOV2		ENSDNOG00000004826
	elephant	scaf 33 2,165,397-2,166,706	loxAfr3		ENSLAFG00000002657
	opossum	chr 2 187,950,053-187,951,689	BROAD05		ENSMODG00000017368
S100A2	human	chr 1 153,533,584-153,540,366	GRCh37		ENSG00000196754
	mouse	chr 3 90,394,169-90,394,277 90,395,223-90,395,430	NCBIM37	predicted	NP_001182689
	dog	chr 7 46,457,830-46,458,909	BROADD2		ENSCAFG00000017547
	cattle	chr 3 16,869,280-16,872,613	UMD3.1	S10A2	ENSBTAG00000037651
	elephant	scaf 33 2,105,516-2,107,174	loxAfr3		ENSLAFG00000026541
	platypus	contig 11856 27,759-29,865	OANA5	F6Q7Q8	ENSOANG00000014147
S100A3	human	chr 1 153,519,805-153,521,848	GRCh37		ENSG00000188015
	mouse	chr 3 90,404,137-90,406,624	NCBIM37		ENSMUSG00000001021
	dog	chr 7 46,468,586-46,469,240	BROADD2		ENSCAFG00000017548
	cattle	chr 3 16,880,769-16,884,904	UMD3.1	A4FUH7	ENSBTAG00000039105
	elephant	scaf 33 2,091,966-2,092,762	loxAfr3		ENSLAFG00000028157
	opossum	chr 3 187,871,908-187,874,350	BROAD05		ENSMODG00000017395
	platypus	contig 11856 20,135-21,548	OANA5	F6Q7F6	ENSOANG00000014146
S100A4	human	chr 1 153,516,089-153,522,612	GRCh37		ENSG00000196154
	mouse	chr 3 90,407,692-90,409,967	NCBIM37		ENSMUSG00000001020
	dog	chr 7 46,471,122-46,473,134	BROADD2		ENSCAFG00000017550
	cattle	chr 3 16,887,102-16,888,570	UMD3.1	S10A4	ENSBTAG00000019203
	armadillo	scaf 81406 2,676-3,717	dasNOV2		ENSDNOG00000018653
	elephant	scaf 33 2,088,093-2,089,345	loxAfr3		ENSLAFG00000015008

	opossum	chr 2 187,864,578-187,865,727	BROAD05		ENSMODG00000017397
	platypus	contig 11856 14,931-15,397	OANA5	F6Q7T6	ENSOANG00000014144
S100A5	human	chr 1 153,509,623-153,514,241	GRCh37		ENSG00000196420
	mouse	chr 3 90,412,445-90,415,702	NCBIM37		ENSMUSG00000001023
	dog	chr 7 46,476,026-46,478,009	BROADD2		ENSCAFG00000017552
	cattle	chr 3 16,891,318-16,894,701	UMD3.1	E1B8S0	ENSBTAG00000000644
	elephant	scaf 33 2,082,026-2,084,572	loxAfr3		ENSLAFG00000015005
	opossum	chr 2 187,846,662-187,858,463	BROAD05		ENSMODG00000017400
S100A6	human	chr 1 153,507,075-153,508,720	GRCh37		ENSG00000197956
	mouse	chr 3 90,416,816-90,418,336	NCBIM37		ENSMUSG00000001025
	dog	chr 7 46,478,503-46,480,182	BROADD2		ENSCAFG00000017553
	armadillo	scaf 59060 2,839-3,445	dasNOV2		ENSDNOG00000009294
	elephant	scaf 33 2,079,007-2,079,661	loxAfr3		ENSLAFG00000027113
	platypus	contig11856 10,098-11,426	OANA5	F6R394	ENSOANG00000014143
S100A7	human 7A	chr 1 153,388,945-153,395,701	GRCh37		ENSG00000184330
	human 7L2	chr 1 153,409,471-153,412,503	GRCh37		ENSG00000197364
	human 7	chr 1 153,430,220-153,433,177	GRCh37		ENSG00000143556
	mouse	chr 3 90,458,224-90,462,052	NCBIM37		ENSMUSG00000063767
	dog	chr 7 46,508,505-46,510,721	BROADD2	XM_850076.1	ENSCAFG00000017554
	cattle 7	chr 3 16,967,602-16,971,446	UMD3.1	S10A7	ENSBTAG00000008238
	cattle 7A	chr 3 17,086,462-17,089,235	UMD3.1	S100A7	ENSBTAG00000033007
	cattle 15	chr 3 17,137,845-17,140,530	UMD3.1	E1BPR8	ENSBTAG00000024437
	elephant 7A	scaf 33 1,451,124-1,453,601	loxAfr3	novel	ENSLAFG00000032266
	elephant 7L2	scaf 33 1,819,975-1,822,053	loxAfr3	novel	ENSLAFG00000027467
	elephant 7	scaf 33 1,964,341-1,965,961	loxAfr3	novel	ENSLAFG00000026442
	opossum 7A	chr 2 187,744,792-187,750,533	BROAD05	XM_001372135.1	ENSMODG00000017402
S100A8	human	chr 1 153,362,508-153,363,664	GRCh37		ENSG00000143546
	mouse	chr 3 90,472,993-90,473,956	NCBIM37		ENSMUSG00000056054
	dog	chr 7 46,518,213-46,518,630	BROADD2	COLQLO	ENSCAFG00000017557
	cattle	chr 3 17,146,918-17,147,898	UMD3.1	S10A8	ENSBTAG00000012640
	elephant	scaf 33 1,443,095-1,443,500	loxAfr3		ENSLAFG00000028077
	opossum	chr 2 187,727,621-187,728,606	BROAD05		ENSMODG00000017403
S100A9	human	chr 1 153,330,330-153,333,503	GRCh37		ENSG00000163220
	mouse	chr 3 90,496,554-90,499,643	NCBIM37		ENSMUSG00000056071
	dog	chr 7 46,544,157-46,546,227	BROADD2		ENSCAFG00000017558
	cattle	chr 3 17,176,217-17,179,005	UMD3.1		ENSBTAG00000006505
	armadillo	scaf 12075 46,052-47,949	dasNOV2		ENSDNOG00000016750
	elephant	scaf 33 1,401,492-1,403,766	loxAfr3		ENSLAFG00000030986
	opossum	chr 2 187,669,504-187,671,573	BROAD05		ENSMODG00000017410
S100A12	human	chr 1 153,346,184-153,348,125	GRCh37		ENSG00000163221
	dog	chr 7 46,531,797-46,532,462	BROADD2		ENSCAFG00000023324

	cattle	chr 3 17,163,820-17,165,262	UMD3.1	S10AC	ENSBTAG00000012638
	armadillo	scaf 4927 53,473-54,247	dasNOV2		ENSDNOG00000016751
	elephant	scaf 33 1,427,267-1,428,062	loxAfr3		ENSLAFG00000000593
	opossum	chr 2 187,688,712-187,690,101	BROAD05		ENSMODG00000017406
S100A13	human	chr 1 153,591,263-153,606,873	GRCh37		ENSG00000189171
	mouse	chr 3 90,318,357-90,328,503	NCBIM37		ENSMUSG00000042312
	dog	chr 7 46,410,463-46,419,158	BROADD2		ENSCAFG00000017542
	cattle	chr 3 16,818,414-16,824,125	UMD3.1	S10AD	ENSBTAG00000021378
	armadillo	scaf 6778 37,104-50,040	dasNOV2		ENSDNOG00000004825
	elephant	scaf 33 2,151,146-2,161,331	loxAfr3		ENSLAFG00000026554
	opossum	chr 2 187,934,436-187,945,057	BROAD05		ENSMODG00000017387
S100A14	human	chr 1 153,586,731-153,589,462	GRCh37		ENSG00000189334
	mouse	chr 3 90,330,778-90,332,755	NCBIM37		ENSMUSG00000042306
	dog	chr 7 46,422,006-46,422,967	BROADD2		ENSCAFG00000017544
	cattle	chr 3 16,827,337-16,829,392	UMD3.1	S10AE	ENSBTAG00000021377
	armadillo	scaf 6778 33,004-33,976	dasNOV2		ENSDNOG00000004820
	elephant	scaf 33 2,146,590-2,147,584	loxAfr3		ENSLAFG00000011995
	opossum	chr 2 187,929,738-187,931,292	BROAD05		ENSMODG00000017390
S100A16	human	chr 1 153,579,362-153,585,644	GRCh37		ENSG00000188643
	mouse	chr 3 90,341,176-90,347,073	NCBIM37		ENSMUSG00000074457
	dog	chr 7 46,424,683-46,429,683	BROADD2		ENSCAFG00000017545
	cattle	chr 3 16,830,511-16,837,422	UMD3.1	S10AG	ENSBTAG00000014204
	elephant	scaf 33 2,139,379-2,140,033	loxAfr3		ENSLAFG00000002656
	opossum	chr 2 187,918,881-187,919,666	BROAD05		ENSMODG00000017391

Supplemental Figure 1: S100A10-A11 and SFTP protein locus synteny. S100 genes are shown in red and S100 fused-type proteins (SFTP) in pink. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. With the exception of gene insertions in the human and mouse indicated in white, black lines denote intergenic space where no other genes were detected. Orthologs between species are vertically aligned and distances are not to scale.

Supplemental Figure 2: S100A1-A9, A12-A14 and A16 locus synteny. S100 genes are shown in red, purple, green and yellow. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. Black lines denote intergenic space where no other genes were detected. Orthologs between species are vertically aligned and distances are not to scale.

Supplemental Figure 3: S100B locus synteny. S100B genes are shown in red. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. Black lines denote intergenic space where no other genes were detected. Dashed lines highlight syntenic blocks that have been inverted or translocated. Orthologs between species are vertically aligned and distances are not to scale.

Supplemental Figure 4: S100G locus synteny. S100G genes are shown in yellow. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. Black lines denote intergenic space where no other genes were detected. Dashed lines highlight syntenic blocks that have been inverted or translocated. Orthologs between species are vertically aligned and distances are not to scale.

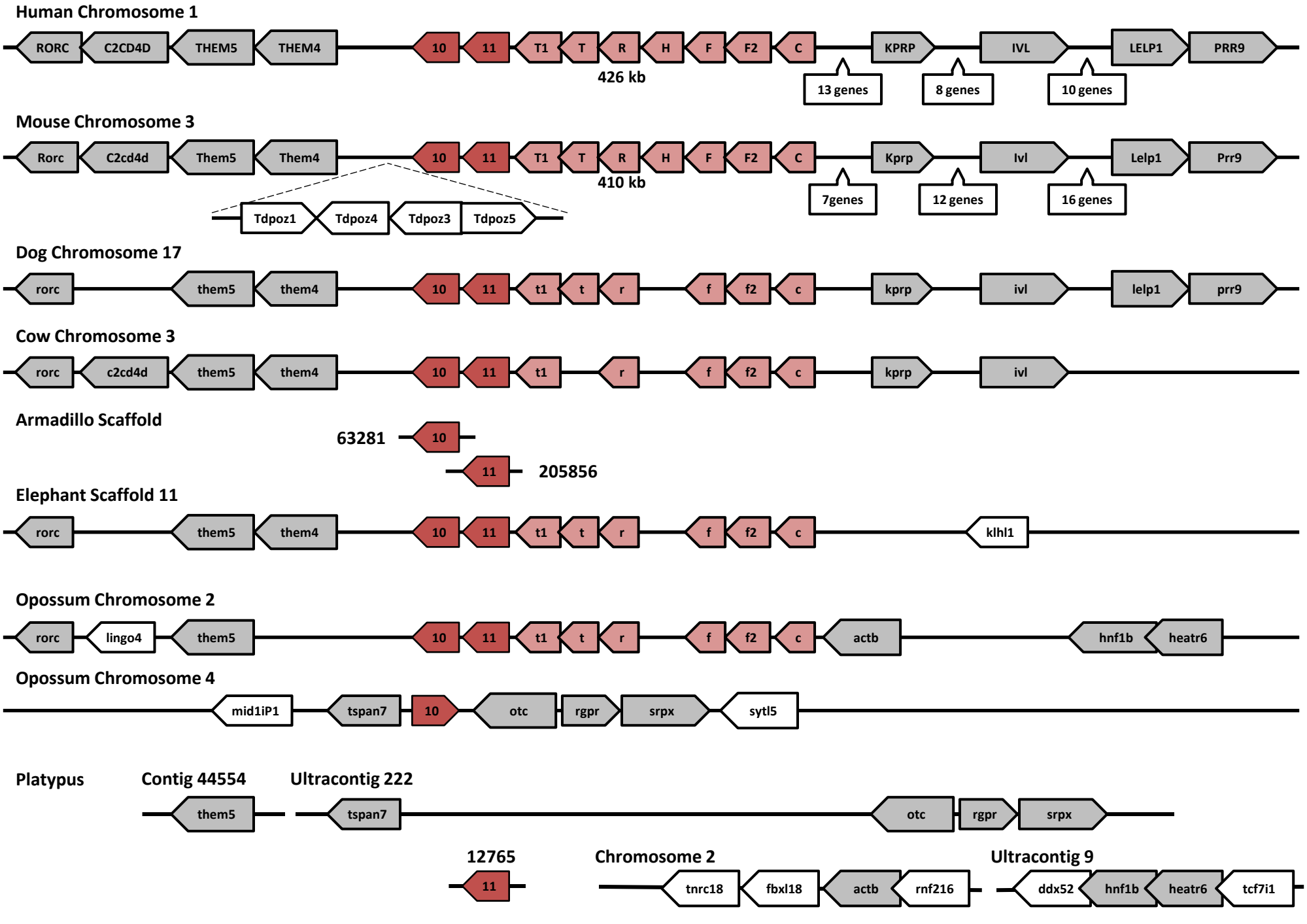
Supplemental Figure 5: S100P locus synteny. S100P genes are shown in red. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. Black lines denote intergenic space where no other genes were detected. Orthologs between species are vertically aligned and distances are not to scale.

Supplemental Figure 6: S100Z locus synteny. S100Z genes are shown in red. Syntenic genes found in at least two mammalian genomes are shown in grey. Gene polygons point in transcriptional orientation. Black lines denote intergenic space where no other genes were detected, except where blocks of genes inserted below. Dashed arrow highlight a syntenic gene that has been translocated. Orthologs between species are vertically aligned and distances are not to scale.

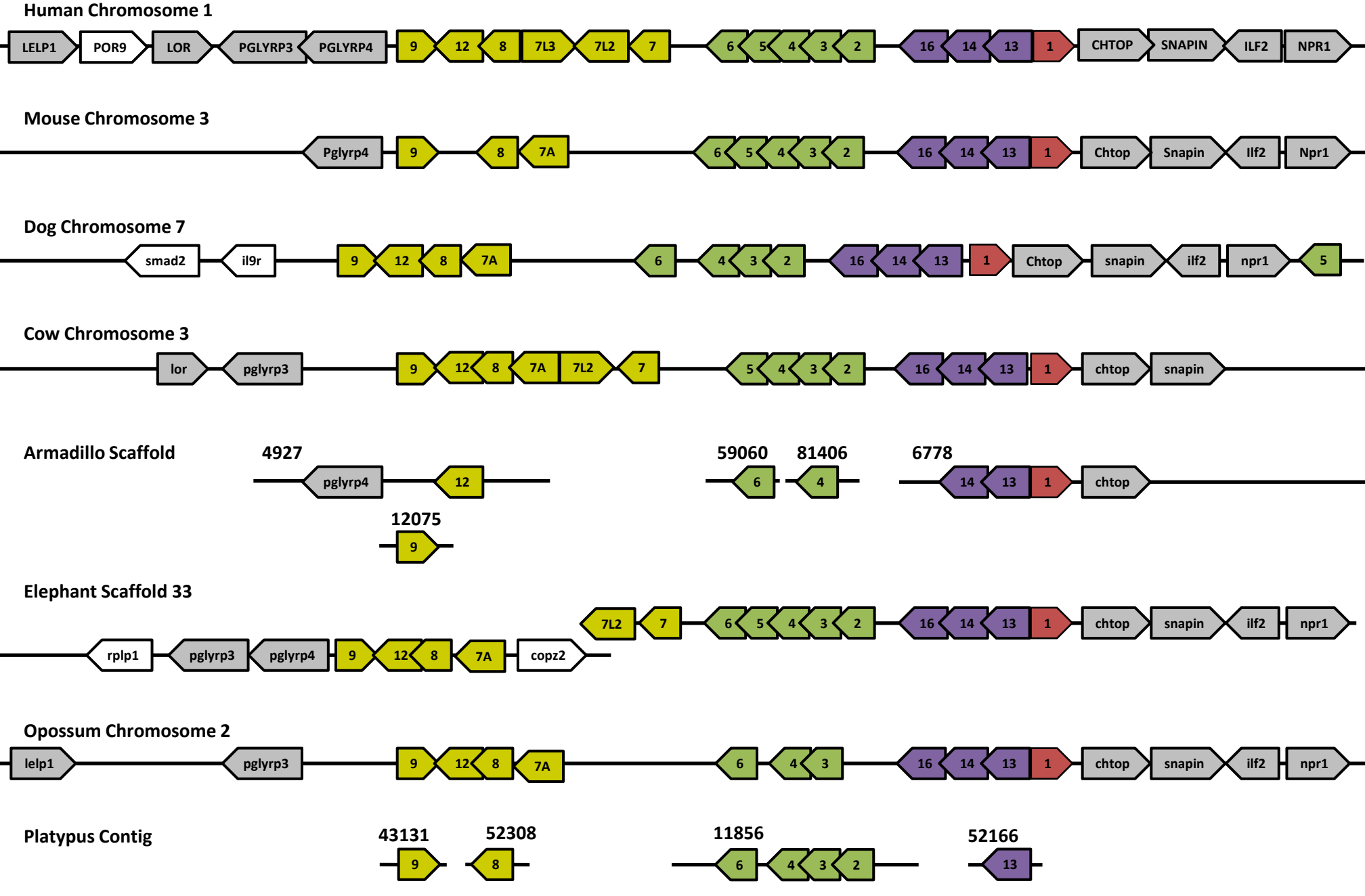
Supplemental Figure 7: Pairwise identity matrix for mammalian S100A7 loci. Number of predicted amino acid substitutions per site using a Dayhoff matrix based model analyzing 108 positions between 12 sequences. Darker pink in the heatmap indicates greater divergence between the two sequences.

Supplemental Figure 8: Amino acid alignments for mammalian S100A7 loci. Amino acid sequences of mammalian S100A7 family members. Dashes indicate gaps inserted for alignment to other family members. The S100 (pseudo) EF-hand Ca^{2+} binding loop is in dark blue, the C-terminal EF-hand Ca^{2+} binding loop in light blue, the linker region connecting the EF-hands in green, alpha-helices in black, and non-helical regions in gray.

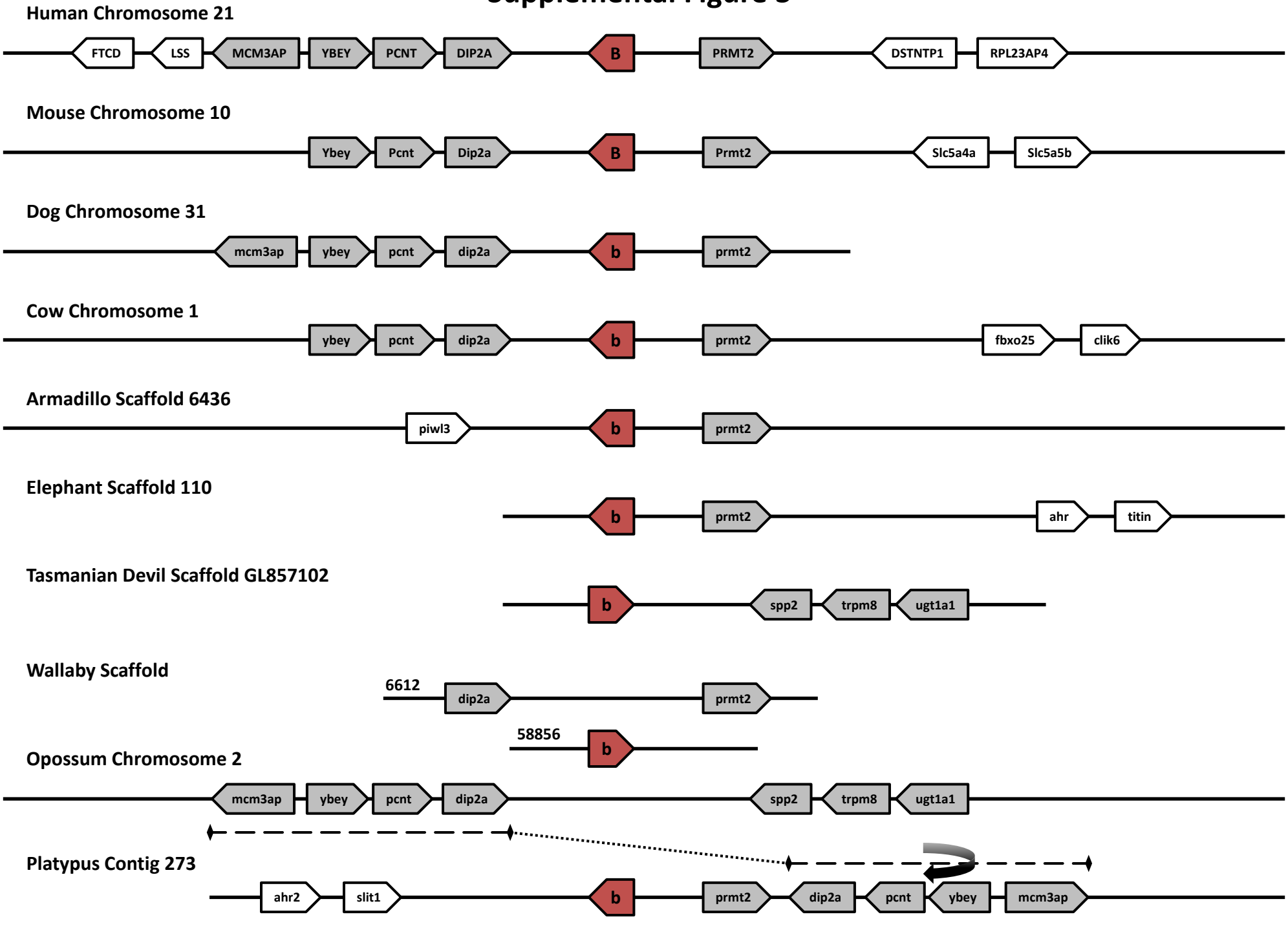
Supplemental Figure 1



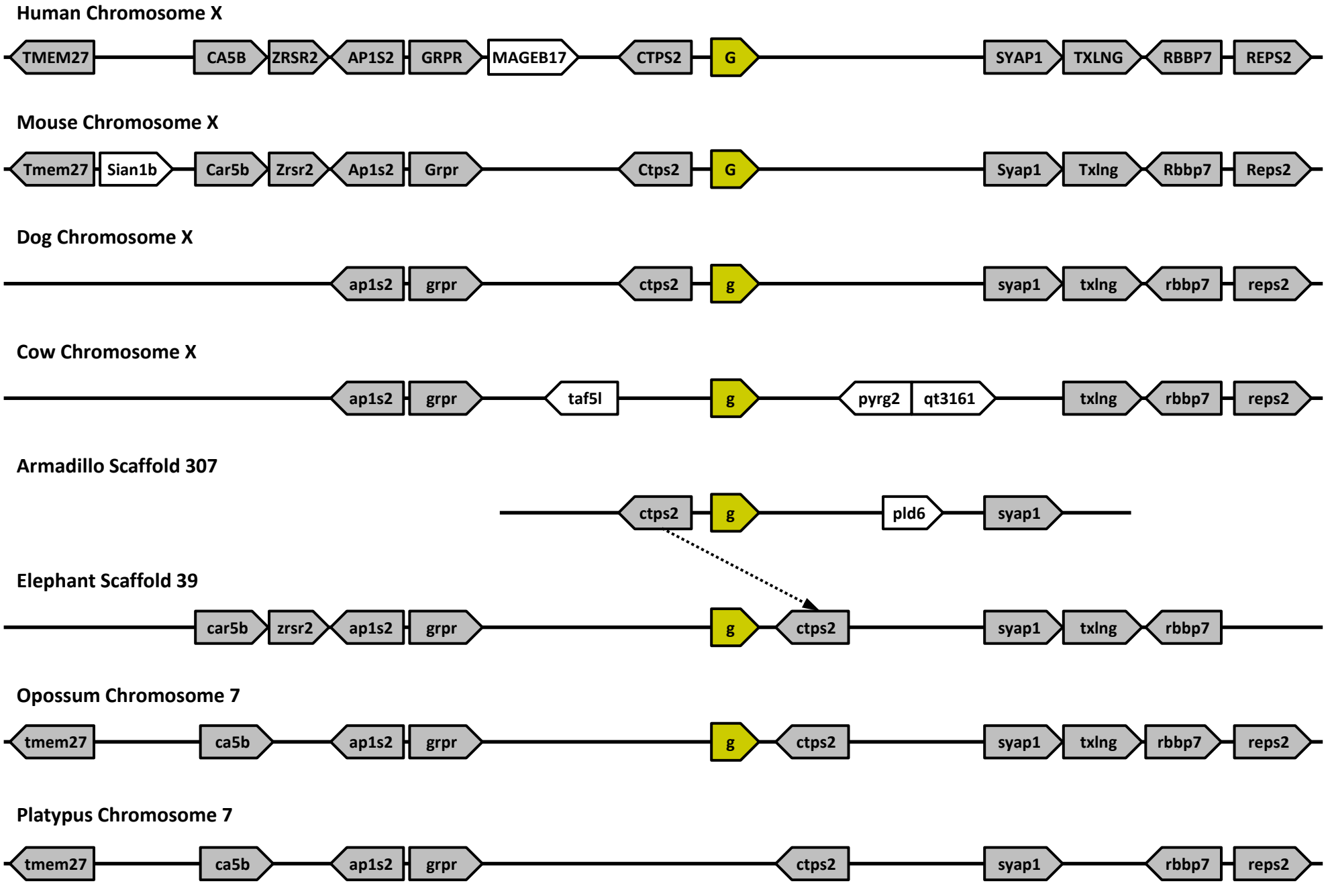
Supplemental Figure 2



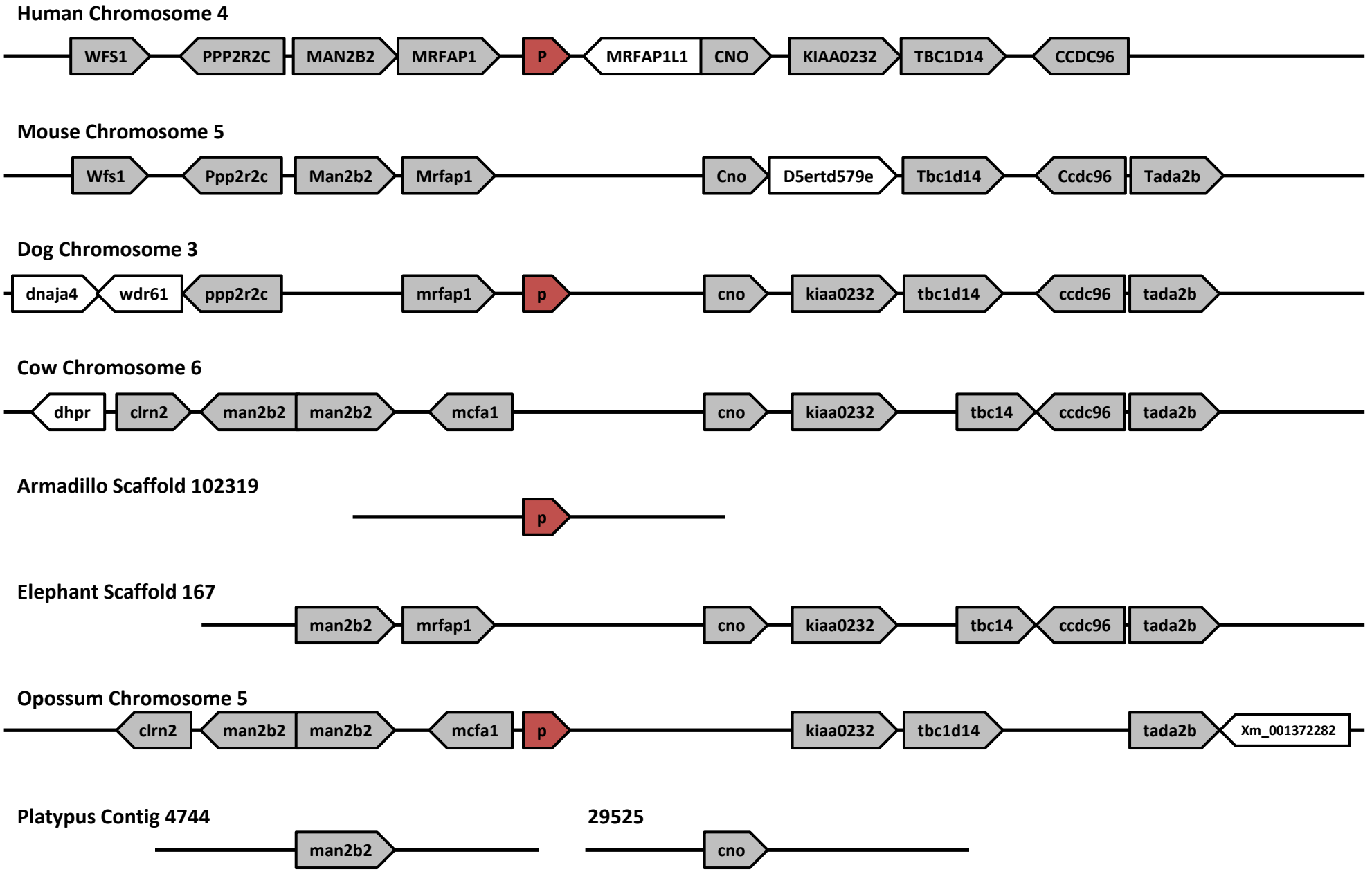
Supplemental Figure 3



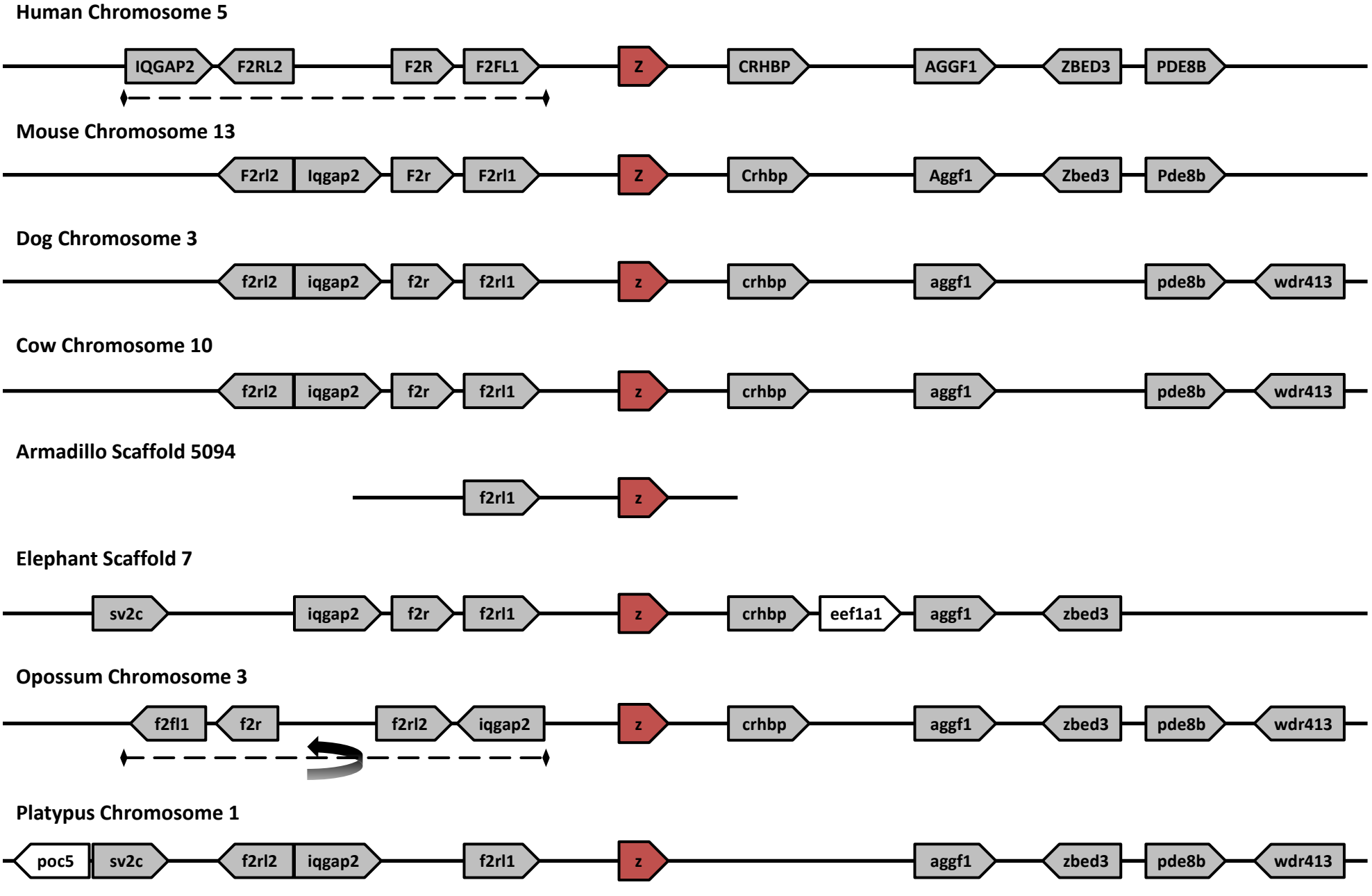
Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7

	hA7	hA7A	hA7L2	mA7	dA7	cA7 t	cA7 c	cA7A	eA7A	eA7L2	eA7
human_S100A7A	0.060										
human_S100A7L2	0.649	0.624									
mouse_S100A7	1.515	1.528	2.013								
dog_S100A7	1.647	1.660	1.972	0.301							
cattle_S100A7(t)	0.423	0.438	0.717	1.463	1.590						
cattle_S100A7(c)	0.386	0.406	0.713	1.497	1.639	0.088					
cattle_S100A7A	1.440	1.430	2.034	0.157	0.211	1.475	1.524				
elephant_S100A7A	1.520	1.497	1.981	0.247	0.261	1.479	1.486	0.179			
elephant_S100A7L2	0.529	0.542	0.663	1.666	1.870	0.532	0.507	1.753	1.700		
elephant_S100A7	0.436	0.459	0.766	1.559	1.685	0.523	0.529	1.577	1.548	0.173	
opossum_S100A7A	1.495	1.498	1.904	0.374	0.481	1.445	1.440	0.378	0.456	1.588	1.507

Supplemental Figure 8