Expansion, retention and loss in the Acyl-CoA Synthetase *"Bubblegum"* (*Acsbg*) gene family in vertebrate history

Mónica Lopes-Marques[1*], André M. Machado[1], Raquel Ruivo[1], Elza Fonseca[1,2], Estela Carvalho[1], and L. Filipe C. Castro[1,2*]

[1]Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto (U.Porto), Matosinhos, Portugal

[2]Faculty of Sciences (FCUP), Department of Biology, University of Porto (U.Porto), Porto, Portugal

* Address correspondence to:

Mónica Lopes-Marques & Luís Filipe Costa Castro
Interdisciplinary Centre for Marine and Environmental Research (CIIMAR)
Av. General Norton de Matos s/n, 4450-208 Matosinhos, Portugal
Tel.: +351223401800
Fax: +351223390608
Email: monicaslm@hotmail.com; filipe.castro@ciimar.up.pt

Keywords: Acyl-CoA synthetases; fatty acid activation; ACSBG; gene duplication; vertebrates

ABSTRACT

Fatty acids (FAs) constitute a considerable fraction of all lipid molecules with a fundamental role in numerous physiological processes. In animals, the majority of complex lipid molecules are derived from the transformation of FAs through several biochemical pathways. Yet, for FAs to enroll in these pathways they require an activation step. FA activation is catalyzed by the rate limiting action of Acyl-CoA synthases. Several Acyl-CoA enzyme families have been previously described and classified according to the chain length of FA they process. Here, we address the evolutionary history of the ACSBG gene family which activates, FA with more than 16 carbons. Currently, two different ACSBG gene families, *ACSBG1* and *ACSBG2*, are recognized in vertebrates. We provide evidence that a wider and unequal *ACSBG* gene repertoire is present in vertebrate lineages. We identify a novel *ACSBG-like* gene lineage which occurs specifically in amphibians, ray finned fish, coelacanths and chondrichthyes named *ACSBG3*. Also, we show that the *ACSBG2* gene lineage duplicated in the Theria ancestor. Our findings, thus offer a far richer understanding on FA activation in vertebrates and provide key insights into the relevance of comparative and functional analysis to perceive physiological differences, namely those related with lipid metabolic pathways.

1. Introduction

Lipids represent a complex group of biomolecules present in all living organisms, playing a key role in numerous biological processes, such as inflammatory response, reproduction, biological membranes and energy sourcing and storage. Additionally, they participate in the overall the homoeostasis as signal molecules, cofactors, and endogenous ligands for nuclear receptors (Wall et al., 2010; Robinson and Mazurak, 2013; Grygiel-Górniak, 2014). Aside from sterol lipids, all remaining lipids are obtained from the endogenous elaboration of fatty acid (FA) molecules (Watkins et al., 2007). Yet, for FAs to enroll in any anabolic and catabolic process they require an activation step. Thus, FA activation is a critical rate limiting step of FA metabolism. FA activation was first recognized in 1948 and referred to as “sparking” or “priming” (Grafflin and Green, 1948; Knox et al., 1948). This enzymatic step consists of a two-step thioesterification reaction catalyzed by Acyl-CoA synthetase (ACS), resulting in a thioester with coenzyme A (CoA) (Watkins et al., 2007).

Several ACS involved in FA activation have been previously identified and organized according to the degree of unsaturation and chain length of the FAs favored as substrate: the short-chain ACS-Family (*ACSS*), medium-chain ACS-Family (*ACSM*), long-chain ACS-Family (*ACSL*), very long-chain ACS-Family (*ACSVL*), Bubblegum ACS-Family (*ACSBG*) and ACSFamily (*ACSF*) (Watkins et al., 2007; Soupene and Kuypers, 2008). Although some substrate preference overlap is observed, these enzymes also differ in tissue distribution and subcellular location, an indication of their highly specific role in FA metabolism (Watkins et al., 2007). ACS enzymes have been found to have a wide taxonomic distribution, with homologues ranging from Eubacteria to Plants and Metazoa, a clear indication of their pivotal role in lipid metabolism (Hisanaga et al., 2004; Soupene and Kuypers, 2008).

Despite their wide taxonomic occurrence, the genetic repertoire of ACS has been found to vary, namely in vertebrates (Castro et al., 2012; Lopes-Marques et al., 2013). For example, some studies have disclosed that both ACSL and ACSS gene family composition and function were shaped by events of gene/genome duplication in combination with differential loss. Moreover, multi-genome comparisons across a wide range of vertebrate species revealed novel and previously uncharacterized ACS enzymes (Castro et al., 2012; Lopes-Marques et al., 2013). The present work seeks to build on previous findings and further extend, the knowledge regarding the genetic repertoire and distribution ACS in vertebrates namely the ACS *Bubblegum* (ACSBG) gene family.

ACSBG enzymes, also known as lipidosin, activate FA with C16 to C24 (Moriya-Sato et al., 2000; Steinberg et al., 2000; Pei et al., 2003). Presently, 2 members of the ACSBG gene family have been identified and characterized in mammals, ACSBG1 and ACSBG2 (Pei et al., 2003; Watkins et al., 2007). Similarly, to the previously described ACS enzymes, both ACSBG members display conserved sequence motifs, such as the putative ATP-AMP signature motif for ATP binding (Motif I) and a motif for FA binding, characteristic of the ACS gene family (Motif II) (Moriya-Sato et al., 2000; Watkins et al., 2007). Notably, all known ACS, with the exception of human ACSBG2, contain a highly conserved arginine (Arg-R) in Motif II (Pei et al., 2006). The replacement of this Arg by histidine (His-H) in Human ACSBG2 was found to confer a biphasic pH optimum (pH 6.5 and pH 7.5) to the enzyme, in contrast to the monophasic activity at pH 7-7.5 of the mouse orthologue (Pei et al., 2006). Yet, due to the degree of conservation of both Motifs I and II, these have previously been used to seek and identify potential ACS enzymes (Steinberg et al., 2000; Watkins et al., 2007).

ACSBG enzymes have been suggested to play a significant role in brain development and reproduction (Moriya-Sato et al., 2000; Tang et al., 2001; Pei et al., 2006). Previous reports with the *D. melanogaster* bubblegum mutant (termed bubblegum due to the bubbly appearance of the lamina, a result of neurodegeneration and dilation of the photoreceptor axons) and mouse, associated the disruption of *ACSBG1* to X-linked adrenoleukodytrophy (X-ALD) (Min and Benzer, 1999; Moriya-Sato et al., 2000). X-ALD is characterized by the accumulation of high levels of very long FA in plasma and tissues, accompanied by neurodegeneration (Min and Benzer, 1999; Moriya-Sato et al., 2000; Moser et al., 2002). On the other hand, *ACSBG2* plays an important role in spermatogenesis and testicular development, being associated to male infertility (Zheng et al., 2005; Fraisl et al., 2006). In agreement gene expression of *ACSBG1* is found to be mainly restricted to brain, adrenal gland, gonads, spleen in mouse and human (Moriya-Sato et al., 2000). In contrast, *ACSBG2* showed a more exclusive expression pattern being highly expressed in the testis, followed by medulla and spinal cord (Pei et al., 2006) .

Here, using a combination of phylogenetics, comparative genomics and gene expression analysis we deduced the evolutionary history of the *ACSBG* gene family in all major vertebrate lineages. Our findings illustrate the importance of comparative analysis to address the role of adaptive evolution in the shaping of lipid metabolic modifications between lineages.

2. Materials & Methods

2.1. Database search and phylogenetic analysis

NCBI GenBank release 220 June and release 221 August 2017 and Ensembl release 89 May and release 90 August 2017, databases were searched using tblastn and blastp to recover ACSBG-like sequences using human ACSBG1 (NP_055977) and ACSBG2 (NP_112186) amino acid sequences as query. All major vertebrate lineages such as mammals, birds, reptiles, amphibians, coelacanths, teleost fish, cartilaginous fish and cyclostomes were searched. Additionally, the following invertebrate lineages basal to chordates were also explored cephalochordates, hemichordates, Mollusca (Supplementary material 1).

Our search retrieved 121 ACSBG-like sequences; sharing a minimum 70% pairwise identity with corresponding query sequence for mammals; 60% for bird's reptiles and amphibians; 50% identity for teleost's and chondrichthyes and finally 40% identity for invertebrates. The collected sequences were aligned and inspected with partial sequences being removed, leaving a 119 full ORF or near full ORF sequences for phylogenetic analysis (Supplementary material 1). Amino acid sequences were aligned in MAFFT with the L-INS-I method (Katoh et al., 2005; Katoh and Toh, 2008). In the resulting alignment, all columns containing 90% gaps were stripped leaving a total of 787 positions for phylogenetic analysis. A second sequence alignment containing 121 ACSBG-like sequences including the truncated sequences of *Xenopus tropicalis* and *Xenopus leavis* was performed using the same method leaving a total of 788 positions for phylogenetic analysis. Both alignments were then individually submitted to PhyML3.0 server (Guindon et al., 2010), with evolutionary model determined automatically, resulting in the selection JTT+G+I in both cases. The branch support for phylogenetic trees was calculated using aBayes. The resulting trees were visualized and edited in Fig. Tree V1.3.1 available at http://tree.bio.ed.ac.uk/software/figtree/ and rooted with the invertebrate sequences.

2.2. Synteny and Paralogy analysis

Using as reference the human and teleost *ACSBG loci*, synteny maps of the genomic neighbourhoods of the *ACSBG1, ACSBG2* and *ACSBG3* gene were assembled in a set of species representative of the major lineages analysed. The following genome assemblies available in NCBI were accessed for *Homo sapiens* - GCF_000001405.33, *Monodelphis domestica* - GCF_000002295.2, *Gallus gallus* - GCF_000002315.4, *Pelodiscus sinensis* - GCF_000230535.1, *X. tropicalis* - GCF_000004195.3, *X. laevis* - GCF_001663975.1, *Latimeria chalumnae* - GCF_000225785.1, *Oryzias latipes* - GCA_000313675.1, *Astyanax mexicanus* - GCA_000372685.2, *Danio rerio* - GCA_000002035.4, *Lepisosteus oculatus* - GCA_000242695.1, *Callorhinchus milii* - GCA_000165045.2, *Branchiostoma floridae* - GCA_000003815.1 and *S. kowalevskii* - GCF_000003605.2. For *Tetraodon nigroviridis* and *Petromyzon marinus* genome assemblies TETRAODON 8.0, Mar 2007 and Pmarinus_7.0, Jan 2011 available in Ensembl were accessed. Paralogy analysis of the *ACSBG locus* was conducted using the reconstructed ancestral chordate genome as described in (Putnam et al., 2008).

2.3. RNA ISOLATION AND ASCBG TISSUE EXPRESSION PANEL IN *X. TROPICALIS*

X. tropicalis (African clawed frog) tissues (brain, skin, heart, liver, spleen, pancreas, kidney, intestine, testis and ovary) were kindly provided by O. Brochain (CNRS, Orsay). Total RNA was purified using the Illustra RNAspin Mini RNA Isolation Kit animal tissues protocol (GE Healthcare) with on-column DNase I digestion. RNA quality was assessed by electrophoresis and its concentration determined using a microplate spectrophotometer (Take 3 and Synergy HT Multi-Mode Microplate Reader, Biotek). First-strand cDNA was synthesized from 250ng RNA using the iScriptTM cDNA Synthesis Kit (Bio-Rad), according to the manufacturer's instructions.

Forward and reverse primers sets were designed to flank an intron and to avoid genomic DNA amplification. Primers sets were created for the following genes *ACSBG1* - Forward-5'TTTGCCAGGATGTTGGAAGT3', Reverse-5'AAAGCTTCCACGTGCTCTGT 3', annealing at 57°; *ACSBG2* - Forward-5' CTTTTCTGGGGACGTCATGT 3', Reverse-5' TTGGAACCTGCTCTTTGAGG 3', annealing at 55° and *ASCGB3* - Forward-5' TGCAGTCTTTGCTACGTTGG 3' reverse-5' ACAAACAGAGCTCCCCTGTG 3', annealing at 57°. To assess the quality of *X. tropicalis* cDNA two sets of primers targeting housekeeping genes were included for β-actin – Forward-5' GGTCGCCCAAGACATCAG3´, Reverse-5´GCATACAGGGACAACACA annealing at 57º and for EEF1A1 – Forward-

5´TCGTTAAGGAAGTCAGCACA3´ and Reverse5´CATGGTGCATTTCAACAGAT3´ annealing at 57º. PCR reactions were all performed using 2 µl of *X. tropicalis* cDNA and Phusion® Flash high-fidelity Master Mix (FINNZYMES). PCR parameters were as follows: initial denaturation at 98°C for 10 s, followed by 30 cycles of denaturation at 98°C for 1 s, annealing for 5 s and elongation at 72°C for 10s and a final step of elongation at 72°C for 1 min. PCR products were then loaded onto 2% agarose gel stained with GelRed and run in TBE buffer at 80 V.

2.4. ACSBG EXPRESSION ANALYSIS THROUGH RNA-SEQ

The RNA-Seq analysis was performed using a collection of tissues datasets from seven species Human (*H. sapiens*), mouse (*M. musculus*), chicken (*G. gallus*), western clawed frog (*X. tropicalis*), zebrafish (*D. rerio*), spotted gar (*L. oculatus*) and elephant shark (*C. milii*), available in National Centre for Biotechnology (NCBI) Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/) (Supplementary material 2). To standardize datasets from different sources, all files were converted to FASTQ file format and sequence quality trimming was performed using Trimmomatic v 0.36 (Bolger et al., 2014). Reads with 36bp in length and an average score of 20 phred were selected for further analysis.

Reference sequences and respective annotation files of each specie were collected from NCBI and Ensembl (Release 89) (Yates et al., 2016) (supplementary material 3). For both elephant shark and western clawed frog the reference sequences for RNAseq mapping were retrieved from NCBI (ftp://ftp.ncbi.nih.gov/genomes/refseq/), while for the remaining species the reference sequences for mapping were retrieved from Ensembl database (ftp://ftp.ensembl.org/pub/release-89/) (Supplementary material 3). Trimmed and groomed reads from each dataset were mapped to their respective reference using Bowtie2 (Langmead and Salzberg, 2012), and the transcript quantification was calculated in transcript per million (TPM), with RSEM v.1.2.31 software (Li and Dewey, 2011). TPM values for each gene were taken as evidence of relative gene expression, low TPM values (< 0.5) were considered unreliable and substituted with zero. To complete this exploratory gene expression analysis, the TPM values were $\log_2$-transformed after adding a value of one.

3. Results and discussion

3.1. Database mining and phylogenetic analysis reveals novel members ACSBG gene family

Initial blast searches identified ACSBG-like sequences and recovered a larger than anticipated number of sequence hits. ACSBG1 and ACSBG2-like sequences were found in species from the following vertebrate lineages: mammals, birds, reptiles, amphibians, holostei, coelacantiforms and teleostei. In chondrichthyans and cyclostomes no ACSBG1-like sequences were retrieved. Moreover, an additional uncharacterized ACSBG2-like sequence was also retrieved in some mammalian species. Database searches also recovered a novel set of ACSBG-like sequences in four gnathostome lineages: amphibians (Western clawed frog and African clawed frog) chondrichthyans (elephant shark), holostei (spotted gar), coelacantiforms (coelacanth) and teleostei. However, amphibian sequences were considerably shorter than the remaining ACSBG, thus being excluded from the main phylogenetic analysis.

To disclose the orthology of these various sequences a phylogenetic analysis was performed. The resulting tree topology displays 3 well supported clades in vertebrates. The first group contained all ACSBG1 sequences from mammals, reptiles, birds, amphibians, coelacanths and teleost fish, with no representatives of chondrichtyes and cyclostomes. Besides the ACSBG1 clade, we find a sister clade comprising ACSBG2 sequences. This contains ACSBG2 previously described in mammals (ACSBG2a) and an uncharacterized ACSBG2-like (ACSBG2b) identified in the present work. The tree topology suggests that the both mammalian ACSBG2 sequences are related by a duplication event that took place in the ancestor of Theria. Out grouping the mammalian ACSBG2 sequences we find the ACSBG2 sequences from birds, reptiles, amphibians, coelacanths, chondrichthyes, teleost fish and cyclostomes. Thus, *bona fide* orthologues of ACSBG2 are represented across all major vertebrate classes. Within the third clade, we find a novel uncharacterized group of ACSBG sequences which have sequence representatives in coelacanths, teleost fish, holostei and chondrichthyes. We name this novel sequence ACSBG3. Finally, placed basally to all vertebrate sequences we find invertebrate ACSBG sequences. The present tree topology provides robust indications that the diversification of the ACSBG gene family occurred in the vertebrate ancestor. A second phylogenetic analysis was run separately to include the uncharacterized short *Xenopus sp.* ACSBG sequences (Supplementary material 4). Although the overall tree topology is conserved with the main analysis (Fig. 1), we find that the *Xenopus sp.* ACSBG sequences are placed basally to all vertebrate clades hindering the identification of their orthology. This placing of *Xenopus sp*. ACSBG-like sequences correlates to the highly divergent nature observed in the sequence alignment, with these amphibian sequences being considerably shorter and displaying a poor conservation of the AMP-binding motif (see section 3.3).

Additionally, we find that mammalian ACSBG2a and 2b sequences are placed in long branches in both phylogenetic analysis (Fig.1 and Supplementary material 4) suggesting an accelerated evolution and divergence of these sequences further analysed in section 3.3.

3.2. A NOVEL ACSBG GENE, ACSBG3, IS AN OHNOLOG GONE MISSING IN AMNIOTES

Phylogenetic analysis suggests that the ACSBG gene family expanded in the vertebrate ancestor. This time frame coincides with the proposed timing of two round of whole genome duplication (2R WGD) in the ancestral vertebrate approximately 500MYA (Ohno, 1970; Dehal and Boore, 2005). Yet, while it is generally accepted that all gnathostomes underwent the 2R-WGD, the extent of these genome duplications in cyclostomes still remains a matter of debate (Smith and Keinath, 2015).

To complement the phylogenetic analysis, validate events of gene duplication/loss, resolve the orthology of the *Xenopus sp.* uncharacterized ACSBG and the origin of ACSBG3 sequences, the genomic *locus* of each *Acsbg* gene was examined in a set of representative species (Fig. 2). Comparative synteny analysis of the ACSBG1 *locus* reveals a high degree of conservation of neighbouring gene families throughout all the analysed lineages (Fig.2A). *ACSBG1* is localized in human chromosome 15, being flanked by gene families such as the *IDH3A, CIB2, WDR61* and *CRABP1*. These flanking gene families are also present in the vicinity of ACSBG1 *locus* in the all analysed lineages (Fig. 2A). In the case of *C. milii* although no *ACSBG1* gene was found, synteny analysis reveals that the *locus* organization is conserved, suggesting gene loss in this lineage (Fig. 2A). Regarding the cyclostomes, extensive blast searches did not retrieve an ACSBG1 sequence; synteny analysis uncovered a fragmented *locus* segregated into at least two distinct scaffolds. Therefore, the absence of this gene in cyclostomes may be attributed to poor genome coverage or to gene loss (Fig. 2A).

The human ACSBG2 gene resides in chromosome 19 and is flanked by the following gene families: *RFX2, RANBP3, MLLT1* and *ACER1.* The *locus* architecture is conserved in all species analysed. In cyclostomes, the *ACSBG2 locus* is disjointed and distributed among several scaffolds thus the absence of ACSBG2 in cyclostomes remains similarly to ACSBG1 unresolved (Fig. 2B). In the case of mammalian duplicates, we find that *ACSBG2a* and *ACSBG2b* are located side by side in the *M. domestica.* Synteny analysis of this *locus* in other mammals presenting both *ACSBG2a* and *2b* (data not shown) is coincident with the observation for *M. domestica* supporting the hypothesis that *ACSBG2a* and *2b* arose through tandem duplication in the ancestor of therian mammals, with ACSBG2b being later lost in Haplorhini. Finally, we find that in the *ACSBG3 locus*, despite the lesser conservation, some neighbouring gene families such as *HINT2, SPAG8, RGP1* and *GBA2* are preserved in the majority of the analysed lineages. Using these conserved neighbouring gene families, the corresponding *locus* was mapped in birds and mammals to address the loss of *ACSBG3* in these lineages (Fig. 2C). ACSBG3 is also absent in reptiles and the analysis of the *locus* revealed that it is fragmented in various species examined (*Anolis carolinesis*, *Thamnophis sirtalis*, *Alligator mississippiensis* and *Chrysemys picta*), hindering the validation of ACSBG3 loss in this lineage.

We next investigated the synteny maps for the single ACSBG *locus* from two invertebrate cephalochordates. Here we find that the *B. floridae locus* retains a conserved gene family arrangement, namely with the presence of *HERC1-like* gene, whose orthologue is found in the vicinity of vertebrate *ACSBG1* (Fig. 2 A and D indicated in red). Similarly, the hemichordate *S. kowalevskii* also displays a conserved neighbouring gene family, *CHRNA3* with the vertebrate orthologue placed in the *ACSBG1 locus* (Fig. 2A and D indicated in red). Finally, to address the hypothesis that ACSBG gene expansion took place with the 2R WGD the location of ACSBG and neighbouring genes (with described paralogues underlined genes in Fig.2A, B, C and D) were mapped to the predicted ancestral paralogons as described by Putnam *et al* 2008 (Putnam et al., 2008). Next, ancestral paralogons were mapped back to the same ancestral linkage group, LG2 (Putnam et al., 2008) indicating that the *ACSBG loci* are related by genome duplication, strongly suggesting that vertebrate ACBG diversity arose with the 2R WGD (Fig. 2 F).

3.3. Sequence analysis and gene expression

To further characterize the novel *ACSBG2b* and *ACSBG3* a sequence alignment was performed to inspect the typical ACS enzyme motifs (Watkins et al., 2007). The analysis of this alignment revealed that the predicted AMP-binding domain (Motif I Fig. 3A and 3B), a highly conserved motif in all ACS enzymes from bacteria to humans (Black et al., 1997; Steinberg et al., 2000; Weimar et al., 2002; Karan et al., 2003), is conserved in the vast majority of the sequences collected with the exception of the novel ACSBG2b (Fig.3B) and the ACSBG3 sequence in *Xenopus sp.*(see Supplementary material 5 for alignment of the full 121 sequences). An indication that residues within Motif I play a fundamental role in ACS catalytic activity was found in previous studies were the mutation of residues within Motif I (positions 1, 2, 4, 5 and 10) in *E. coli* considerably reduced the catalytic activity, while the replacement of residues 1 and 5 in *S. cerevisiae* resulted in a minor reduction of enzymatic activity (Fig. 3A grey arrows) (Weimar et al., 2002; Zou et al., 2002). Thus, the low conservation of this motif in the mammalian-specific ACSBG2b strongly suggests that these enzymes may show an alternative function or *modus operandi*. In the analysis of the *Xenopus sp.* ACSBG3, we find that this motif differs from the remaining ACSBG3 identified, being disrupted with the deletion of 3bp (Supplementary material 5). Again, this observation suggests an alternative role for the enzyme given that AMP binding is essential for FA activation. Regarding Motif II, also known as the ACS signature-motif and proposed to be involved in acyl chain length specificity (Black et al., 1997), we find that again ACSBG2b displays a divergent sequence when compared to the remaining ACSBG analysed here. Interestingly, we observe that ACSBG2b presents an Asparagine residue (Asn-N) instead of the highly conserved Arginine (Fig. 3A and B black arrow). Notably, human ACSBG2a harbours a Histidine in this position, representing the single case described to date of an ACSBG without an Arginine (Pei et al., 2006). Reverse mutation of the Histidine within Motif II in human ACSBG2a showed that this residue assumes a critical role in determining the optimal pH for this enzyme (Pei et al., 2006). Additionally, this replacement (Asn) is only observed for placental mammals, with marsupials retaining the conserved Arginine (Fig 3 B). Next, Motif V (KXX(R,K) is a conserved motif found in several members of the ACS enzymes families and contains a conserved K- Lys demonstrated to be essential for the catalytic function of ACS in *S. enterica* propionyl-coA synthetase and ACS activity of murine ACSF2 (Horswill and Escalante-Semerena, 2002). Here we find that Motif V is conserved in all recovered ACSBG sequences with the exception of *Xenopus sp* ACSBG3 due to the short nature of these sequences (see Supplementary material 5). Finally, the Motifs III and IV, identified by Hisanaga *et al* 2008 (Hisanaga et al., 2004), are found to be conserved in the majority of analysed sequences, with the exception of a conservative replacement in Motif IV in ACSBG2b. The highly conserved Histidine is replaced by biochemically similar residue, tyrosine, having a minor or no predicted impact.

In an attempt to infer the function of the newly identified *ACSBG2b* and *ACSBG3*, and address the retention of these genes in a restricted number of lineages, we next performed an expression analysis using available RNA-Seq SRAs (Fig. 3C). Similarly, to previous reports (Moriya-Sato et al., 2000; Tang et al., 2001; Pei et al., 2006), relative expression profiles reveal that *ACSBG1* expression is mainly limited to brain and gonads, with the exception of *D. rerio* for which the liver stands as the main expression site. On the other hand, the expression profile of non-mammalian vertebrate *ACSBG2* was found to be more extensive than in mammals (Pei et al., 2006) with expression detected in all analysed tissues of *C. milii*, *D. rerio*, *L. oculatus*, *G. gallus*, and *X. tropicalis*. Interestingly, the expression analysis of mammalian specific duplicates *ACSBG2a* and *ACBG2b* shows a confined expression of the duplicates essentially in testis, with a relatively low expression of *ACSBG2a* detected in human kidney. Regarding the gene expression profile of *ACSBG3 in X. tropicalis, L. oculatus*, and in *ACSBG3a* of *C. milli*, we find a localized and high relative expression in ovary and in testis. Semi-quantitative PCR expression analysis of ACSBG3 from *X. tropicalis* is in accordance with *in silico* RNAseq analysis (Fig. 3D), with ACSBG1 expression confined to brain and testis, while ACSBG2 is detected all tissues except ovary and finally ACSBG3 being restricted to testis and brain (Fig. 3D).

High expression of *ACSBG3* in gonads is indicative that this enzyme may play an important role in reproduction similarly to the role of *ACSBG2* (Pei et al., 2006). Finally, for *ACSBG3* in *C. milii* no expression was detected in any of the analysed tissues.

3.4. Evolutionary history of ACSBG gene family

Using a multi-comparative approach, including database searches, phylogenetic and synteny analysis we have uncovered a larger than anticipated genetic repertoire of *ACSBG* genes in vertebrates. We find that the initial expansion of the *ACSBG* gene family from which arose *ACSBG1 ACSBG2* and *ACSBG3* is coincident with the 2R WGD, with representative gene orthologues present in several gnathostome lineages (Fig. 4). The detailed analysis of the *ACSBG* gene repertoire revealed a differential retention of *ACSBG3,* with this paralogue being lost in birds, mammals and possibly in reptiles, while being retained in teleosts, amphibians and chondrichthyes. The identification of additional ACSBG enzymes in teleosts correlates with previous studies, where differential paralogue retention led to the maintenance of extra ACS enzyme paralogues, namely *ACSL2* and *ACSS1b* in teleosts (Fraisl et al., 2006; Wall et al., 2010). The preservation of duplicated genes is often observed when the corresponding transcript, in this case ACS, is in high demand (Zhang, 2003). Thus, one may hypothesize that the preservation of additional ACS duplicates in teleosts is a means to fulfil a high demand of FA activation given that FA oxidation is considered to be the main energy source in this lineage (Tocher, 2003). Finally, further duplications were observed in the ancestor of mammals with the tandem duplication of *ACSBG2* and also in specific lineages such as the *ACSBG3* in *C. milii* and *T. nigroviridis* and *ACSBG1 L. chalumnae* (Fig. 4).

4. Conclusion

Our findings suggest that FA activation metabolic modules, including the ACSBG gene family, have significantly diversified upon vertebrate radiation as a consequence of genome duplication, lineage specific duplication and losses.

ACKNOWLEDGMENTS

This work was supported by Project INNOVMAR—Innovation and Sustainability in the Management and Exploitation of Marine Resources (reference NORTE-01-0145-FEDER-000035, within Research Line NOVELMAR/INSEAFOOD/ECOSERVICES), supported by North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). We acknowledge Fundação para a Ciência e a Tecnologia for the support to EF (SFRH/BD/79305/2011).

References

Black, P.N., Zhang, Q., Weimar, J.D. and DiRusso, C.C., 1997. Mutational Analysis of a Fatty Acyl-Coenzyme A Synthetase Signature Motif Identifies Seven Amino Acid Residues That Modulate Fatty Acid Substrate Specificity. Journal of Biological Chemistry 272, 4896-4903.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-20.

Castro, L.F.C., Lopes-Marques, M., Wilson, J.M., Rocha, E., Reis-Henriques, M.A., Santos, M.M. and Cunha, I., 2012. A novel Acetyl-CoA synthetase short-chain subfamily member 1 (Acss1) gene indicates a dynamic history of paralogue retention and loss in vertebrates. Gene 497, 249-255.

Dehal, P. and Boore, J.L., 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. PLOS Biology 3, e314.

Fraisl, P., Tanaka, H., Forss-Petter, S., Lassmann, H., Nishimune, Y. and Berger, J., 2006. A novel mammalian bubblegum-related acyl-CoA synthetase restricted to testes and possibly involved in spermatogenesis. Arch Biochem Biophys 451, 23-33.

Grafflin, A.L. and Green, D.E., 1948. Studies on the cyclophorase system; the complete oxidation of fatty acids. J Biol Chem 176, 95-115.

Grygiel-Górniak, B., 2014. Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications - a review. Nutrition Journal 13, 17.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59, 307-321.

Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii, Y., Sugahara, M., Kuramitsu, S., Yokoyama, S. and Miyano, M., 2004. Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. J Biol Chem 279, 31717-26.

Horswill, A.R. and Escalante-Semerena, J.C., 2002. Characterization of the Propionyl-CoA Synthetase (PrpE) Enzyme of Salmonella enterica: Residue Lys592 Is Required for Propionyl-AMP Synthesis. Biochemistry 41, 2379-2387.

Karan, D., Lesbats, M., David, J.R. and Capy, P., 2003. Evolution of the AMP-forming Acetyl-CoA Synthetase Gene in the Drosophilidae Family. Journal of Molecular Evolution 57, S297-S303.

Katoh, K., Kuma, K.-i., Toh, H. and Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33, 511-518.

Katoh, K. and Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9, 286-98.

Knox, W.E., Noyce, B.N. and Auerbach, V.H., 1948. Studies on the cyclophorase system: III. Obligatory Sparking of the fatty acid oxidaton. Journal of Biological Chemistry 176, 117-122.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-9.

Li, B. and Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Lopes-Marques, M., Cunha, I., Reis-Henriques, M.A., Santos, M.M. and Castro, L.F.C., 2013. Diversity and history of the long-chain acyl-CoA synthetase (Acsl) gene family in vertebrates. BMC Evolutionary Biology 13, 271.

Min, K.-T. and Benzer, S., 1999. Preventing Neurodegeneration in the *Drosophila* Mutant *bubblegum*. Science 284, 1985-1988.

Moriya-Sato, A., Hida, A., Inagawa-Ogashiwa, M., Wada, M.R., Sugiyama, K., Shimizu, J., Yabuki, T., Seyama, Y. and Hashimoto, N., 2000. Novel Acyl-CoA Synthetase in Adrenoleukodystrophy Target Tissues. Biochemical and Biophysical Research Communications 279, 62-68.

Moser, H.W., Smith, K.D., Watkins, P.A., Powers, J. and Moser, A.B., 2002. X-linked adrenoleukodystrophy. In The metabolic and molecular bases of inherited disease. McGraw Hill.

Ohno, S., 1970. Evolution by Gene Duplication, 1 ed. Springer Berlin Heidelberg, New York.

Pavlidis, P. and Noble, W.S., 2003. Matrix2png: a utility for visualizing matrix data. Bioinformatics 19, 295-6.

Pei, Z., Jia, Z. and Watkins, P.A., 2006. The second member of the human and murine bubblegum family is a testis- and brainstem-specific acyl-CoA synthetase. J Biol Chem 281, 6632-41.

Pei, Z., Oey, N.A., Zuidervaart, M.M., Jia, Z., Li, Y., Steinberg, S.J., Smith, K.D. and Watkins, P.A., 2003. The acyl-CoA synthetase "bubblegum" (lipidosin): further characterization and role in neuronal fatty acid beta-oxidation. J Biol Chem 278, 47070-8.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutierrez, E., Dubchak, I., Garcia-

Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N. and Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Robinson, L. and Mazurak, V., 2013. N-3 Polyunsaturated Fatty Acids: Relationship to Inflammation in Healthy Adults and Adults Exhibiting Features of Metabolic Syndrome. Lipids 48, 319-332.

Smith, J.J. and Keinath, M.C., 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Research 25, 1081-1090.

Soupene, E. and Kuypers, F.A., 2008. Mammalian long-chain acyl-CoA synthetases. Exp Biol Med (Maywood) 233, 507-21.

Steinberg, S.J., Morgenthaler, J., Heinzer, A.K., Smith, K.D. and Watkins, P.A., 2000. Very Long-chain Acyl-CoA Synthetases: Human "Bubblegum" represents a new family of proteins capable of activating Very Long chain Fatty Acids. Journal of Biological Chemistry 275, 35162-35169.

Tang, P.-Z., Tsai-Morris, C.-H. and Dufau, M.L., 2001. Cloning and characterization of a hormonally regulated rat long chain acyl-CoA synthetase. Proceedings of the National Academy of Sciences 98, 6581-6586.

Tocher, D.R., 2003. Metabolism and Functions of Lipids and Fatty Acids in Teleost Fish. Reviews in Fisheries Science 11, 107-184.

Wall, R., Ross, R.P., Fitzgerald, G.F. and Stanton, C., 2010. Fatty acids from fish: the anti-inflammatory potential of long-chain omega-3 fatty acids. Nutrition Reviews 68, 280-289.

Watkins, P.A., Maiguel, D., Jia, Z. and Pevsner, J., 2007. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. Journal of Lipid Research 48, 2736-2750.

Weimar, J.D., DiRusso, C.C., Delio, R. and Black, P.N., 2002. Functional Role of Fatty Acyl-Coenzyme A Synthetase in the Transmembrane Movement and Activation of Exogenous Long-chain Fatty Acids: Amino Acid residues within the ATP/AMP signature motif of Escherichia coli FadD are required for enzyme activity and Fatty acid transport. Journal of Biological Chemistry 277, 29369-29376.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R. and Flicek, P., 2016. Ensembl 2016. Nucleic Acids Research 44, D710-D716.

Zhang, J., 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution 18, 292-298.

Zheng, Y., Zhou, Z.M., Min, X., Li, J.M. and Sha, J.H., 2005. Identification and characterization of the BGR-like gene with a potential role in human testicular development/spermatogenesis. Asian J Androl 7, 21-32.

Zou, Z., DiRusso, C.C., Ctrnacta, V. and Black, P.N., 2002. Fatty acid transport in Saccharomyces cerevisiae. Directed mutagenesis of FAT1 distinguishes the biochemical activities associated with Fat1p. J Biol Chem 277, 31062-71.

Figure captions

Figure 1: Maximum likelihood phylogenetic analysis of ACSBG amino acid sequences rooted with the invertebrate clade. Numbers at nodes indicate posterior probabilities calculated using aBayes.

Figure 2: Comparative genomic maps of vertebrate *ACSBG1* (A) *ACSBG2* (B) and *ACSBG3* (C and D) gene *loci.* Paralogy analysis and invertebrate genomic maps of ACSBG (E and F).

Figure 3: Sequence alignment and ACS Motif analysis. **A**- Sequence logo graphs of the consensus sequences of all ACSBG sequences recovered excluding mammal specific ACSBG2b acyl-coenzyme A sequences, totalizing 104 sequences. **B**- Sequence logo graphs of the consensus sequences of all mammalian specific ACSBG2b (17 sequences). Overall height of the stack reflects the degree of conservation, the height of each letter represents relative frequency of a given residue in a specific position. Black arrow highlights the highly conserved arginine residue (Pei et al., 2006) and corresponding position in ACSBG2b **C**- Heatmap of the relative expression of *ACSBG1 ACSBG2* and *ACSBG3* obtained from RNA-seq analysis and visualized using Matrix2png (Pavlidis and Noble, 2003). **D-** Tissue expression profile of *X. tropicalis* ACSBG genes.

Figure 4: Proposed evolutionary history of the *ACSBG* gene family in vertebrates. Yellow corresponds to *ACSBG1* blue *ACSBG2* and green *ACSBG3*. Grey full lined circles with question marks indicate unknown or unresolved if gene is present, grey circles dashed lined indicate gene loss. Black line under genes indicates tandem duplication.

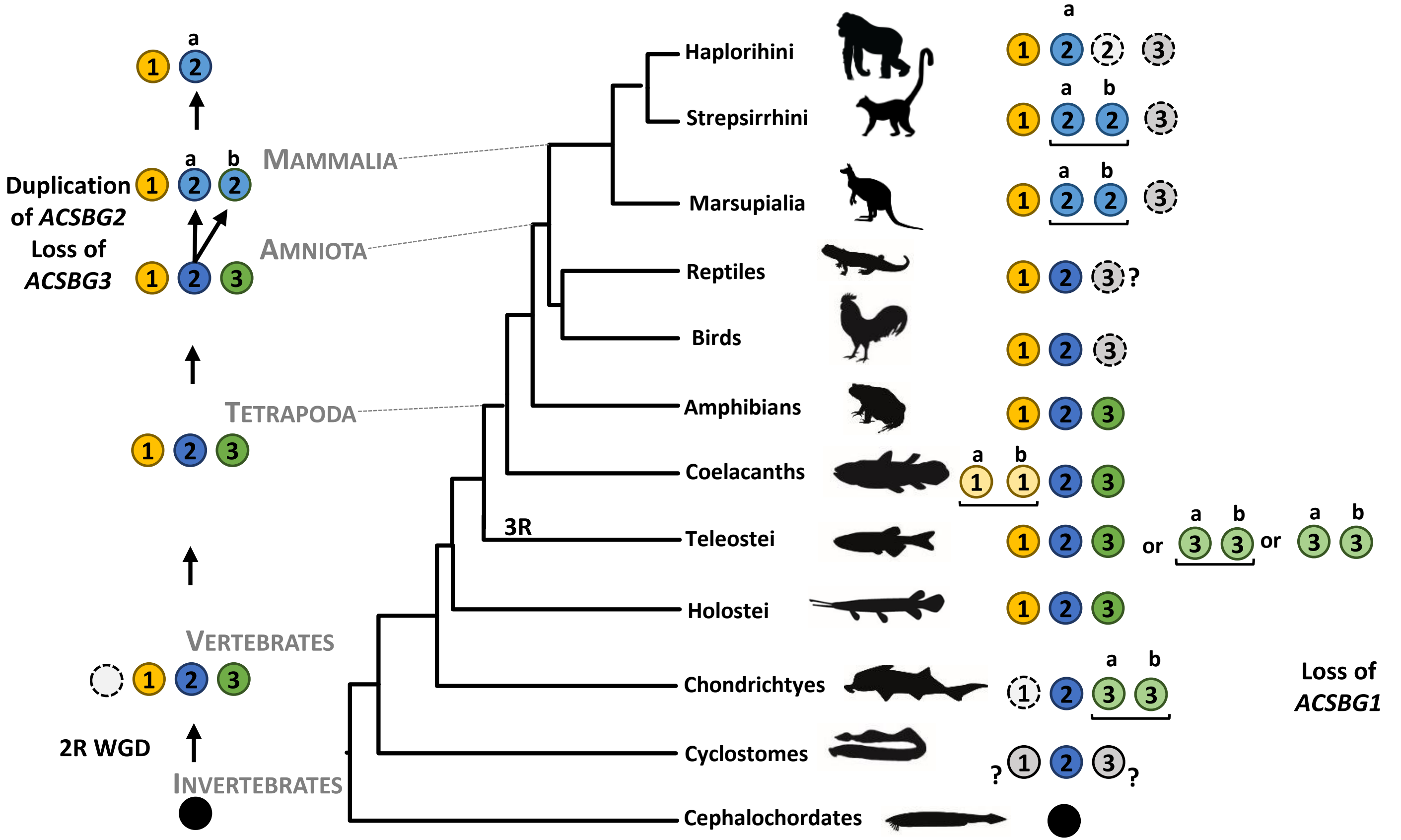
Supplementary material 1: Table containing ACSBG sequences accession numbers.

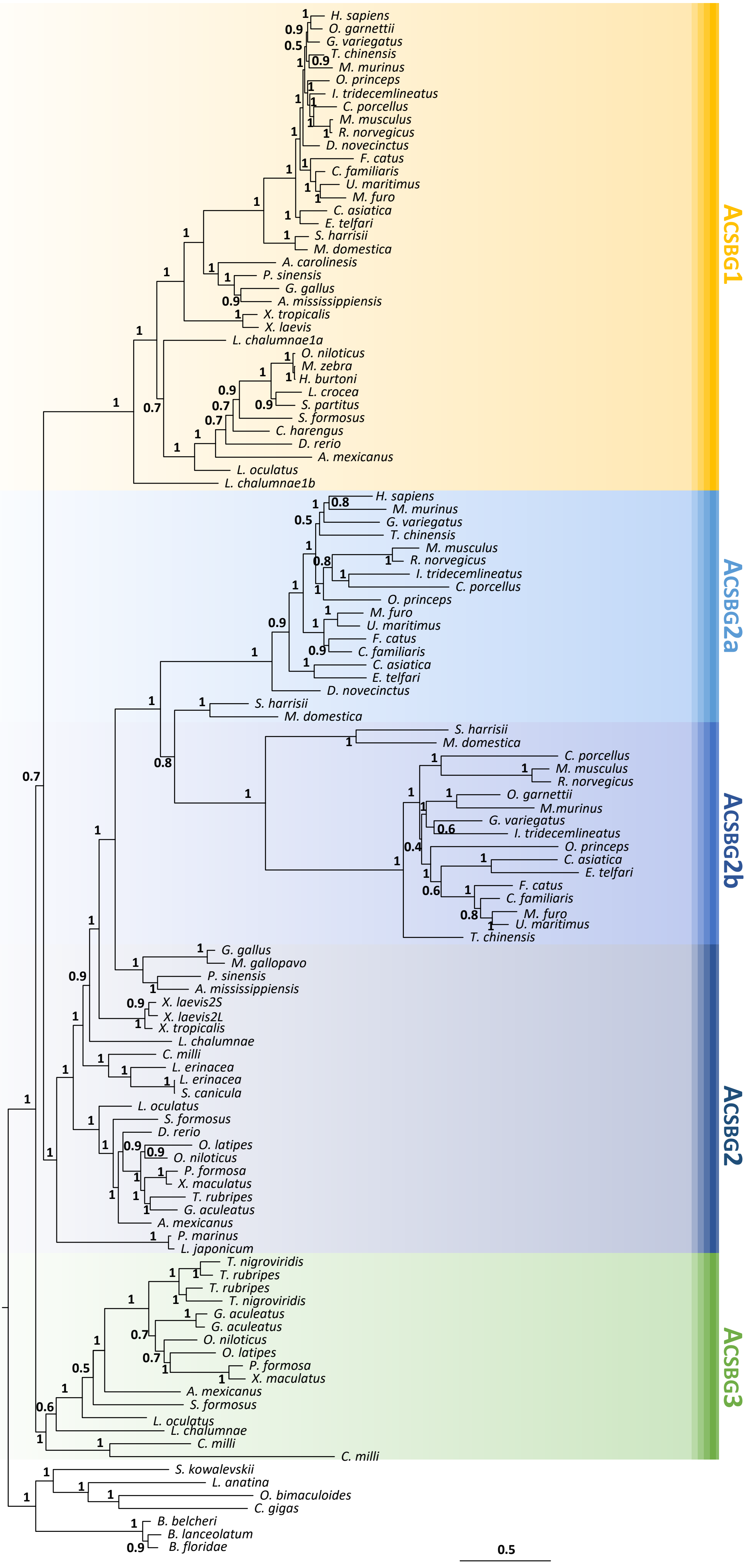
Supplementary material 2: Accession numbers of the RNAseq files retrieved for expression analysis.

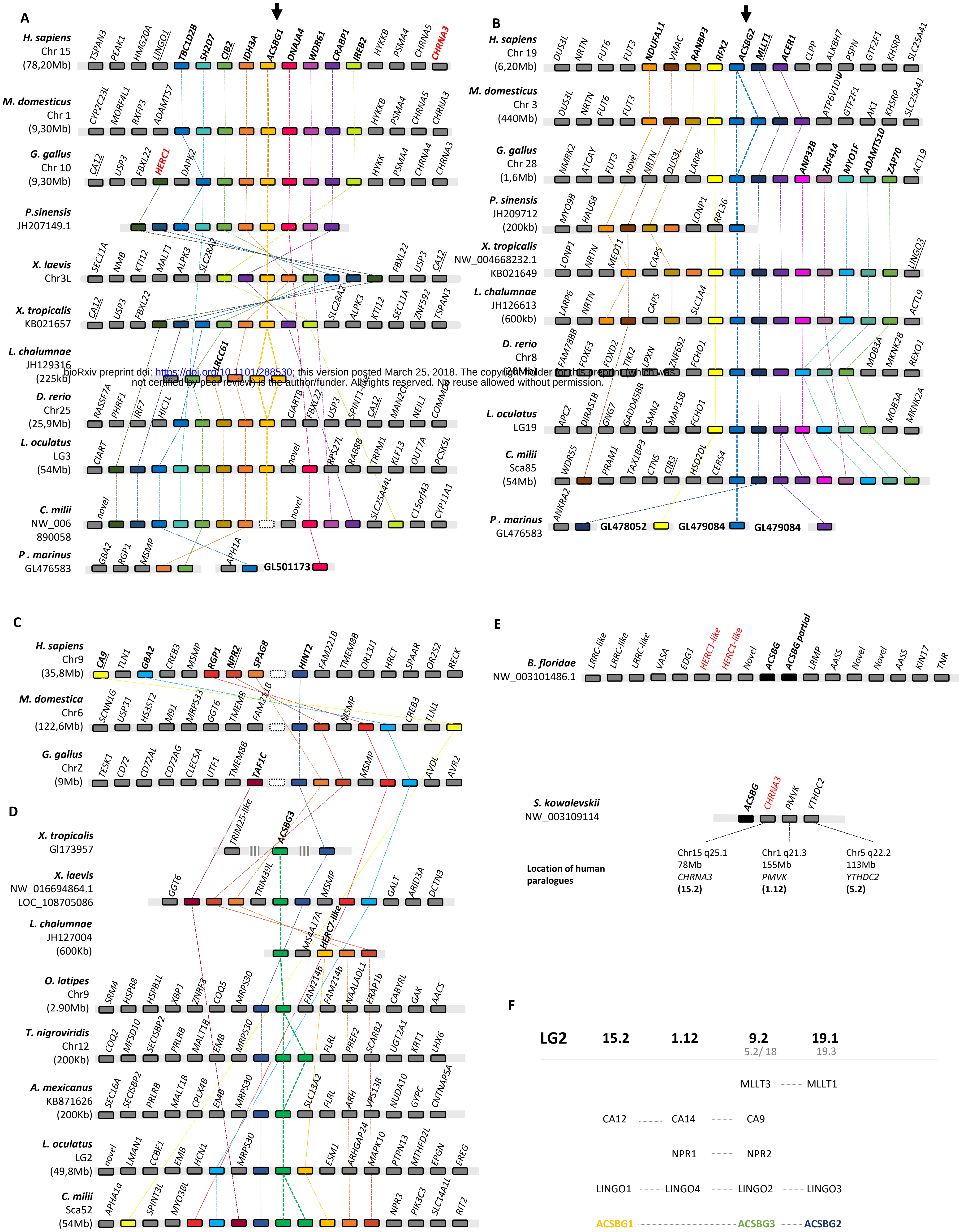
Supplementary material 3: Genome and GTF files retrieved from Ensemble database (Release 89) and Transcriptome files retrieved from NCBI used on this study and accession numbers of reference genes.

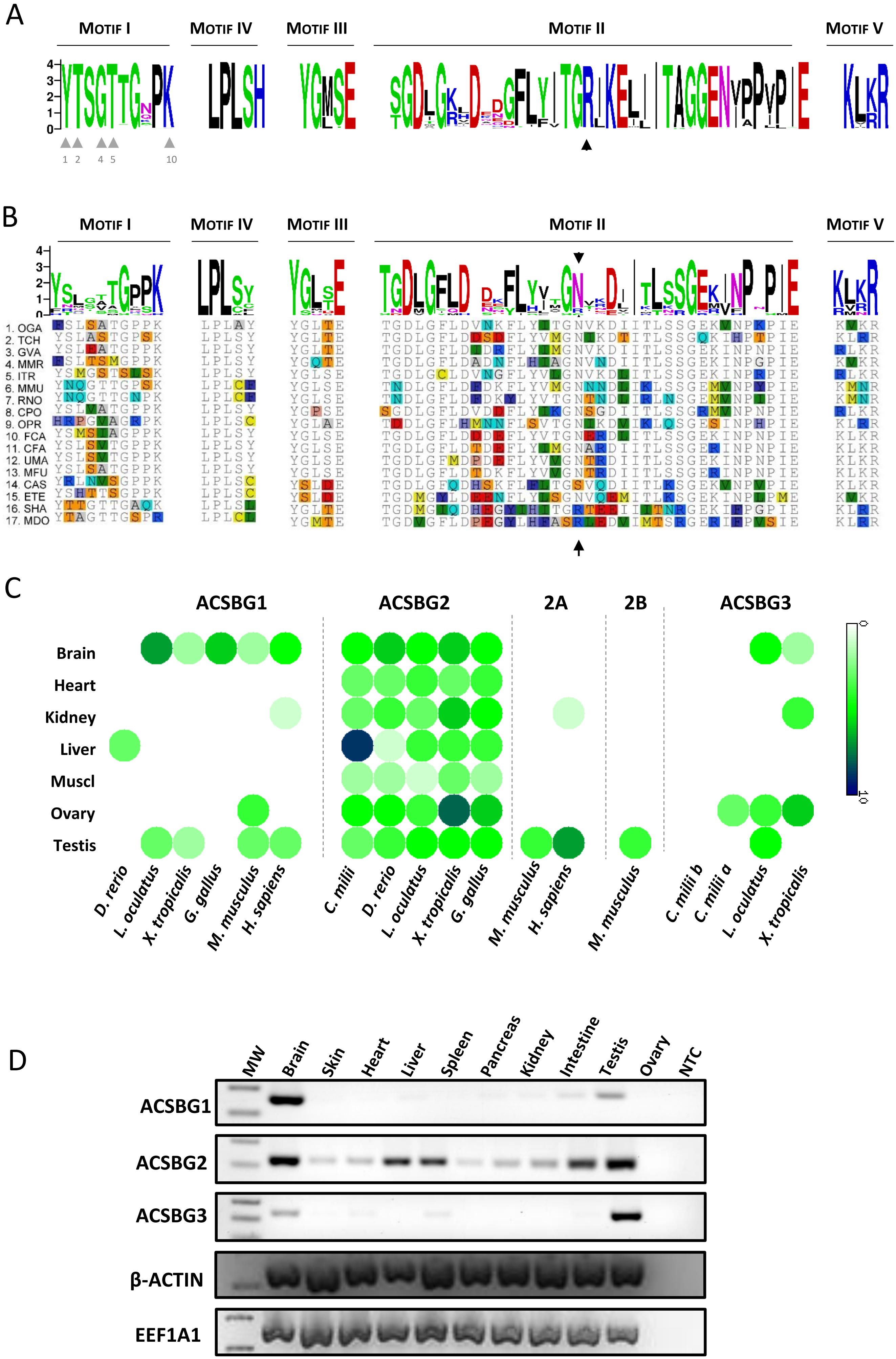
Supplementary material 4: Complementary phylogenetic analysis of ACSBG sequences including amphibian uncharacterized truncated ACSBG-like sequences indicated in red.

Supplementary material 5: Motif sequence alignment of the full dataset used 121 ACSBG sequences. Red box highlights mammal specific ACSBG2b.


a
1 2
Duplication of ACSBG2
Loss of ACSBG3
a b
1 2 2
1 2 3
MAMMALIA
AMNIOTA
TETRAPODA
1 2 3
VERTEBRATES
1 2 3
2R WGD
INVERTEBRATES
3R
Haplorihini
Strepsirrhini
Marsupialia
Reptiles
Birds
Amphibians
Coelacanths
Teleostei
Holostei
Chondrichtyes
Cyclostomes
Cephalochordates
a
1 2 2 3
a b
1 2 2 3
a b
1 2 2 3
1 2 3 ?
1 2 3
1 2 3
a b
1 1 2 3
1 2 3 or
a b
3 3 or
a b
3 3
1 2 3
a b
1 2 3 3
? 1 2 3 ?
Loss of ACSBG1


ACSBG1
1
H. sapiens
0.9
O. garnettii
0.5
G. variegatus
0.9
T. chinensis
1
M. murinus
O. princeps
1
I. tridecemlineatus
1
1
C. porcellus
1
M. musculus
1
R. norvegicus
1
D. novecinctus
1
F. catus
1
C. familiaris
1
U. maritimus
1
M. furo
1
C. asiatica
E. telfari
1
1
S. harrisii
M. domestica
1
A. carolinesis
1
1
P. sinensis
G. gallus
0.9
A. mississippiensis
1
X. tropicalis
X. laevis
1
L. chalumnae1a
1
O. niloticus
M. zebra
1
1
H. burtoni
0.9
L. crocea
0.7
0.9
S. partitus
0.7
S. formosus
1
C. harengus
1
D. rerio
A. mexicanus
L. oculatus
L. chalumnae1b
0.7
1
ACSBG2a
0.8
H. sapiens
1
M. murinus
0.5
G. variegatus
T. chinensis
1
M. musculus
0.8
1
R. norvegicus
1
I. tridecemlineatus
1
C. porcellus
1
O. princeps
1
M. furo
1
U. maritimus
F. catus
0.9
0.9
C. familiaris
1
C. asiatica
E. telfari
1
D. novecinctus
1
1
S. harrisii
M. domestica
0.8
ACSBG2b
S. harrisii
1
M. domestica
1
C. porcellus
1
M. musculus
R. norvegicus
1
1
O. garnettii
1
M.murinus
G. variegatus
0.6
I. tridecemlineatus
0.4
O. princeps
1
C. asiatica
1
E. telfari
1
F. catus
0.6
C. familiaris
0.8
M. furo
1
U. maritimus
T. chinensis
1
1
ACSBG2
1
G. gallus
1
M. gallopavo
P. sinensis
1
A. mississippiensis
1
0.9
X. laevis2S
0.9
X. laevis2L
1
X. tropicalis
L. chalumnae
1
C. milli
1
L. erinacea
1
L. erinacea
1
S. canicula
L. oculatus
1
S. formosus
D. rerio
0.9
0.9
O. latipes
1
O. niloticus
1
P. formosa
1
X. maculatus
1
T. rubripes
1
G. aculeatus
1
A. mexicanus
1
1
P. marinus
L. japonicum
0.7
1
ACSBG3
T. nigroviridis
1
1
T. rubripes
T. rubripes
1
1
T. nigroviridis
1
G. aculeatus
G. aculeatus
0.7
O. niloticus
1
0.7
O. latipes
1
P. formosa
1
X. maculatus
0.5
A. mexicanus
1
S. formosus
0.6
L. oculatus
L. chalumnae
1
1
C. milli
C. milli
1
S. kowalevskii
1
L. anatina
1
1
O. bimaculoides
C. gigas
1
B. belcheri
B. lanceolatum
0.9
B. floridae
0.5


A
H. sapiens
Chr 15
(78,20Mb)
TSPAN3
PEAK1
HMG20A
LINGO1
TBC1D2B
SH2D7
CIB2
IDH3A
ACSBG1
DNAJA4
WDR61
CRABP1
IREB2
HYKK8
PSMA4
CHRNA5
CHRNA3
M. domesticus
Chr 1
(9,30Mb)
CYP2C23L
MORF4L1
RXFP3
ADAMTS7
HYKK8
PSMA4
CHRNA5
CHRNA3
G. gallus
Chr 10
(9,30Mb)
CA12
USP3
FBXL22
HERC1
DAPK2
HYKK
PSMA4
CHRNA4
CHRNA3
P.sinensis
JH207149.1
X. laevis
Chr3L
SEC11A
NMB
KTI12
MALT1
ALPK3
SLC28A2
FBXL22
USP3
CA12
X. tropicalis
KB021657
CA12
USP3
FBXL22
SLC28A2
ALPK3
KTI12
SEC11A
ZNF592
TSPAN3
L. chalumnae
JH129316
(225kb)
LRRC61
D. rerio
Chr25
(25,9Mb)
RASSF7A
PHRF1
IRF7
HIC1L
CIARTB
FBXL22
USP3
SPINT1-like
CA12
MAN2C1
NEIL1
COMMD
L. oculatus
LG3
(54Mb)
CIART
novel
RPS27L
RAB8B
TRPM1
KLF13
OUT7A
PCSK5L
C. milii
NW_006
890058
novel
novel
SLC25A44L
C15orf43
CYP11A1
P . marinus
GL476583
GBA2
RGP1
MSMP
APH1A
GL501173
B
H. sapiens
Chr 19
(6,20Mb)
DUS3L
NRTN
FUT6
FUT3
NDUFA11
VMAC
RANBP3
RFX2
ACSBG2
MLLT1
ACER1
CLPP
ALKBH7
PSPN
GTF2F1
KHSRP
SLC25A41
M. domesticus
Chr 3
(440Mb)
DUS3L
NRTN
FUT6
FUT3
ATP6V1D1V
GTF2F1
AK1
KHSRP
SLC25A41
G. gallus
Chr 28
(1,6Mb)
NMRK2
ATCAY
FUT3
novel
NRTN
DUS3L
LARP6
ANP32B
ZNF414
MYO1F
ADAMTS10
ZAP70
ACTL9
P. sinensis
JH209712
(200kb)
MYO9B
HAUS8
LONP1
RPL36
X. tropicalis
NW_004668232.1
KB021649
LONP1
NRTN
MED11
CAPS
LINGO3
L. chalumnae
JH126613
(600kb)
LARP6
NRTN
CAPS
SLC1A4
ACTL9
D. rerio
Chr8
(20Mb)
FAM78BB
FOXE3
FOXD2
TIKI2
LPXN
ZNF692
FCHO1
MOB3A
MKNK2B
REXO1
L. oculatus
LG19
APC2
DIRAS1B
GNG7
GADD45BB
SMN2
MAP1SB
FCHO1
MOB3A
MKNK2A
C. milii
Sca85
(54Mb)
WDR55
PRAM1
TAX1BP3
CTNS
CIB3
HSD2DL
CERS4
P . marinus
GL476583
ANKRA2
GL478052
GL479084
GL479084
C
H. sapiens
Chr9
(35,8Mb)
CA9
TLN1
GBA2
CREB3
MSMP
RGP1
NPR2
SPAG8
HINT2
FAM221B
TMEM8B
OR13J1
HRCT
SPAAR
OR2S2
RECK
M. domestica
Chr6
(122,6Mb)
SCNN1G
USP31
HS3ST2
M91
MRPS33
GGT6
TMEM8
FAM221B
MSMP
CREB3
TLN1
G. gallus
ChrZ
(9Mb)
TESK1
CD72
CD72AL
CD72AG
CLEC5A
UTF1
TMEM8B
TAF1C
MSMP
AVDL
AVR2
D
X. tropicalis
GI173957
TRIM25-like
ACSBG3
X. laevis
NW_016694864.1
LOC_108705086
GGT6
TRIM39L
MSMP
GALT
ARID3A
DCTN3
L. chalumnae
JH127004
(600Kb)
MS4A17A
HERC7-like
O. latipes
Chr9
(2.90Mb)
SRM4
HSPB8
HSPB1L
XBP1
ZNRF3
COQ5
MRPS30
FAM214b
FAM214B
NAALADL1
ERAP1b
CABYRL
GAK
AACS
T. nigroviridis
Chr12
(200Kb)
COQ2
MFSD10
SECISBP2
PRLRB
MALT1B
EMB
MRPS30
FLRL
PREF2
SCARB2
UGT2A1
KRT1
LHX6
A. mexicanus
KB871626
(200Kb)
SEC16A
SECISBP2
PRLRB
MALT1B
CPLX4B
EMB
MRPS30
SLC13A2
FLRL
ARH
VPS13B
NUDA10
GYPC
CNTNAP5A
L. oculatus
LG2
(49,8Mb)
novel
LMAN1
CCBE1
EMB
HCN1
MRPS30
ESM1
ARHGAP24
MAPK10
PTPN13
MTHFD2L
EPGN
EREG
C. milii
Sca52
(54Mb)
APHA1a
SPINT3L
MYO3BL
NPR3
PIK3C3
SLC14A1L
RIT2
E
B. floridae
NW_003101486.1
LRRC-like
LRRC-like
LRRC-like
VASA
EDG1
HERC1-like
HERC1-like
Novel
ACSBG
ACSBG partial
LRMP
AASS
Novel
Novel
AASS
KIN17
TNR
S. kowalevskii
NW_003109114
ACSBG
CHRNA3
PMVK
YTHDC2
Location of human paralogues
Chr15 q25.1
78Mb
CHRNA3
(15.2)
Chr1 q21.3
155Mb
PMVK
(1.12)
Chr5 q22.2
113Mb
YTHDC2
(5.2)
F
LG2
15.2
1.12
9.2
5.2/ 18
19.1
19.3
MLLT3
MLLT1
CA12
CA14
CA9
NPR1
NPR2
LINGO1
LINGO4
LINGO2
LINGO3
ACSBG1
ACSBG3
ACSBG2


A
Motif I
Motif IV
Motif III
Motif II
Motif V
1 2 4 5 10
B
Motif I
Motif IV
Motif III
Motif II
Motif V
1. OGA
2. TCH
3. GVA
4. MMR
5. ITR
6. MMU
7. RNO
8. CPO
9. OPR
10. FCA
11. CFA
12. UMA
13. MFU
14. CAS
15. ETE
16. SHA
17. MDO
C
ACSBG1
ACSBG2
2A
2B
ACSBG3
Brain
Heart
Kidney
Liver
Muscl
Ovary
Testis
0
10
D. rerio
L. oculatus
X. tropicalis
G. gallus
M. musculus
H. sapiens
C. milii
D. rerio
L. oculatus
X. tropicalis
G. gallus
M. musculus
H. sapiens
M. musculus
C. milii b
C. milii a
L. oculatus
X. tropicalis
D
MW
Brain
Skin
Heart
Liver
Spleen
Pancreas
Kidney
Intestine
Testis
Ovary
NTC
ACSBG1
ACSBG2
ACSBG3
β-ACTIN
EEF1A1