

RESEARCH ARTICLE

Open Access

The evolution of *S100A7*: an unusual gene expansion in *Myotis* bats



Ana Águeda-Pinto^{1,2}, L. Filipe C. Castro^{2,3} and Pedro J. Esteves^{1,2,4*} 

Abstract

Background: The *S100A7* gene, also called psoriasin, was first described as an upregulated protein in psoriatic skin. For the past years, the importance of this protein as a key effector of innate immunity has been clearly established, not only due to its importance protecting against bacteria skin insult in humans, but also because of its important role in amplifying inflammatory processes. Given the importance of *S100A7* in host defense, *S100A7* genes have been mostly studied in humans. Here we provide a detailed analysis of the evolution of the gene family encoding for the *S100A7* protein in mammals.

Results: Examination of several mammalian genomes revealed an unexpected variation in the copy number of *S100A7*. Among the most representative mammalian groups, we report that multiple events of duplication, gene loss and high mutation rates are shaping the evolution of this gene family. An unexpected result comes from *Myotis* species (order Chiroptera), where we found an outstanding *S100A7* gene radiation, resulting in more than 10 copies in *M. lucifugus* and 5 copies in *M. brandtii*. These findings suggest a unique adaptive road in these species and are suggestive of special role of this protein in their immune system.

Conclusions: We found different evolutionary histories among different mammalian groups. Overall, our results suggest that this gene family is evolving under the birth-and-death model of evolution. To our knowledge, this work represents the first detailed analysis of phylogenetic relationships of *S100A7* within mammals and therefore will pave the way to further clarify their unique function in the immune system.

Keywords: *S100A7* (psoriasin), Chiroptera, Gene duplication, Model of concerted and birth-and-death evolution

Background

S100A7 (S100 Calcium Binding Protein A7) is a member of the S100 protein family, sharing the typical EF-hand helix-loop-helix domain that is responsible for the calcium binding function [1]. *S100A7* was first identified as an upregulated protein in psoriatic keratinocytes, having an important role in hyperplasia and in inflammatory processes observed in psoriatic skin lesions [2]. *S100A7* is not only overexpressed in psoriatic skin, but also in atopic dermatitis, Darier disease, mycosis fungoides and skin cancer, suggesting a major role for this protein in inflammation and keratinocyte differentiation [3–7]. Several studies show that high levels of *S100A7* expression

occur in response to several cytokines, including IL-17, IL-22 and TNF (Tumor Necrosis Factor) [8–10]. Following *S100A7* stimulation, this protein can activate different cellular signaling pathways, both in keratinocytes and neutrophils, such as AP-1 (activator protein 1), NF- κ B (nuclear factor kappa-light-chain-enhancer of activated B cells), and STAT3 (signal transducer and activator of transcription 3), resulting in the upregulation of multiple pro-inflammatory cytokines and in an amplifying inflammatory process [11, 12]. Moreover, it has been established that after secretion by keratinocytes, *S100A7* proteins also work as a chemotactic agent towards neutrophils and T-cells [13, 14]. Besides its chemotaxis properties, *S100A7* is also an important *Escherichia coli*-cidal antimicrobial protein. *S100A7* is expressed in areas with high bacterial colonization, being one of the main antimicrobial proteins expressed in normal skin [15]. Moreover, it is the main *Escherichia coli*-killer compound, explaining why *E. coli* rarely infects healthy skin,

* Correspondence: pjesteves@cibio.up.pt

¹CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos genéticos, Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal

²Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre S/N, 4169-007 Porto, Portugal

Full list of author information is available at the end of the article



even in areas where its concentration is high [16, 17]. Given its antimicrobial properties, along with its ability to enhance the host defense functions at sites of infection and inflammation, *S100A7* has been considered a key effector of innate immunity.

Genomic *loci* analyses showed that most *S100A* genes are clustered on a single chromosome with a conserved arrangement among eutherians (e.g. human and mouse). In humans, *S100A7* genes are located in the epidermal differentiation complex (EDC) on chromosome 1q21 that comprises several other genes predominantly expressed in the skin [18, 19]. So far, the largest number of *S100A7* copies was described in primates, with *Homo sapiens* having three functional genes and two nonfunctional [20].

Several studies support that extant eutherian species have two genes, *S100A7* and *S100A15*, likely to have originated via gene duplication after the split from marsupials around 150 million years ago (Mya) (Fig. 1) [21]. In some eutherian mammals, like human and chimpanzee, *S100A7* experienced several events of duplication, while in rodents this gene lineage was lost [22, 23]. Interestingly, the ancient *S100A15* is preserved in rodents, while was lost in the human lineage [23, 24]. Overall, *S100A7* has most likely undergone multiple events of gene rearrangement and duplication during mammalian evolution. The variable number of *S100A7* copies found in Primates along with the importance of this protein in the innate immune system, prompted us to investigate its molecular evolution among eutherian mammals. Moreover, a comparative and phylogenetic approach involving a wider range of eutherian mammals should shed some light on the evolution of *S100A7* gene family. In this study, we show that after the radiation of *S100A7* from marsupials, this gene experienced a dynamic pattern of gene duplication and loss. Surprisingly, in the *Myotis* family (Chiroptera), we show strong evidences of multiple events of duplication, followed by events of recombination and acceleration of point mutations. On the basis of the obtained

results, we propose a birth-and-death model of evolution acting in this gene family.

Results

The *S100A7* repertoire in mammals

In this study, our main goal was to explore the evolutionary patterns of *S100A7* genes in mammals and the evolutionary forces that are shaping this gene family. For this, human *S100A7* and *S100A7A* were used to perform blastn searches and collect *S100A7-like* sequences from several available genomes. Our database search detected the presence of *S100A7* genes in seven eutherian orders, including order Artiodactyla (12 species), order Carnivora (four species), order Chiroptera (7 species), order Primates (12 species), order Cetacea (one species), order Perissodactyla (two species) and superorder Afrotheria (two species). In total, 82 sequences from 40 species were included in this study (Table 1 and Additional file 1).

Besides the complete *S100A7* coding sequences, a detailed sequence analysis suggested a number of inconsistent annotations in the NCBI and Ensembl databases. For example, we found a *S100A7A* (XM_006501635.3) and a *S100A7* annotation (AY582964.1) in mouse with 100% identity between them, suggesting that these sequences are in fact the same gene. Moreover, we observed that *S100A7* from mouse presented a low degree of similarity when aligned with the *S100A7* coding sequences from other mammals (Additional file 3). In fact, after further analysis, we found that the annotated *S100A7* from mouse was in fact a *S100A15* gene. This situation was also found in the brown rat (*Rattus norvegicus*, NM_001109471.1). Given the high degree of similarity of these genes with the *S100A15* gene from other mammals (Additional file 4), here we assume that these sequences correspond to a *S100A15* gene. Three annotated *S100A7* for the minke whale (*Balaenoptera acutorostrata*, XM_007178590.1, XM_007178589.1 and XM_007198423.1) were found. The second and the third correspond to partial sequences and, for

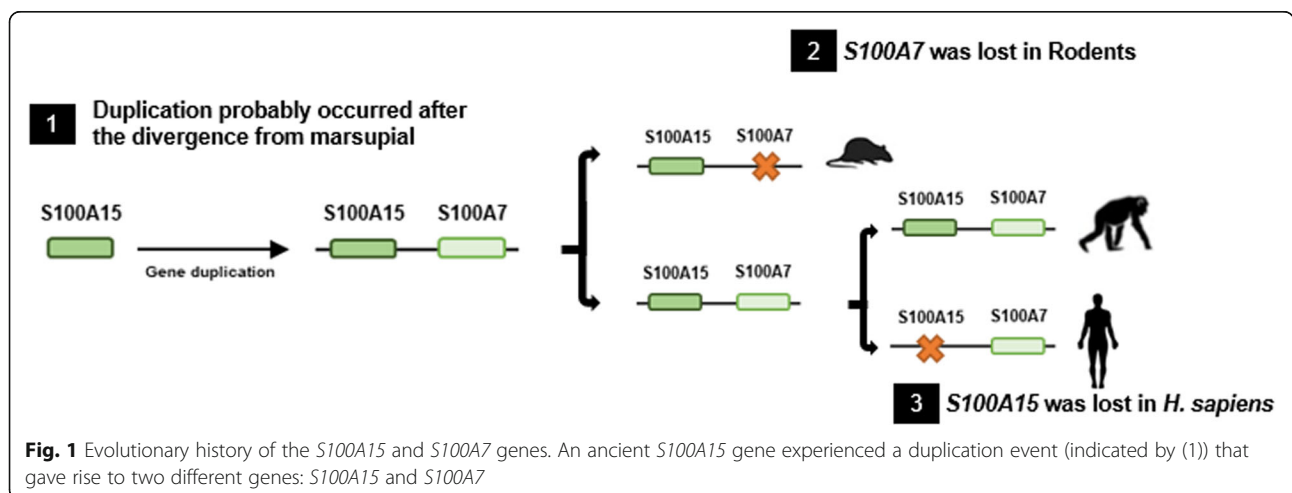


Table 1 Number of *S100A7* copies found in different mammalian orders

| Order | Family | Species name | <i>S100A7</i> | |
|-------------------------|-----------------|-----------------------------------|-----------------------|---|
| Primates | Hominidae | <i>Homo sapiens</i> | 5 ^a | |
| | | <i>Pan troglodytes</i> | 2 | |
| | | <i>Pan paniscus</i> | 2 | |
| | | <i>Gorilla gorilla</i> | 2 | |
| | Hylobatidae | <i>Nomascus leucogenys</i> | 2 | |
| | Cercopithecidae | <i>Papio anubis</i> | 1 | |
| | | <i>Macaca mulatta</i> | 1 | |
| | | <i>M. fascicularis</i> | 2 | |
| | | <i>M. nemestrina</i> | 1 | |
| | | <i>Rhinopithecus bieti</i> | 2 | |
| | Cebidae | <i>Cebus capucinus</i> | 1 | |
| | | <i>Saimiri boliviensis</i> | 1 | |
| | Chiroptera | Vespertilionidae | <i>Myotis davidii</i> | 3 |
| <i>Myotis brandtii</i> | | | 5 | |
| <i>Myotis lucifugus</i> | | | 13 | |
| <i>Eptesicus fuscus</i> | | | 1 | |
| Miniopteridae | | <i>Miniopterus natalensis</i> | 2 | |
| Pteropodidae | | <i>Pteropus vampyrus</i> | 1 | |
| | | <i>Pteropus alecto</i> | 2 | |
| | | <i>Rousettus aegyptiacus</i> | 3 | |
| Perissodactyla | | Equidae | <i>Equus caballus</i> | 1 |
| | | | <i>Equus asinus</i> | 1 |
| Artiodactyla | Bovidae | <i>Bos taurus</i> | 3 | |
| | | <i>B. indicus</i> | 3 | |
| | | <i>Bison bison</i> | 2 | |
| | | <i>Bubalus bubalis</i> | 3 | |
| | | <i>Pantholops hodgsonii</i> | 2 | |
| | | <i>Capra hircus</i> | 3 | |
| | | <i>Ovis aries</i> | 2 | |
| | Suidae | <i>Sus scrofa</i> | 2 | |
| | Balaenopteridae | <i>Balaenoptera acutorostrata</i> | 3 ^b | |
| | Camelidae | <i>Camelus dromedarius</i> | 1 | |
| | | <i>Camelus ferus</i> | 1 | |
| | | <i>Camelus bactrianus</i> | 1 | |
| | | <i>Vicugna pacos</i> | 1 | |
| Carnivora | Ursidae | <i>Ailuropoda melanoleuca</i> | 1 | |
| | | <i>Ursus arctos</i> | 1 | |
| | | <i>Ursus maritimus</i> | 1 | |
| | Otariidae | <i>Callorhinus ursinus</i> | 1 | |
| Afrotheria | Orycteropodidae | <i>Orycteropus afer afer</i> | 2 | |
| | Elephantidae | <i>Loxodonta africana</i> | 1 ^c | |

^aPresence of two described pseudogenes

^bPresence of two partial sequences

^cPartial sequence

this reason, were not included in the phylogenetic analysis. Moreover, the search in the elephant (*Loxodonta africana*) genome also resulted in a partial sequence (XM_007948587.1) and was also not used in this study.

It was previously reported that *S100A7* from mouse and rat may have been lost after their separation marsupials [22, 23]. In fact, and as mentioned before, our results support these findings since no functional *S100A7* sequences were found in rodents (mouse and rat). Moreover, after a careful analysis of the mouse genome, we found remnants of the *S100A7* gene (Additional file 5) at the *S100* locus, supporting further the gene loss hypothesis. The same was also true in human genome, where we also found a partial sequence of *S100A15* gene. Moreover, the partial sequences found in mouse and human are in the same locus as the remaining *S100* genes (Additional file 5). Our database research also revealed that species like *Canis lupus* (family Canidae), *Acinonyx jubatus* and *Panthera pardus* (both from family Felidae) presented no *S100A7* annotation. However, when we further investigated the genomic region corresponding to the *S100A7* locus [22] for several canids and felines species we found that *S100A15* gene is still present (Additional file 4).

On the other hand, we found several events of *S100A7* duplication in other eutherian species (Table 1 and Additional file 1). Interestingly, blast searches at Chiroptera genomes hinted at a high number of *S100A7* genes, with six *S100A7* genes being assigned into the Pteropodidae family (suborder Megachiroptera) while the remaining sequences were assigned to the Vespertilionidae family (suborder Microchiroptera). Interestingly, the Little Brown Bat (*Myotis lucifugus*), a representative of the Vespertilionidae family (genera *Myotis*), presented 13 copies of *S100A7* gene, while *M. brandtii* presented five copies, suggesting that a rapid evolution of *S100A7* is occurring in these species. Besides the complete *S100A7* sequences, six other incomplete sequences were identified for the Chiroptera order (see Additional file 6). All the incomplete sequences belong to suborder Microchiroptera, being distributed by Megadermatidae, Rhinolophidae, Mormoopidae, Hipposideridae and Phyllostomidae families. However, these incomplete sequences were not used in the recombination and phylogenetic analyses performed in this study.

Phylogenetic reconstruction suggests specific *S100A7* expansions in mammals

Before further phylogenetic analyses, the retrieved sequences were screened using the RDP [25] to look for any evidence of recombination in the alignment. For the 82 complete sequences of *S100A7*, only *M. brandtii_A7(3)* presented a consistent recombination breakpoint (nucleotide positions 158–183) with strong statistical support for 5 different methods (*p*-values < 0.05) (Table 2). *M. lucifugus_A7(4)* and *M. brandtii_A7(4)* were identified as the donors of this sequence (Table 2). From the obtained results, it seems that the recombination

Table 2 Results of the recombination analysis of the *S100A7* genes using RDP

| Sequence | Most likely donor sequence | | Breakpoint Start | Methods (average <i>P</i> -value) | | | | | |
|--------------------------|----------------------------|--------------------------|---------------------|-----------------------------------|------------------------|------------------------|------------------------|-------------------------|------|
| | Major Parent | Minor Parent | | GENECONV | BootScan | MaxChi | Chimaera | SIScan | 3Seq |
| <i>M. brandtii_A7(3)</i> | <i>M. lucifugus_A7(4)</i> | <i>M. brandtii_A7(4)</i> | 158–183 | 2.53×10^{-02} | 1.05×10^{-02} | 6.05×10^{-07} | 1.05×10^{-07} | 1.147×10^{-06} | – |

event involved an ancestral *S100A7* gene of *M. brandtii*, very similar to the sequence of *M. lucifugus_A7(4)*, which in the process of evolution might have been lost (Fig. 2). Nevertheless, we cannot rule out the hypothesis that this gene can be present in the genome of *M. brandtii*, but it is not available in public databases. Taking into consideration that recombination events might interfere with the phylogenetic relationships of *S100A7* sequences, *M. brandtii_A7(3)* was not included in the phylogenetic analysis.

In the *S100A7* phylogenetic reconstruction, a total of seven mammalian orders are represented. It can be observed a concordant topology with the accepted evolutionary relationships of the eutherian mammals [26, 27]. However, the monophyly of these groups is not statistically supported by the bootstraps analysis (Fig. 3). Given the topology of the phylogenetic tree, it seems that after the radiation of the *S100A7* ancestor gene it continued to be subject to gene duplication and loss. We have previously shown that *S100A7* gene family has been shaped by multiple events of duplication over ~ 35 Mya of primate evolution [28], predating the evolutionary origins of the divergence of Platyrrhini and Catarrhini primates. In Fig. 3, it is clear that these duplication events are not specific to Primate lineage. For example, regarding the order Chiroptera, it is possible to observe that in the Pteropodidae family, *P. vampyrus* maintained a single copy of *S100A7*, while in *P. alecto* and *R. aegyptiacus* the single copy ancestor of *S100A7* experienced duplication, resulting in a total of two and three

genes, respectively. In Vespertilionidae, our ML tree further suggests that a radiation of *S100A7* occurred before the divergence of early *Myotis* species (around 20 Mya [29–31]). Strikingly, in *M. lucifugus*, at least 10 copies can be found, which suggest the occurrence of multiple duplication events. Accordingly, duplication events can also be observed in family Bovidae (Fig. 3).

Genomic *loci* analysis clearly showed that the *S100A* genes clustered in close proximity in a single chromosome in mammals (human, mouse and opossum): the *S100A3*, *S100A4*, *S100A5* and *S100A6* form a single cluster, next to *S100A7*, *S100A15*, *S100A8*, *S100A12* and *S100A9* genes [32, 33]. Thus, we next analyzed the genomic organization of *S100A7* genes and respecting flanking genes from five different species: human, mouse, domestic cow (*B. taurus*), flying fox (*P. alecto*) and microbat (*M. lucifugus*) (Fig. 4). The analysis of the flanking genes revealed that in each of these five mammals, *S100A7* multi-copies are clustered between *S100A9*, *S100A12* and *S100A8* and *S100A15*, *S100A6* and *S100A5*, suggesting a conserved gene rearrangement between these species. However, it should be notice that these gene clusters are in opposite orientation from the human *S100A7* gene locus (Fig. 4). Regarding *M. lucifugus'* *S100A7* copies, we found that these genes are dispersed for three different scaffolds: gl429918, gl429970 and gl431328. Among these genes, five are in close proximity with the expected two clusters in the scaffold gl429918, following the same arrangement as the *S100A7* copies of

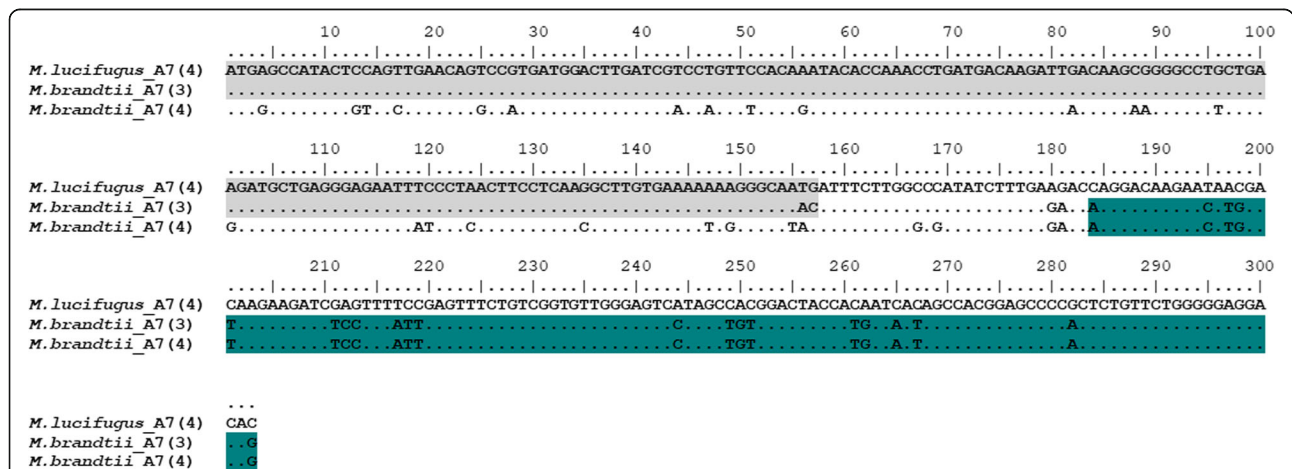


Fig. 2 Nucleotide alignment of the recombinant *S100A7* sequence *M. brandtii_A7(3)* and the parental sequences as defined by RDP (*M. lucifugus_A7(4)* and *M. brandtii_A7(4)*). Nucleotide positions 158–183 shows the recombination breakpoint found by five different methods (*p*-values < 0.05). The grey region highlights the recombination area between *M. lucifugus_A7(4)* and *M. brandtii_A7(3)* and the blue region highlights the recombination area between *M. brandtii_A7(4)* and *M. brandtii_A7(3)*

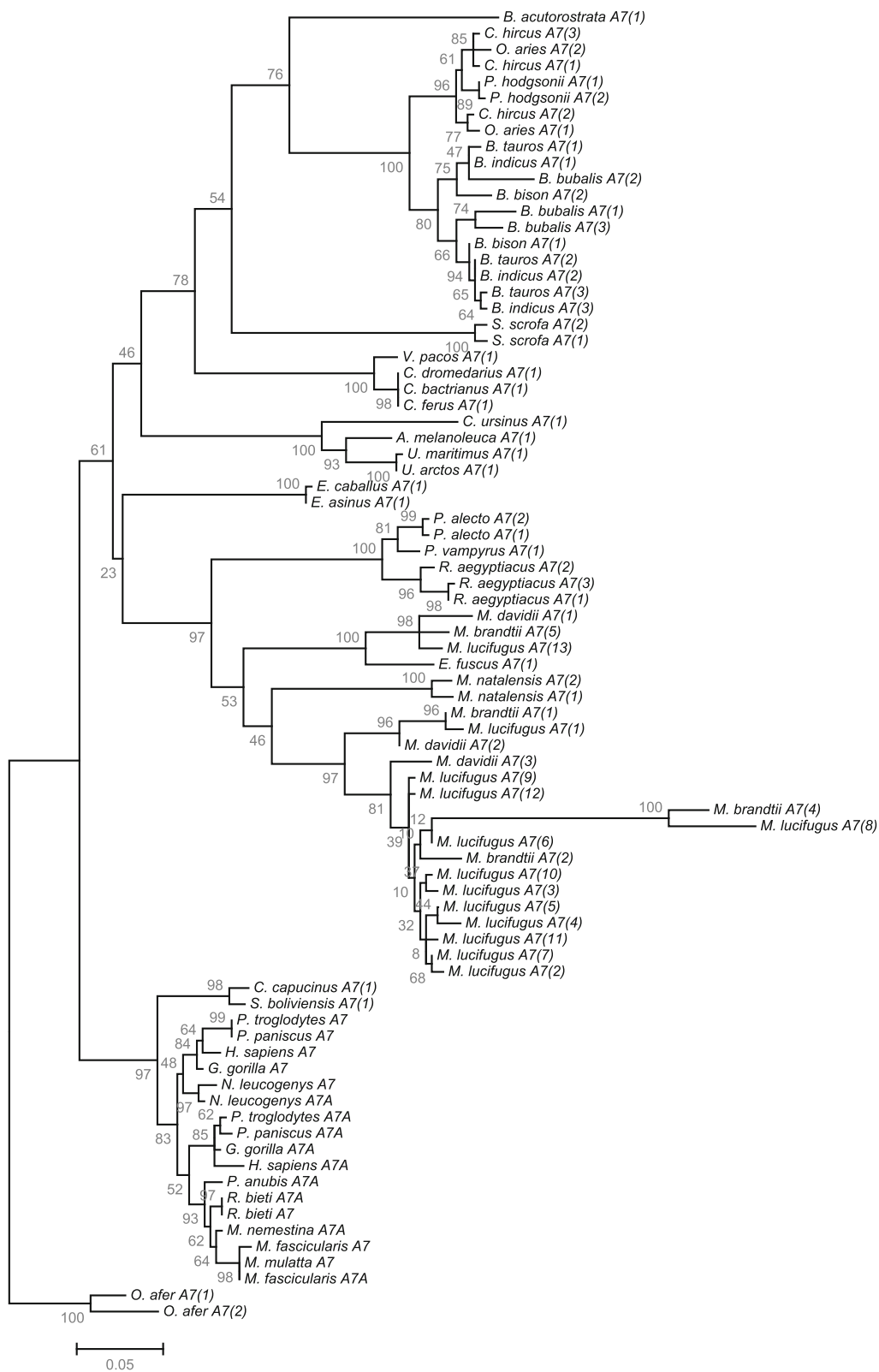
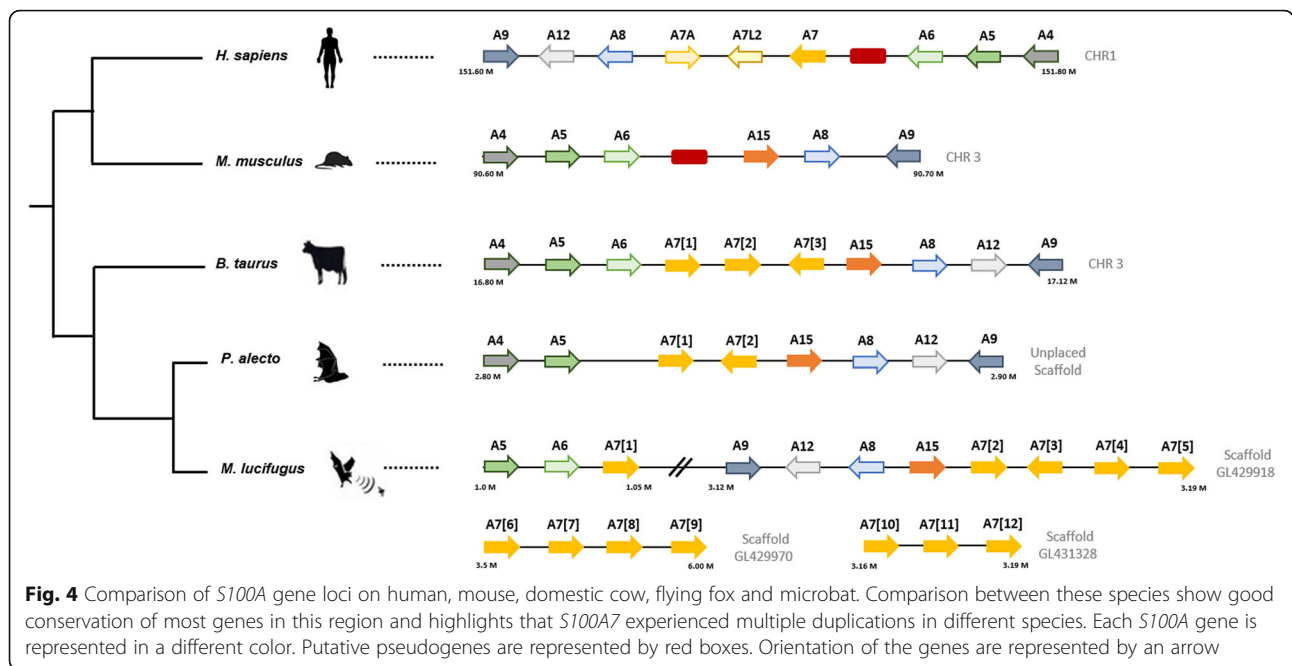


Fig. 3 Phylogenetic analysis of *S100A7* genes in eutherian mammals. The analysis were performed with 1000 generations and 1000 bootstrap searches. Bootstrap values (%) are indicated on the branches. The abbreviations correspond to the ones shown in Additional file 1



the other mammalian species (Fig. 4). Since the scaffold gl429918 ends right after the position of the *M. lucifugus_A7(5)* gene, it is possible that the position of the remaining genes in two other scaffolds do not allow us to infer the true location of these genes in the genome. Overall, the species that presented multiple copies of the *S100A7* gene revealed a conserved synteny among them and, therefore, further support the hypothesis that multiple events of duplication are shaping the evolution of this gene.

Evidence of a high mutation rate in specific *S100A7* gene lineages

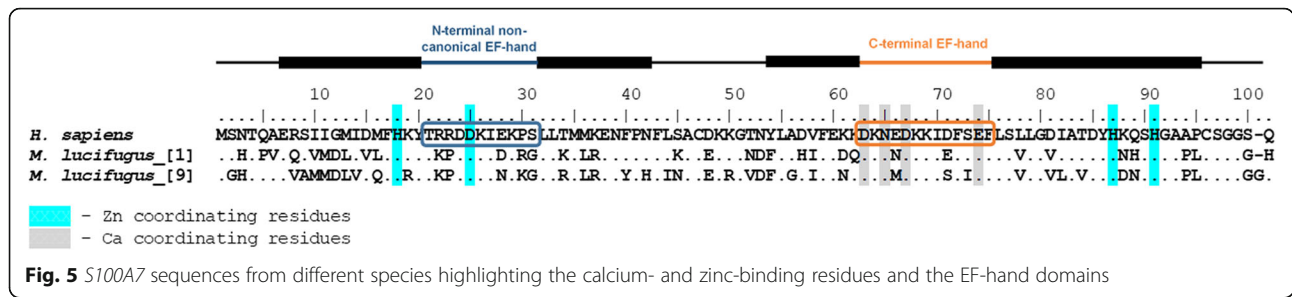
From the obtained ML tree, and taking in consideration the longer branch lengths, it appears that *M. lucifugus_A7(8)* and *M. brandtii_A7(4)* might be experiencing an acceleration of the mutation rate, compared to the remaining *S100A7* sequences (Fig. 3). In the Tajima’s test, the hypothesis that the compared genes have equal rates of mutation was rejected (Table 3). In fact, from the Tajima’s test results it is possible to observe that *M. lucifugus_A7(8)* and *M. brandtii_A7(4)* are experiencing a rate of mutation two times higher than the other *S100A7* genes.

Given the high mutation rates observed, selective pressures acting on this branch were calculated using codeml of the PAML 4.9 package [34]. Codeml results suggested an intensification of selective pressures, when compared to the remaining branches, with an estimation of $\omega = 1.44$, indicating that this branch is under positive selection (Additional file 7). Moreover, RELAX analysis also confirmed an intensification of the mutation rates (Additional file 7). Considering the *S100A7* protein structure, it is interesting to observe that despite *M. lucifugus_A7* [8] presents around 32 and 44% of amino acid differences from *M. lucifugus_A7* [4] and *H. sapiens_A7*, respectively, none of these amino acids fall within the Ca^{2+} and Zn^{2+} coordinating residues [35] (Fig. 5), suggesting a maintenance of the core function of this protein. However, although no ORF disrupting mutations were found in the coding sequences of these genes, we can not rule out the hypothesis that the accumulation of several non-synonymous mutations might indicate a pseudo-genization process acting on *M. lucifugus_A7* [8] and *M. brandtii_A7* [4].

Table 3 Results of Tajima’s relative rate test

| Sequence A | Sequence B | Outgroup | Identical sites | Divergent sites | Sequence A Specific | Sequence B Specific | Outgroup Specific | X ² test | p-value ¹ |
|---------------------------|----------------------------|--------------------------|-----------------|-----------------|---------------------|---------------------|-------------------|---------------------|----------------------|
| <i>M. lucifugus_A7(8)</i> | <i>M. lucifugus_A7(13)</i> | <i>E. caballus_A7(1)</i> | 196 | 6 | 49 | 24 | 28 | 8.56 | 0.003 |
| <i>M. brandtii_A7(4)</i> | <i>M. brandtii_A7(5)</i> | <i>E. asinus_A7(1)</i> | 196 | 6 | 47 | 27 | 27 | 5.41 | 0.020 |

¹P-value <0.05 was used to reject the hypothesis of equal rates between the three lineages



Discussion

In the past years, *S100A7* has emerged as a key regulatory component in the immune system, not only because of its ability to defend host from invading pathogens, but also for having an important role in inflammation and keratinocyte differentiation [2, 14, 15]. So far, humans were the species with more *S100A7* genes described to date, presenting three functional copies and two nonfunctional ones [19, 20]. Although this gene family has been the focus of several studies, their evolutionary history among mammals has not been studied before. In this study, we present an extensive analysis, combining genome analysis, phylogenetics and comparative genomics to provide the complete history of the evolution of this gene.

Dynamic evolution of the *S100A7* gene in mammals

From our results, the *S100A7* repertoire varies significantly among mammals (Fig. 3). Studies suggest that after the split from marsupial, the *S100A15* underwent a duplication event, resulting in a *S100A7* ancestor gene present only in the eutherian mammals [22]. Accordingly, we did not find any evidence of *S100A7* in marsupial species. Moreover, we found that rodents and some Carnivora lineages also lack *S100A7*. The most parsimonious explanation for the obtained data is that during evolution, these lineages lost the *S100A7* gene. Interestingly, they retain the *S100A15* gene orthologue (Additional file 4 and [22]). It has been proposed that the mouse *S100A15* parallels the structure, gene expression and protein processing patterns that the *S100A7* protein from humans [23]. In fact, the loss of this protein in the mentioned mammals might suggest a compensatory effect of *S100A15*. However, further studies are needed to validate this theory.

Expansion of *S100A7* in bats

Although bats represent one of the largest and diverse group of mammals and are prone to various emerging infectious diseases, little is still known about their immune system [36–38]. Vespertilionidae is the largest family in Chiroptera, with more than 400 known species [39]. However, the high number of *S100A7* genes found in this family appears to be the result of several duplicated genes in three species: *M. brandtii*, *M. lucifugus* and *M. davidii*, with

S100A7 copy numbers ranging from three to 13. In bats, a comparable polymorphism to that observed for *S100A7* comes from the major histocompatibility complex (MHC) class II DRB genes, which are known to play a major role in immune system [40, 41]. Indeed, a wide range of variability among different species of bats was also found in MHC loci, where up to 10 DRB loci were found in sac-winged bat (*Saccopteryx bilineata*) while in genus *Noctilio* only one locus was found [42]. Moreover, a study focusing in bitter taste receptor genes (*Tas2rs*) showed a lineage-specific duplication of several of these receptors that only occurred in *Myotis* bats, with the new copies of *Tas2rs* presenting functional differentiations and functional innovations following duplication [43]. It is known that bats have a high propensity to tolerate massive diseases originated by a high array of pathogens [37, 44–47]. For example, bats harbor a higher proportion of zoonotic viruses than all other mammalian orders [48, 49]. A recent study using the Egyptian Rousette (*Rousettus aegyptiacus*) shows that enhanced tolerance to infection might be a result of an expansion and diversification of several loci, as well as unique adaptations in type I interferons responses and natural killer cell signaling pathways [37].

Besides the multiple events of duplication that occurred in *S100A7* gene family in Chiroptera order, our results also show strong evidences of recombination and high rates of mutation. In fact, the phylogenetic tree presented longer branches in *M. lucifugus*_A7(8) and *M. brandtii*_A7(4) coding sequences when compared with the remaining *S100A7* genes. From our results, we observed that these genes are experiencing a high mutation rate, comparatively to the remaining *S100A7* from bats (Tables 2 and 3). Similar to our results, a previous study also reported evidences of episodes of positive selection acting on the Toll-like receptor 8 (TLR8), shaping its' diversity throughout bats evolution [50]. Interestingly, the authors showed that positive selection is especially strong in codons located in the ligand-binding ectodomain, which might contribute to a variation in the ability of different bats to recognize molecular patterns of virus [50]. In S100 proteins, upon Ca²⁺ binding, the EF-hand motif undergoes a conformational change. This change in conformation is responsible for the exposure of a

hydrophobic surface that is fundamental for target recognition [1, 35]. The same way, it is known that the protective role of S100A7 against *E. coli* is highly dependent of the binding of Zn^{2+} [16]. However, it is interesting to note that despite the high mutation rates observed in these proteins, their Ca^{2+} and Zn^{2+} motifs are not modified (Fig. 5). From the alignment of these genes, both copies appear to be functional as no obvious deleterious mutations or early stop codons were found. However, since no expression patterns are currently available for these genes, we hypothesise that a recent pseudogenization event might have affected the promoter or other regulatory region (UTR), resulting in the high mutation rates observed.

Given the wide geographic distribution of *Myotis* (except Antarctica) and the importance of S100A7 in the immune system, we suggest that these duplication events might have an important role in protecting these species when exploiting new environmental settings.

Birth-and-death evolution model

Our results raised the puzzling question on how and why this complex evolutionary pattern arose in eutherian mammals. Gene duplication is a fundamental process in genome evolution [51, 52]. While some young duplicates are degraded in the process of evolution, some duplicate pairs are able to survive in a long-term. While still controversial, several mechanisms can explain the maintenance of duplicated genes in the genome, including neofunctionalization, subfunctionalization, and increased gene-dosage advantage [51, 53]. A recent theory suggest that young duplicates might be controlled by dosage balance and tight co-regulation of tandem duplicates, allowing the initial survival of the duplicates, followed by its slower adaptation to the genome [54–56]. In our results, we find that repeated gene duplications occurred in a relative quick succession, resulting in several genes closely related. The sequence similarity found in several multigene families are usually a result of conversions and other recombination events, resulting in homogenization of all members of a multigene family, even in the presence of mutations [57]. Yet, the S100A7 multigene family also displays highly divergent sequences such as that of *M. brandtii_A7(3)*, *M. brandtii_A7(4)* and *M. lucifugus_A7(8)*. Moreover, there is also evidence for the presence of two S100A7 pseudogenes in *H. sapiens* [20] and several events of gene loss among the eutherian mammals. Therefore, we propose a model of concerted and birth-and-death evolution to explain the obtained results and the evolution of S100A7 multigene family [57, 58]. This was not the first report of a multigene family with important roles in the immune system having evolved under the birth-and-death model of evolution [59]. In fact, Nei and collaborators suggested that given the importance of some gene families in immune function, natural selection might be a major force of diversification, acting in favor of functional diversification, helping hosts to cope with illness and infection [59].

Conclusion

In eutherian mammals, multiple duplications, gene loss, recombination and acceleration of point mutations characterize the evolution of the S100A7 multigene family. Moreover, we suggest a model of concerted and birth-and-death evolution to better explain the evolution of this gene family. Interestingly, several species presented multiple copies of this gene, suggesting that S100A7 may present a special role in the immune defenses of these species. Nevertheless, future studies are needed to fully elucidate the functional roles of these duplicated S100A7 genes in the innate immunity of mammals.

Methods

Identification of S100A7 sequences

S100A7 coding sequences were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov>) and Ensembl (<http://www.ensembl.org/index.html>) databases through blastn searches using as reference the S100A7 and S100A7A genes from human. To include all major mammalian orders we analyzed the genomes of several eutherian species: order Artiodactyla, Cetacea, Carnivora, Chiroptera, Perissodactyla, Primates and from the Superorder Afrotheria. According to previous reports, no S100A7 genes were found in the order Marsupialia, Monotremata and Rodentia [22, 33]. On the other hand, studies also suggest that S100A15 is not present in human [22, 24]. To confirm the presence of gene loss remnants in both rodents and human, S100A7 from human and S100A15 from mouse were used to perform blastn searches in mouse (*GRCm38.p6*) and human (*GRCh38.p12*) genomes, respectively.

Our database search resulted in a final dataset of 82 complete coding sequences, including several duplicated S100A7 genes in major mammalian orders (Additional file 1). The obtained sequences were aligned with Clustal W [60] implemented in BioEdit v7.2.6.1 using default parameters [61] and manually inspected with the exclusion of gaps and partial sequences (Additional file 2). Nucleotide sequences translation into amino acids was also performed using BioEdit.

Recombination and phylogenetic analysis

The sequence alignment was screened for recombination using GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq methods implemented in the RDP software (version 4.95) [25] under the following parameters: sequences were set to linear, Bonferroni correction, highest acceptable *P* value of 0.05 and 100 permutations. Only recombination events detected by three or more methods were considered.

To increase the reliability of the phylogenetic analysis, sequences with evidence of recombination were not included in the phylogeny. In order to infer the phylogenetic relationships of the S100A7 genes in mammals,

evolutionary analyses were conducted in MEGA7 [62] using a Maximum Likelihood (ML) tree based on the Tamura-Nei model [63]. The reliability of the clusters was tested by performing the bootstrap test of phylogeny, with 1000 bootstrap replications. *S100A7* coding sequences from *Orycteropus afer afer* (superorder Afrotheria) were used as outgroup.

Comparative genomics

To infer the duplication history of *S100A7* genes in several mammals, these genes were mapped into their respective chromosomes. For human and mice comparison of *S100* gene loci previous studies were used to identify the location of several *S100* genes [20, 22]. For *Bos taurus*, *Pteropus alecto* and *M. lucifugus* the specific location of *S100A7* genes along with their neighboring genes were collected from NCBI and Ensembl databases using their available genomes (ARS-UCD1.2, ASM32557v1 and Myo-luc2.0, respectively). The human *S100A7* gene locus was used as a model of comparison.

Evolutionary analysis

To test for statistical significance in molecular evolution of *S100A7* bat sequences, a Tajima's relative test was conducted. For that, the statistical parameters were set as default using MEGA7 [62]. A *P*-value < 0.05 was used to reject the null hypothesis of equal rates between the 3 lineages considered simultaneously (sequence A, sequence B and outgroup).

Selective pressures on specific branches were determined using the phylogenetic tree obtained from the Tamura-Nei model mentioned before and codeml of the PAML 4 package [34]. To test these selective pressures, previously described methodology was used [64, 65]. *P* < 0.05 was used to determine whether or not the alternative hypothesis was significant. We further assessed the strength of natural selection (relaxed or intensified) using DATAMONKEY web server (<http://www.datamonkey.org/> [last accessed January 21, 2019] [66, 67]. The *M. lucifugus_A7(8)* and *M. brandtii_A7(4)* branch was the “test” branch, and all the primate branches were assigned as “reference” branches.

Additional files

Additional file 1: List of the sequences of the *S100A7* genes used in this study. All sequences are available from NCBI and Ensembl databases. (PDF 109 kb)

Additional file 2: Alignment of *S100A7* coding sequences from several eutherian mammals. The abbreviations correspond to the ones shown in Additional file 1. Dots = identity with *B. acutorostrata_A7(1)* coding sequence. (PDF 275 kb)

Additional file 3: Alignment between *S100A15* and *S100A7* proteins. Alignment between two *S100A15* proteins from mouse, one *S100A15* from *P. troglodytes* and two *S100A7* proteins from *H. sapiens* and *P.*

paniscus. Dots = identity with *M. musculus_S100A7_XM_006501635.3* protein. (PDF 305 kb)

Additional file 4: Alignment of *S100A15* proteins from several eutherian mammals. Dots = identity with *Mus musculus S10015* protein. (PDF 172 kb)

Additional file 5: Search for the gene loss remnants of *S100A7* and *S100A15* from mouse and human, respectively. (PDF 193 kb)

Additional file 6: Alignment of partial sequences *S100A7* from bats. The abbreviations correspond to the following species: *M. lyra* - *Megaderma lyra*; *R. ferrumequinum* - *Rhinolopus ferrumequinum*; *R. sinicus* - *Rhinolophus sinicus*; *P. patnellii* - *Pteronotus patnellii*; *H. arimiger* - *Hipposideros arimiger*; *D. rotundus* - *Desmodus rotundus*. Dots = identity with *M. lucifugus_A7(1)* sequence. (PDF 149 kb)

Additional file 7: Selection (dN/dS) analyses. (PDF 331 kb)

Abbreviations

MHC: Major histocompatibility complex; ML: Maximum likelihood; Mya: Million years ago; *S100A15*: *S100* Calcium Binding Protein A15; *S100A7*: *S100* Calcium Binding Protein A7; *Tas2rs*: Taste receptor genes; *TLRs*: Toll-like receptors; ω : non-synonymous to synonymous substitution ratios

Acknowledgements

The authors would like to thank Joana Abrantes and Ana M. Lopes for the constructive comments to the final version of the manuscript and to Ana Pinheiro and Fabiana Neves for the discussion regarding the evolutionary analysis.

Funding

FCT - Foundation for Science and Technology supported the doctoral fellowship of AAP (ref. SFRH/BD/128752/2017) and the investigator grant of PJE (IF/00376/2015). This article is a result of the project NORTE-01-0145-FEDER-000007, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). The funders had no role in the design of the study, collection, analysis and interpretation of data, and in writing the manuscript.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files). The data used for this study are available on NCBI and Ensembl under the accession numbers provided in Additional file 1. Data can be obtained from <https://www.ncbi.nlm.nih.gov> and <http://www.ensembl.org/index.html>.

Authors' contributions

AAP collected the coding sequences, analysed the obtained data and wrote the first draft of the manuscript, LFCC and PJE helped to conceptualize the idea, interpreted the data and contributed towards the final write-up of the work. All authors contributed comments to the final version of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos genéticos, Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão,

Portugal. ²Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre S/N, 4169-007 Porto, Portugal. ³CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Av. general Norton de Matos S/N, 4450-208 Matosinhos, Portugal. ⁴CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Rua Central de Gandra 1317, 4585-116 Gandra, Portugal.

Received: 7 February 2019 Accepted: 7 May 2019

Published online: 14 May 2019

References

1. Schafer BW, Heizmann CW. The S100 family of EF-hand calcium-binding proteins: functions and pathology. Trends BiochemSci. 1996;21(4):134–40.
2. Madsen P, Rasmussen HH, Leffers H, Honoré B, Dejgaard K, Olsen E, Kiil J, Walbum E, Andersen AH, Basse B, et al. Molecular cloning, occurrence, and expression of a novel partially secreted protein “Psoriasin” that is highly up-regulated in psoriatic skin. J Invest Dermatol. 1991;97(4):701–12.
3. Eckert RL, Broome A-M, Ruse M, Robinson N, Ryan D, Lee K. S100 proteins in the epidermis. J Invest Dermatol. 2004;123(1):23–33.
4. Algermissen B, Sitzmann J, LeMotte P, Czarnetzki B. Differential expression of CRABP II, psoriasin and cytokeratin 1 mRNA in human skin diseases. Arch Dermatol Res. 1996;288(8):426–30.
5. Glaser R, Meyer-Hoffert U, Harder J, Cordes J, Wittersheim M, Kobliakova J, Folster-Holst R, Proksch E, Schroder JM, Schwarz T. The antimicrobial protein Psoriasin (S100A7) is upregulated in atopic dermatitis and after experimental skin barrier disruption. J Invest Dermatol. 2009;129(3):641–9.
6. Donato R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. Int J Biochem Cell Biol. 2001;33(7):637–68.
7. Salama I, Malone PS, Mihaimed F, Jones JL. A review of the S100 proteins in cancer. Ejsso. 2008;34(4):357–64.
8. Boniface K, Bernard F-X, Garcia M, Gurney AL, Lecron J-C, Morel F. IL-22 inhibits epidermal differentiation and induces Proinflammatory gene expression and migration of human keratinocytes. J Immunol. 2005;174(6):3695–702.
9. Wolk K, Kunz S, Witte E, Friedrich M, Asadullah K, Sabat R. IL-22 increases the innate immunity of tissues. Immunity. 2004;21(2):241–54.
10. Guttman-Yassky E, Lowes MA, Fuentes-Duculan J, Zaba LC, Cardinale I, Nogralas KE, Khatcherian A, Novitskaya I, Carucci JA, Bergman R, et al. Low expression of the IL-23/Th17 pathway in atopic dermatitis compared to psoriasis. J Immunol. 2008;181(10):7420–7.
11. Leclerc E, Fritz G, Vetter SW, Heizmann CW. Binding of S100 proteins to RAGE: an update. Biochimica et Biophysica Acta - Molecular Cell Research. 2009;1793(6):993–1007.
12. Zackular JP, Chazin JW, Skaar EP. Nutritional immunity: S100 proteins at the host-pathogen interface. J Biol Chem. 2015;290(31):18991–8.
13. Hoffmann HJ, Olsen E, Etzerodt M, Madsen P, Thøgersen HC, Kruse T, Celis JE. Psoriasin binds calcium and is upregulated by calcium to levels that resemble those observed in normal skin. J Invest Dermatol. 1994;103(3):370–5.
14. Tan J, Vorum H, Larsen CG, Madsen P, Rasmussen HH, Gesser B, Etzerodt M, Honoré B, Celis JE, Thestrup-Pedersen K. Psoriasin: a novel chemotactic protein. J Invest Dermatol. 1996;107(1):5–10.
15. Schröder JM, Harder J. Antimicrobial skin peptides and proteins. Cellular and Molecular Life Sciences CMLS. 2006;63(4):469–86.
16. Gläser R, Harder J, Lange H, Bartels J, Christophers E, Schroder JM. Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. Nat Immunol. 2005;6(1):57–64.
17. Lee KC, Eckert RL. S100A7 (Psoriasin) - mechanism of antibacterial action in wounds. J Invest Dermatol. 2007;127(4):945–57.
18. Mischke D, Korge BP, Marenholz I, Volz A, Ziegler A. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex (“epidermal differentiation complex”) on human chromosome 1q21. J Invest Dermatol. 1996;106(5):989–92.
19. Zimmer DB, Eubanks JO, Ramakrishnan D, Criscitiello MF. Evolution of the S100 family of calcium sensor proteins. Cell Calcium. 2013;53(3):170–9.
20. Kulski JK, Lim CP, Dunn DS, Bellgard M. Genomic and phylogenetic analysis of the S100A7 (psoriasin) gene duplications within the region of the S100 gene cluster on human chromosome 1q21. J Mol Evol. 2003;56(4):397–406.
21. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, et al. Genome of the marsupial

Monodelphis domestica reveals innovation in non-coding sequences. Nature. 2007;447:167.

22. Kwek JHL, Wynne A, Lefèvre C, Familiari M, Nicholas KR, Sharp JA. Molecular evolution of a novel marsupial S100 protein (S100A19) which is expressed at specific stages of mammary gland and gut development. Mol Phylogenet Evol. 2013;69(1):4–16.
23. Wolf R, Voscopoulos CJ, FitzGerald PC, Goldsmith P, Cattaïsson C, Gnsior M, Walz M, Ruzicka T, Yuspa SH. The mouse S100A15 Ortholog parallels genomic organization, structure, gene expression, and protein-processing pattern of the human S100A7/A15 subfamily during epidermal maturation. J Invest Dermatol. 2006;126(7):1600–8.
24. Hahn Y, Jeong S, Lee B. Inactivation of MOXD2 and S100A15A by exon deletion during human evolution. Mol Biol Evol. 2007;24(10):2203–12.
25. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26(19):2462–3.
26. Kemp TS, Kemp T. The origin and evolution of mammals: Oxford University Press on Demand; 2005.
27. Foote M, Hunter JP, Janis CM, Sepkoski JJ. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science. 1999;283(5406):1310–4.
28. Schrago CG, Russo CAM. Timing the origin of New World monkeys. Mol Biol Evol. 2003;20(10):1620–5.
29. Platt RN, Mangum SF, Ray DA. Pinpointing the vesper bat transposon evolution using the *Miniopterus natalensis* genome. Mob DNA. 2016;7:12.
30. Ruedi M, Mayer F. Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic Ecomorphological convergences. Mol Phylogenet Evol. 2001;21(3):436–48.
31. Lack JB, Roehrs ZP, Stanley JCE, Ruedi M, Van Den Bussche RA. Molecular phylogenetics of *Myotis* indicate familial-level divergence for the genus *Cistugo* (Chiroptera). J Mammal. 2010;91(4):976–92.
32. Marenholz I, Heizmann CW, Fritz G. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). Biochem Biophys Res Commun. 2004;322(4):1111–22.
33. Shang X, Cheng H, Zhou R. Chromosomal mapping, differential origin and evolution of the S100 gene family. Genetics, selection, evolution : GSE. 2008;40(4):449–64.
34. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
35. Fritz G, Heizmann CW. 3D structures of the calcium and zinc binding s100 proteins. Encyclopedia of Inorganic and Bioinorganic Chemistry 2011.
36. Corbet GB, Hill JE: A world list of mammalian species: Natural History Museum publications; 1991.
37. Pavlovich SS, Lovett SP, Koroleva G, Guito JC, Arnold CE, Nagle ER, Kulcsar K, Lee A, Thibaud-Nissen F, Hume AJ, et al. The Egyptian *Rousettus* genome reveals unexpected features of bat antiviral immunity. Cell. 2018;173(5):1098–1110.e1018.
38. Simmons N: Order Chiroptera. Mammal species of the world: A taxonomic and geographical reference Smithsonian Institution Press, Washington, DC 2005, 1.
39. Hooper SR, Van den Bussche RA. Molecular phylogenetics of the chiropteran family Vespertilionidae. Acta Chiropt. 2003;5:1–59.
40. Handunnethi L, Ramagopalan SV, Ebers GC, Knight JC. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes Immun. 2010;11(2):99–112.
41. Schwensow N, Fietz J, Dausmann KH, Sommer S. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. Heredity. 2007;99(3):265–77.
42. Schad J, Voigt CC, Greiner S, Dechmann DKN, Sommer S. Independent evolution of functional MHC class II DRB genes in New World bat species. Immunogenetics. 2012;64(7):535–47.
43. Jiao H, Wang Y, Zhang L, Jiang P, Zhao H. Lineage-specific duplication and adaptive evolution of bitter taste receptor genes in bats. Mol Ecol. 2018.
44. Jia G, Zhang Y, Wu T, Zhang S, Wang Y. Fruit bats as a natural reservoir of zoonotic viruses. Chin Sci Bull. 2003;48(12):1179–82.
45. Wong S, Lau S, Woo P, Yuen KY. Bats as a continuing source of emerging infections in humans. Rev Med Virol. 2007;17(2):67–91.
46. Schaefer J, Perkins SL, Decher J, Leendertz FH, Fahr J, Weber N, Matuschewski K. High diversity of west African bat malaria parasites and a tight link with rodent *Plasmodium* taxa. Proc Natl Acad Sci. 2013;110(43):17415–9.
47. Mühldorfer K. Bats and bacterial pathogens: a review. Zoonoses Public Health. 2013;60(1):93–103.

48. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, Mills JN, Timonin ME, Willis CKR, Cunningham AA, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc R Soc B-Biol Sci.* 2013;280(1756):9.
49. Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. *Nature.* 2017;546(7660):646.
50. Schad J, Voigt CC. Adaptive evolution of virus-sensing toll-like receptor 8 in bats. *Immunogenetics.* 2016;68(10):783–95.
51. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 2010;11:97.
52. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. *J Genet.* 2013;92(1):155–61.
53. Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol.* 2005;5:28.
54. Lan X, Pritchard JK. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science.* 2016;352(6288):1009–13.
55. Zahn LM. Evolutionary maintenance of gene duplications. *Science.* 2016; 352(6288):949–51.
56. Rogozin IB. Complexity of gene expression evolution after duplication: protein dosage rebalancing. *Genetics Research International.* 2014;2014:8.
57. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A.* 1997;94(15):7799–806.
58. Karev GP, Wolf YI, Koonin EV. Birth and death models of genome evolution. In: *Power Laws, scale-free networks and genome biology.* Boston, MA: Springer US; 2006. p. 65–85.
59. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 2005;39:121–52.
60. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673–80.
61. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. In: *Nucleic acids symposium series: 1999.* 95–98.
62. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
63. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993;10(3):512–26.
64. Neves F, Águeda-Pinto A, Pinheiro A, Abrantes J, Esteves PJ. Strong selection of the TLR2 coding region among the Lagomorpha suggests an evolutionary history that differs from other mammals. *Immunogenetics.* 2019.
65. Huang Q, Wu Y, Qin C, He W, Wei X. Phylogenetic and structural analysis of the phospholipase A2 gene family in vertebrates. *Int J Mol Med.* 2015;35(3):587–96.
66. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21(10):2531–3.
67. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 2015;32(3):820–32.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

