Supporting Information

Than *et al.* 10.1073/pnas.0807606105

SI Text

Nucleotide Isolation, Reverse Transcription, and Sequence Assembly. The *LGALS1* gene coding sequences were obtained from *Colobus guereza*, *Cercopithecus nictitans*, and *Varecia variegata* using genomic DNA as template. *Homo sapiens*, *Cercopithecus cephus*, *Macaca mulatta*, *Papio anubis*, and *Ateles fusciceps* coding sequences were obtained from cDNA templates. These sequences, along with available vertebrate *LGALS1* nucleotide sequences (Ensembl version 47, National Center for Biotechnology Information GenBank and Trace Archive) were used for phylogenetic analyses. Supporting information (SI) Dataset S1 shows the species used and their accession numbers.

Total RNA was isolated from *H. sapiens*, *C. cephus*, *P. anubis*, and *A. fusciceps* placentas, as well as from *H. sapiens* extraembryonic membranes using TRIzol Reagent (Invitrogen), RNeasy MinElute Cleanup kit, and RNase-Free DNase (Qiagen) according to the manufacturers' recommendations. The 28S/18S ratio and the RNA integrity number were assessed using a Bioanalyzer 2100 (Agilent Technologies). Total RNA was then reverse transcribed with a TaqMan Reverse Transcription Reagent kit using random hexamers (Applied Biosystems) for *H. sapiens*, SuperScript III Reverse Transcriptase (Invitrogen) for *P. anubis*, and BD PowerScript Reverse Transcriptase (BD Biosciences) for *A. fusciceps* samples. Genomic DNA was isolated from fresh frozen tissues of *C. guereza* and *V. variegata* using the DNAeasy Kit (Qiagen).

The amplification of *LGALS1* exon sequences from genomic and complementary DNAs were performed by "primer walk" using multiple primer combinations. Primers (Dataset S7) were designed with Primer3 software (1) based on the human *LGALS1* gene (NC_000022) and mRNA sequences (NM_002305) available at GenBank and also determined by this study. PCR was performed using reagents in the TaqPCR Core Kit (Qiagen). *A. fusciceps* cDNA was amplified by using a SMART RACE cDNA Amplification Kit (BD Biosciences). Thermal cycling conditions for PCR were based on the manufacturer's recommendation and were performed on a Mastercycler gradient or a Mastercycler ep gradient S (Eppendorf).

The *LGALS1* coding sequences determined in this study were obtained from genomic DNA ($n = 3$) or cDNA ($n = 5$). Sequence reads were imported into Sequencher 4.6 (Gene Codes), separate files were created for each species, and trimmed sequences were assembled into contigs with assembly parameters of at least 75% minimum match and a 10 base overlap between sequence fragments. The eight newly generated sequences were deposited into GenBank (Dataset S1).

Nucleotide Sequence Analyses. *LGALS1* coding sequences were obtained for 34 species as described earlier. From the multiple related coding sequences in *G. gallus* and *X. laevis*, those with the highest sequence similarity to *H. sapiens LGALS1* were selected for inclusion in the study. Multiple sequence alignments of promoter regions, coding nucleotide, and deduced amino acid sequences were performed using MacClade 4.08 (Sinauer Associates) (2), MEGA 4.0 (3), and/or MUSCLE (4). Clustal alignments implemented in MEGA were curated visually. The evolutionary conservation of the 2-kb promoter regions and first exons in 12 species was examined by pair-wise *p*-distance comparison with MEGA 4.0 and by "phylogenetic shadowing" with eSHADOW (5), setting the sliding window size for 5 bp and the hidden Markov model island parameters as follows: eS, 0.85; eF, 0.77; T, 0.1. The most conserved 500 bp comprising the first exons and the 5′ upstream sequences were compared using Foot-Printer 3.0 searching 8- to 12-bp sequences conserved with a maximum of two mutations per sequence and one mutation per branch. Only a subset of species was used for footprinter analysis because of the low quality of the draft genome sequences. Transcriptional Element Search System and Footer 2.0 were used to predict putative transcription factor binding sites. To test whether the amount of *cis* element evolution was greater on the placental stem compared with the therian stem lineage, we compared the proportion of *cis* element gains and losses per million years of evolutionary time (6) on these two lineages.

Phylogenetic Analysis. The aligned sequences were analyzed using maximum parsimony algorithms implemented in PAUP* version 4b.10 (7) and MacClade version 4.08 (2) and the maximum likelihood codon model implemented in PAML version 3.15 (8). Parsimony inferred ancestral sequences and character state transformations were inferred using the ACCTRAN and DELTRAN algorithms. Maximum likelihood (i.e., empirical Bayes) marginal reconstruction of ancestral sequences was implemented using the codon model in codeml. The gene tree was inferred by a heuristic search of 500 bootstrap replicates. The starting trees in each replicate were generated using random addition of sequences, and the branch swapping was performed using the tree bisection-reconnection algorithm. The strength of selection on the *LGALS1* gene was measured using the nonsynonymous/synonymous rate ratio ($dN/dS = \omega$). Using a classic branch test (9) of adaptive evolution as implemented in PAML version 3.15 (8), one, two, and three ω models were compared using likelihood ratio tests to determine whether different ratios existed for placental mammal crown group, the stem placental lineage, and non-placental species. The number of nonsynonymous and synonymous substitutions were inferred for the stem placental lineage in codeml that allowed us to determine whether the ω ratios were different among these lineages (9). We also performed branch-site tests to determine whether individual codons had experienced selection during the emergence of the placenta. A potential criticism of this type of test is that it returns too many false-positive results (10); however, the improved branch-site test we used circumvents this issue (11). PAML analyses were conducted on a most parsimonious gene tree and on the presumed species tree. To reconcile the gene tree to the species tree (12) in order to infer gene duplications and losses, the Notung 2.5 software (13) was used.

***In Silico LGALS1* Gene Expression Analysis.** Search of the SAGE Genie database (14) yielded 298 libraries with levels of *LGALS1* expression. SAGE data were then exported from these libraries and relative *LGALS1* expression levels (tags per 200,000) were evaluated.

Amino Acid Sequence Analysis. Amino acid substitutions in galectin-1 on the stem of placental mammals were color-coded on the x-ray crystallographic model of human galectin-1 accessed from the MMDB Database (National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health) by Yasara 7.9.3.

Quantitative Real-Time RT-PCR. The *LGALS1* TaqMan Gene Expression Assay (Hs00169327_m1) and *RPLPO* TaqMan Endogenous Control (4326314E; Applied Biosystems) were used for relative quantification of *LGALS1* gene expression in *H. sapiens*

placentas ($n = 6$) and extraembryonic membranes ($n = 6$) containing maternal decidua. Both genes were run in triplicate to allow the assessment of technical variability.

Immunohistochemistry. Immunostaining was performed on 5-μm-thick paraffin sections of *H. sapiens*, *C. nictitans*, *A. fusciceps*, and *V. variegata* placentas with goat anti-human galectin-1 IgG (R&D Systems) using a Discovery autostainer (Ventana Medical Systems). Sections were incubated for 1 h at 1:50 dilution following antigen retrieval in a citrate-based buffer (pH 6.0). A biotinylated horse anti-goat IgG (Vector Laboratories) at 1:400 dilution, HRP-conjugated streptavidin and 3,3′-diaminobenzidine chromogen (Ventana Medical Systems) were used for the detection.

1. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291.
2. Maddison DR, Maddison WP (2000) MacClade 4: Analysis of Phylogeny and Character Evolution (Sinauer & Associates. Sunderland, MA).
3. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
4. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
5. Ovcharenko I, Boffelli D, Loots GG (2004) eShadow: a tool for comparing closely related sequences. *Genome Res* 14:1191–1198.
6. Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.
7. Swofford DL (2002) Phylogenetic analysis using parsimony, *(and other methods). v. 4.0b10 (Sinauer & Associates. Sunderland, MA).
8. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
9. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573.
10. Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* 99:364–373.
11. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479.
12. Goodman M, *et al.* (1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms. *Syst Zool* 28:132–163.
13. Durand D, Halldorsson BV, Vernot B (2006) A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol* 13:320–335.
14. Boon K, *et al.* (2002) An anatomy of normal and malignant gene expression. *Proc Natl Acad Sci USA* 99:11287–11292.

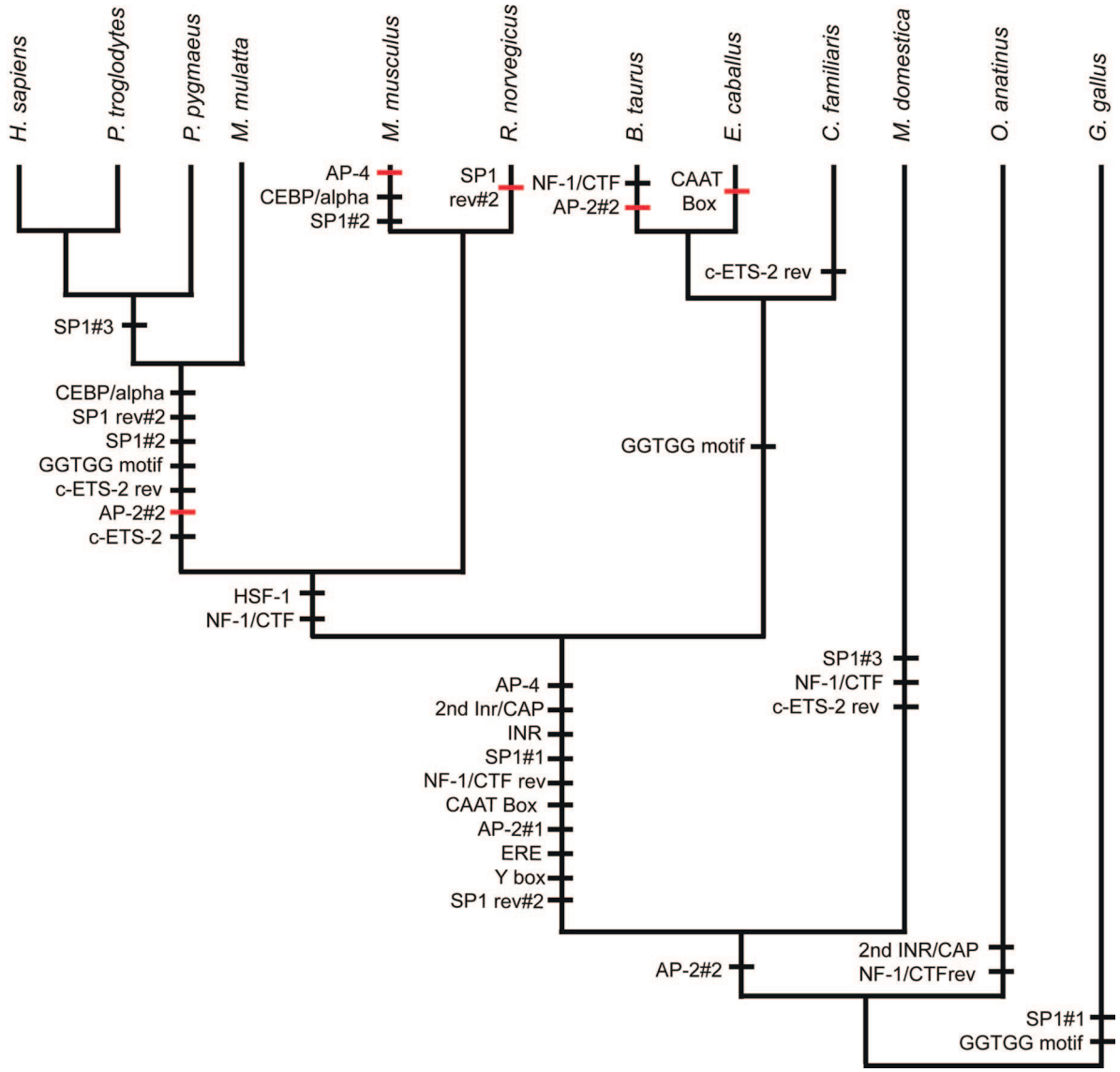

Fig. S1. Evolution of *cis* elements in the *LGALS1* 5′ promoter. A phylogenetic tree depicting the evolutionary relationships among species with publicly available *LGALS1* 5′ promoter sequences. Parsimony inferred gains (black tick marks) and losses (red tick marks) were reconstructed using the DELTRAN algorithm. In total, there were 32 gains and five losses of *cis* elements among the studied taxa. The stem of placental mammals underwent a gain of 10 *cis*-elements including a half estrogen response element. There was also gain of six and loss of one *cis* element on the primate stem. The proportion of *cis* element gains and losses per million years of evolutionary time (1) was significantly higher on the placental stem compared with the therian stem lineage (Fisher's exact test, $P = 0.009$).

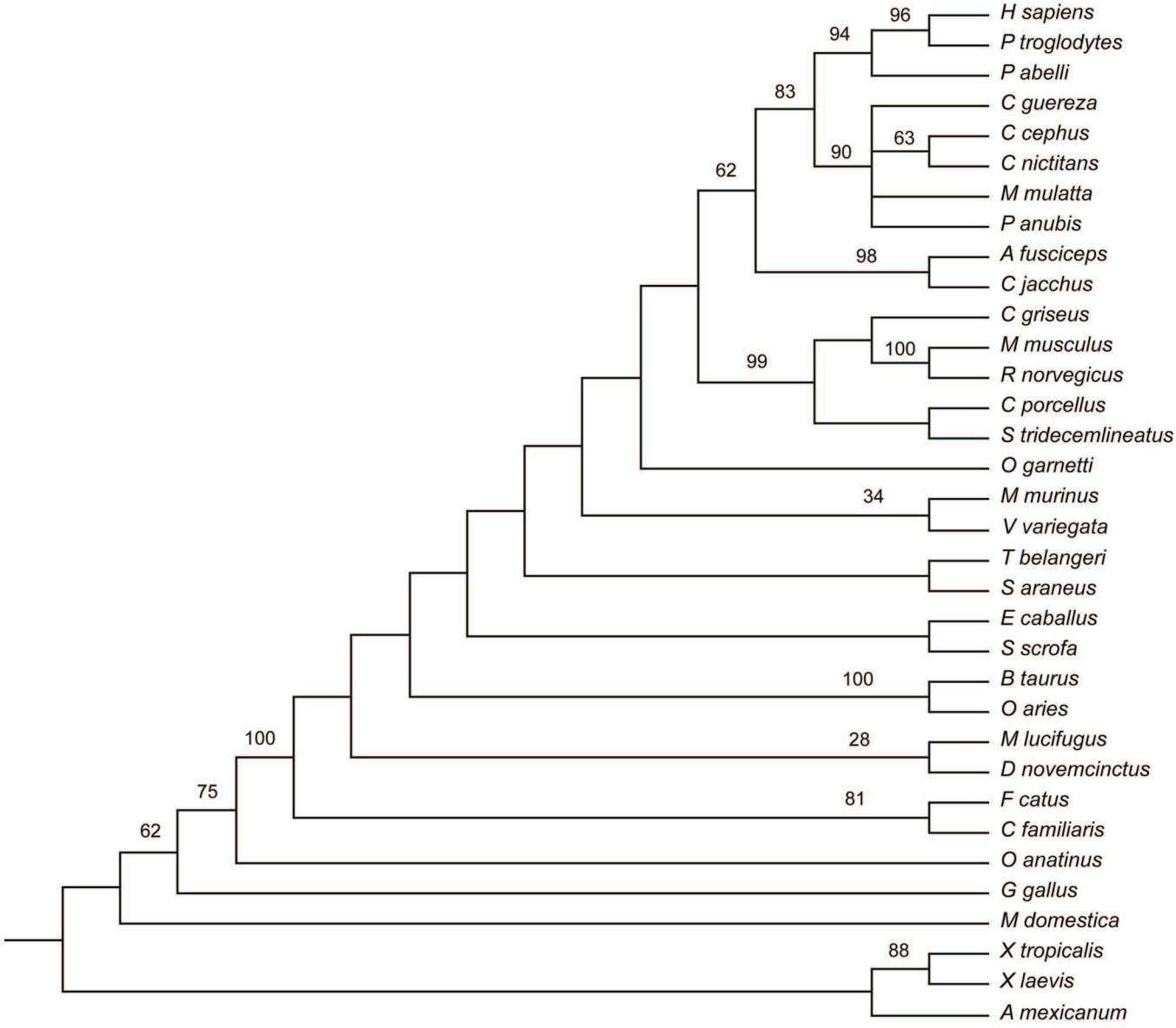

Fig. S2. Most parsimonious gene tree for galectin-1. The tree shown is one of eight most parsimonious trees inferred from nucleotide sequences with a score of 1,061. Bootstrap percent values (out of 500 random addition sequence replicates) of branch support greater than 25 are indicated above branches. As can be seen from the topology presented, neither primates nor artiodactyls were reconstructed as monophyletic orders. Additionally, there are some highly unlikely groupings depicted in the tree (e.g., bat grouping with armadillo). Additionally, monophyletic Eutherian and mammalian clades were not recovered. Explanations for the differences between the gene and species tree include independent gene gain and loss (2), rate heterogeneity and/or lack of sufficient sequence variation in particular lineages.

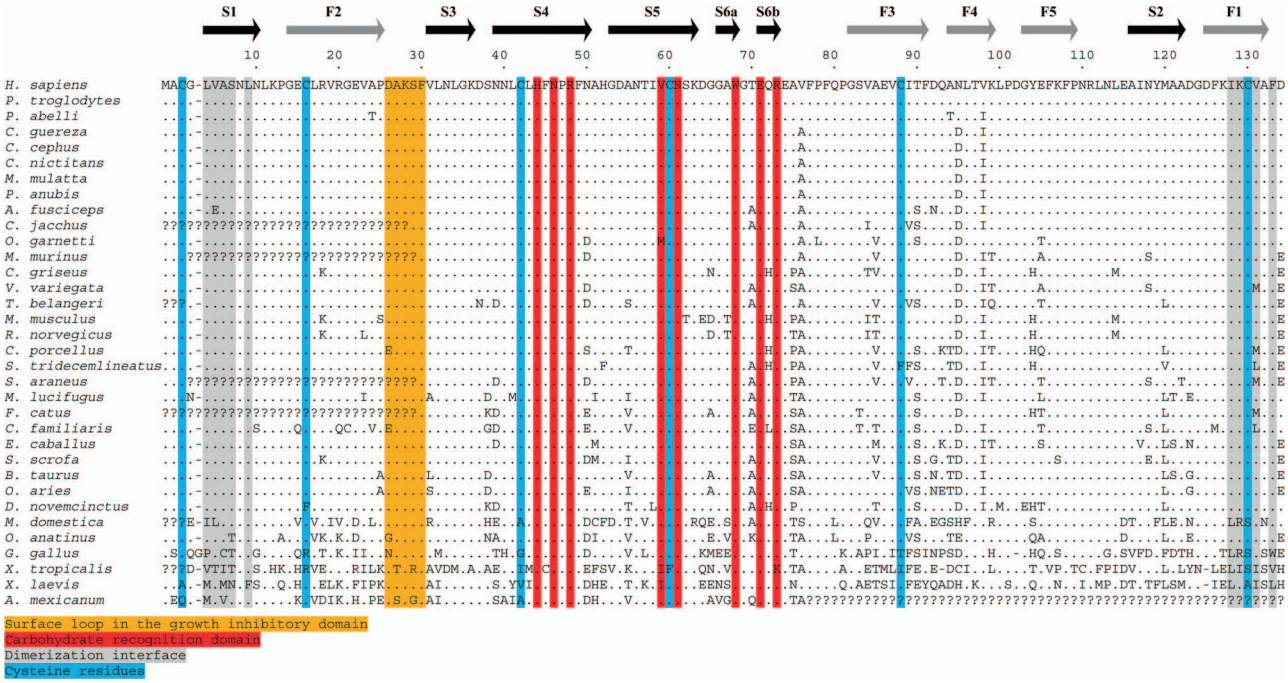

Fig. S3. Multiple sequence alignment of galectin-1 aa sequences. β-strands and residue positions in human galectin-1 are indicated at the top. Residues identical to those in the human sequence are represented with dots, substitutions with the amino acid code of the differing residues, missing data with question marks, and gaps in the WGS sequences with dashes. Residues involved in carbohydrate binding are conserved in all investigated taxa, with only conservative replacements for Val-59 and Arg-73. Two residues in the growth-inhibitory domain are conserved in 100%, two in 93%, and one in 87% in the studied taxa. Seven of nine residues implicated in dimerization of galectin-1 are conserved in 100% in placental mammals with 96% and 78% conservation of two additional residues, respectively. All substitutions in these positions are conservative in lower vertebrates except in two taxa. Cysteines shared by placental mammals are important for the regulation of immune functions adaptive to the redox status.

Other Supporting Information Files

Dataset S1
Dataset S2
Dataset S3
Dataset S4
Dataset S5
Dataset S6
Dataset S7