Haptoglobin Is a Divergent MASP Family Member That Neofunctionalized To Recycle Hemoglobin via CD163 in Mammals

Anthony K. Redmond,*,† Yuko Ohta,‡ Michael F. Criscitiello,§ Daniel J. Macqueen,* Martin F. Flajnik,‡ and Helen Dooley*,‡,¶

In mammals, haptoglobin (Hp) is an acute-phase plasma protein that binds with high affinity to hemoglobin (Hb) released by intravascular hemolysis. The resultant Hp–Hb complexes are bound and cleared by the scavenger receptor CD163, limiting Hb-induced oxidative damage. In this study, we show that Hp is a divergent member of the complement-initiating MASP family of proteins, which emerged in the ancestor of jawed vertebrates. We demonstrate that Hp has been independently lost from multiple vertebrate lineages, that characterized Hb-interacting residues of mammals are poorly conserved in nonmammalian species maintaining Hp, and that the extended loop 3 region of Hp, which mediates CD163 binding, is present only in mammals. We show that the Hb-binding ability of cartilaginous fish (nurse shark, *Ginglymostoma cirratum*; small-spotted catshark, *Scyliorhinus canicula*; and thornback ray, *Raja clavata*) and teleost fish (rainbow trout, *Oncorhynchus mykiss*) Hp is species specific, and where binding does occur it is likely mediated through a different structural mechanism to mammalian Hp. The continued, high-level expression of Hp in cartilaginous fishes in which Hb binding is not evident signals that Hp has (an)other, yet unstudied, role(s) in these species. Previous work indicates that mammalian Hp also has secondary, immunomodulatory functions that are independent of Hb binding; our work suggests these may be remnants of evolutionary more ancient functions, retained after Hb removal became the primary role of Hp in mammals. ***The Journal of Immunology*****, 2018, 201: 2483–2491.**

Haptoglobin (Hp) is an acute-phase plasma protein first discovered in 1938 by Polonovski and Jayle (1). Although there is evidence that Hp has immunomodulatory effects (2–4), its principal recognized function in mammals is the sequestration of free hemoglobin (Hb) released into the bloodstream by damaged erythrocytes (5). Hp binds almost irreversibly to Hb, with the resultant complex being cleared by the scavenger receptor CD163 present on monocytes and macrophages (6). In this way, Hp, aided by the structurally unrelated, heme-binding protein, hemopexin (Hx) (7), limits Hb-induced oxidative damage (Supplemental Fig. 1A).

In most species examined, Hp is produced as a proprotein containing one or two complement control protein (CCP) domains and a C-terminal serine protease (SP) domain. Mammalian pro-Hp is cleaved between the CCP and SP domain in the endoplasmic reticulum by C1r-like protein (C1r-LP) to generate the mature, disulphide-linked protein (8). The enzymatically inactive SP domain (or β-chain) mediates both Hb and CD163 binding (9, 10), whereas the CCP domains (comprising the α-chain) dictate the oligomerization state of Hp in the blood. For example, the human Hp gene exists in two major allelic forms, designated Hp1 and Hp2; the Hp1 allele has a lone CCP and forms disulphide-bonded dimers, whereas the Hp2 allele, which seemingly arose through the nonhomologous recombination of two Hp1 genes after the divergence of *Homo* from other primates (11), has two CCP domains and forms higher-order multimers (10, 12). In contrast, teleost fish Hp (Hp-like) has a short (20 aa) peptide containing a cleavage site for a subtilisin-like proprotein convertase in place of CCP domains and is monomeric in blood (9) (Supplemental Fig. 1B).

Although many evolutionary relationships have been proposed for Hp, the weight of evidence suggests it arose via partial duplication of a MASP family member (9, 13). The members of this family, MASPs 1–3, C1r, and C1s, are initiators of the complement system and play key roles in immune protection. MASP family members are also produced as proproteins, composed of two CUB, one EGF, two CCP, and an SP domain; cleavage between the second CCP and SP domain followed by covalent association of the two chains via a disulphide bond leads to the mature form of each protein. MASP family members are classified into two types based upon their SP domains: MASP-1 has an active site serine encoded by a TCN codon, whereas in all other MASP family members, this serine is encoded by AGY (13–15).

*School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom; †Centre for Genome-Enabled Biology and Medicine, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom; ‡Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201; §Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843; and ¶Institute of Marine and Environmental Technology, Baltimore, MD 21202

ORCIDs: 0000-0003-4262-7832 (M.F.C.); 0000-0002-2570-574X (H.D.).

Received for publication April 9, 2018. Accepted for publication August 14, 2018.

This work was supported by National Institutes of Health Grant RR06603 awarded to M.F.F., Royal Society Research Grant RG130789 awarded to H.D., and a Ph.D. studentship awarded by the Centre for Genome-Enabled Biology and Medicine, University of Aberdeen, to A.K.R.

The sequences presented in this article have been submitted to GenBank under accession numbers HM566086, JN564036, and MG747494.

Address correspondence and reprint requests to Dr. Helen Dooley, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, 701 E. Pratt Street, Baltimore, MD 21202. E-mail address: hdooley@som.umaryland.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: CCP, complement control protein; C1r-LP, C1r-like protein; Hb, hemoglobin; Hp, haptoglobin; Hx, hemopexin; IMAC, immobilized-metal affinity chromatography; LC-MS/MS, liquid chromatography–tandem mass spectrometry; LSB, Laemmli sample buffer; MSA, multiple sequence alignment; PBST, PBS containing 0.05% Tween 20; SEC, size-exclusion chromatography; SP, serine protease; TGS, tris-glycine SDS.

Beyond mammals and teleost fishes, information on Hp is sparse. Although chicken and goose have lost Hp, with the structurally unrelated PIT54 apparently substituting its Hb-binding function, other birds retain both molecules (9). Further, despite reports of an unknown Hb-binding protein in amphibian and reptile serum, Hp is absent from the genomes of *Xenopus* and anole lizard (9, 14, 16). Although the cartilaginous fishes (sharks, skates, rays, and chimaera) are the earliest branching vertebrate taxon to share true orthologs of the tetrameric Hb of mammals (17), it is proposed that Hp arose in the common ancestor of teleost fishes and tetrapods (9, 14). However, during analysis of blood plasma from nurse sharks (*Ginglymostoma cirratum*), we identified one of the major plasma proteins as the shark ortholog of Hp. Subsequently, we re-examined the evolutionary origins and distribution of Hp across vertebrate phylogeny, alongside its sequence-level evolution and associated capacity for Hb binding. Our analyses pinpoint how, and when, Hp became a major player in Hb clearance via CD163 and suggest that in some species, Hp has roles that are independent of Hb binding.

Materials and Methods

Animal maintenance and sampling

Wild-caught nurse sharks (*G. cirratum*) were maintained in artificial seawater at ~28˚C in indoor tanks at the Institute of Marine and Environmental Technology (Baltimore, MD). Captive-bred small-spotted catsharks (*Scyliorhinus canicula*) were maintained in artificial seawater at ~12˚C in indoor tanks at the University of Aberdeen (Aberdeen, U.K.). Little skate (*Leucoraja erinacea*) and thornback ray (*Raja clavata*) samples were obtained from wild-caught animals, and rainbow trout (*Oncorhynchus mykiss*) samples were obtained from purchased, farm-raised fish. Animals were anesthetized with MS-222 (0.16 g/l seawater) before bleeds were harvested from the caudal vein into a 1:10 blood volume of 1000 U/ml porcine heparin reconstituted in shark-modified PBS (mammalian PBS supplemented with 15 ml 5 M NaCl and 100 ml 3.5 M urea/l) or, for trout, mammalian PBS, then spun at 300 g for 10 min to isolate blood plasma and packed RBCs. All procedures were conducted in accordance with University of Maryland School of Medicine Institutional Animal Care and Use Committee protocols and the U.K. Home Office Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

Nurse shark plasma fractionation

Nurse shark plasma was fractionated by passage over a HiPrep 16/60 Sephacryl S-300 high resolution size exclusion column (GE Healthcare Life Sciences) equilibrated in 0.5 M NaCl, 0.05 M $NaPO_4$ (pH 7.2) running buffer, as previously described (18); samples from each fraction were run nonreducing on 5% SDS-PAGE gels in tris-glycine SDS (TGS) running buffer. For two-dimensional analysis, lane 7 from the nonreducing gel was cut out and placed in the well of a 12% SDS-PAGE gel. Excess 2× reducing Laemmli sample buffer (LSB) was added to the well and incubated for 5 min, then the gel was run as above. Bands were visualized following Coomassie Brilliant Blue staining.

N-terminal and internal protein sequencing of the unknown plasma proteins

Protein samples were run on gels prerun for 2 h at 3 mA in TGS with 430 μl of 10 mM reduced glutathione stock added per 100 ml buffer in the upper reservoir. The buffer was removed and replaced with fresh TGS containing 100 μl of 100 mM sodium thioglycolate stock per 100 ml prior to sample loading. The resultant gels were pre-equilibrated, then blotted onto methanol-wetted and equilibrated Immobilon-PSQ PVDF membrane (MilliporeSigma) in CAPS transfer buffer for 2 h at 250 mA in a wet blotting system. Membranes were stained with amido black and air-dried before sending to the Microchemical and Proteomics core facility at Emory University for N-terminal sequencing by Edman degradation or internal sequencing following tryptic digestion. Returned peptide sequences are presented in Table I.

Cloning of cartilaginous fish Hp

N-terminal amino acid sequences for nurse shark P1 and P2 were reverse translated and used to design two degenerate 3′ RACE primers, GcP1-degenF and GcP2-degenF. These primers were used on RACE-primed cDNA synthesized from 2.5 μg of nurse shark total RNA using the SMARTer RACE cDNA Amplification Kit (Clontech Laboratories), as per the manufacturer's instructions. A band of ~300 bp was obtained with the primer GcP2-degenF with liver RNA, allowing design of the gene-specific primer GcP2-F2 to perform nested 3′ RACE PCR. Using this approach, we amplified a transcript of ~1100 bp, which encoded peptide sequences from both P1 and P2. The 5′ end of the transcript was amplified using a nested 5′ RACE approach with the primers GcP2-R1 and GcP2-R2. A single band was obtained using the primers GcP1/2-F1 and GcP1/2-R1, and the full-length sequence was confirmed. Oligo(dT)-primed cDNA was prepared from liver RNA of little skate and small-spotted catshark using illustra Ready-To-Go RT-PCR beads (GE Healthcare Life Sciences) according to the manufacturer's instructions. To confirm the full-length little skate sequence, the primers LeHp-F1 and LeHp-R1 were used, whereas for catshark, the primers ScHp-F1 and ScHp-R1 were used. All primer sequences are detailed in Table II. Full-length Hp sequences can be accessed at GenBank using the following accession numbers: nurse shark, HM566086; little skate, JN564036; and small-spotted catshark, MG747494.

Hb precipitation of plasma proteins

Hb–Sepharose was prepared by lysing packed RBCs in 10 vol of PBS at a 1:10 dilution, then spun at 13 K rpm for 10 min to remove cell debris. The RBC lysates were run on reducing SDS-PAGE against purchased "purified" human Hb and imaged on an Azure Biosystems c500 imaging system, then images were subject to densitometric analysis using ImageJ to assess their Hb content; Hb constituted >95% of the total protein in the human and catshark samples, >80% for nurse shark, and >70% for rainbow trout. Supernatants were diluted in 10 vol of binding buffer and conjugated to activated cyanogen bromide Sepharose (Sigma-Aldrich) according to the manufacturer's instructions. The Hb–Sepharose was washed extensively to remove loosely bound Hb, then equilibrated overnight in mammalian PBS (human and trout) or shark-modified PBS (all cartilaginous fish samples). Plasma samples were diluted 1:10 in PBS, and 1 ml was added to 200 μl species-matched Hb–Sepharose, then incubated for 1 h at room temperature with rotation. The Hb–Sepharose was washed one, five, or ten times with PBS containing 0.05% Tween 20 (PBST). Following the final wash, all supernatant was carefully aspirated, 50 μl of 2× reducing LSB was added to the Sepharose, and the resultant slurry was boiled for 5 min to remove all bound proteins. For immobilized-metal affinity chromatography (IMAC) precipitation, 1 ml of plasma diluted 1:10 in PBS was added to 100 μl Ni^{2+}–Sepharose (Qiagen). The Sepharose was washed four times with 1 ml PBST, twice with 1 ml 20 mM imidazole and four times with PBST before the addition of reducing LSB and boiling as above. Boiled slurries were zip-spun, and supernatants were run on 4–12% Novex Bis-Tris gels (Invitrogen) in MOPS buffer. Bands were visualized by Coomassie Brilliant Blue staining. Bands of interest were excised from the gel with a clean scalpel, then sent to the University of Aberdeen Proteomics Facility for protein identification via liquid chromatography–tandem mass spectrometry (LC-MS/MS) following in-gel tryptic digestion.

Comparative genomic searches

tblastn/blastp searches of genomic/transcriptomic/protein sequence data available for species of interest (Supplemental Table I) were performed using the well-characterized human, pig, and pufferfish Hp protein sequences at a relaxed E-value cutoff of 10. Resulting hits were then blast searched against the National Center for Biotechnology Information nonredundant protein database, and those which had top hits to Hp or other MASP family members were retained for further analysis. These were examined for typical characteristics of Hp (no active site serine, no CUB or EGF domains), and added to phylogenetic analysis of the MASP family and Hp. Loss of Hp was inferred where we could identify non-Hp MASP family members in a species but not Hp itself, despite Hp being the query for the search. Taxa where Hp was not found in our initial screen were subject to additional searches using the most closely related species with a newly-identified Hp.

Phylogenetic analyses

All multiple sequence alignments (MSAs) were generated using PRANK with the "+F" parameter specified (19). All available libraries in the Transitive Consistency Score web server (20) were used to assess alignments for positional reliability, with columns scoring below two discarded to alleviate phylogenetic noise as a result of potential misalignment. IQ-TREE was used to identify the best-fitting model of amino acid substitution for each alignment (21).

Bayesian phylogenetic analyses were performed in BEAST v1.8 (22), specifying a Yule speciation prior, an uncorrelated lognormal relaxed

molecular clock model, the best-fitting amino acid substitution (LG+I+G, JTT+I+G, and WAG+I+G for SP, C1r-LP, and CCP datasets, respectively), and a random start tree. Two Markov chain Monte Carlo runs were executed for each analysis with chain lengths of 5, 10, and 20 million states each for the Hp CCP and C1r-LP, MASP CCP, and SP datasets, respectively. In all cases, chains were sampled every 1000 states before being assessed for sufficient sampling (effective sample sizes for all parameters >200) as well as convergence and mixing in Tracer v1.6 (http://beast.bio.ed.ac.uk/Tracer). Tree files were combined specifying a 20% burn-in using LogCombiner v1.8, and maximum clade credibility trees were then generated in TreeAnnotator v1.8. For the SP domain, dataset Bayesian analyses were repeated as specified above but using less well-fitting models of amino acid substitution (LG+G, WAG+I+G, JTT+I+G) to determine if model perturbation would affect the placement of Hp within the MASP family.

Results

Identification of Hp in shark plasma

During size-exclusion chromatography (SEC) of nurse shark plasma (18) we observed that one of the major plasma proteins eluted from the column at a molecular mass of ~250 kDa but on nonreducing SDS-PAGE had a molecular mass of ~60 kDa (boxed in Fig. 1A). Under reducing conditions, this protein further resolved into two bands of ~35 and ~25 kDa (Fig. 1B; P1 and P2), indicating that the unknown protein was a disulphide-bonded heterodimer that forms noncovalently associated oligomers in plasma. Immunoprecipitation of iodinated plasma proteins with Ig-specific mAbs showed the proteins were not associating with Igs present in the same fractions. P1 and P2 were submitted for N-terminal and internal protein sequencing, then the peptide sequences returned (Table I) were subjected to blast analysis. A single hit, a 20 aa peptide sequence from a similarly sized, heterodimeric, sandbar shark (*Carcharhinus plumbeus*) plasma protein (23), was returned using the P2 N-terminal sequence as the query.

Next, degenerate 5′ RACE primers were designed from the N-terminal sequences, and a partial sequence was obtained from liver RNA with the P2 primer; gene-specific RACE primers were then designed to complete the sequence (Table II). A product of ~1100 bp was identified that encoded both P1 and P2 within a single transcript of 429 aa. blast searches using this sequence identified Hp and MASP family members as the top hits; however, analysis of conserved domains showed the unknown protein, with no CUB or EGF domains, was structurally most like Hp. tblastn searches using this nurse shark MASP/Hp homolog recovered similar transcripts for little skate (*L. erinacea*) in Skatebase (24) and

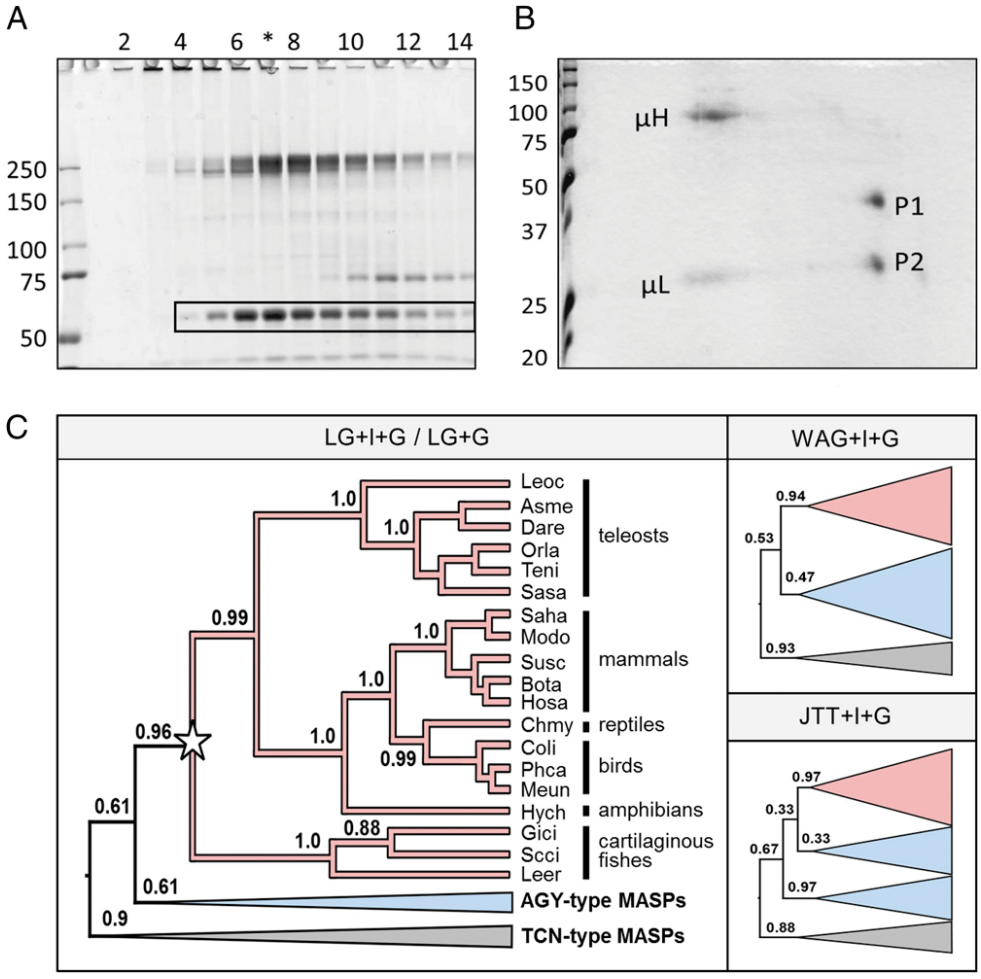


FIGURE 1. Identification of unknown shark plasma proteins as Hp. (**A**) Nonreducing SDS-PAGE of SEC-fractionated shark plasma showing the unknown protein at ~60 kDa (boxed). (**B**) When lane 7 (*) was cut from the gel and run in the second dimension under reducing conditions, this band resolved into two spots of ~35 kDa (P1) and 25 kDa (P2). The H (μH) and L (μL) chains of IgM were also resolved. (**C**) Bayesian relaxed-clock rooting analyses of Hp (red clades) and the MASP family (blue clades, AGY-type; gray clades, TCN-type) strongly indicate that we have cloned cartilaginous fish Hp and, further, that Hp is a MASP family member irrespective of the model of amino acid substitution used. (LG+I+G was the best-fitting model, and this was also tested without the invariant sites parameter [i.e., LG+G]. The less-well fitting, but commonly used, WAG and JTT models were also tested, along with the +I+G parameters.) Four-letter abbreviations for genus and species are used as detailed in Supplemental Table I Full tree topologies, posterior probabilities, species names, and accession numbers for (B) and (C) are presented in Supplemental Fig. 2A.

Table I. N-terminal and internal protein sequencing results for the unknown nurse shark plasma proteins

Peptide	Sequence
P1 N-terminal	VVGGHLVHNGATPxTVLMLGPsgtv
P2 N-terminal	DHVETDHSKVHCGVPV**xIT**hghY
Internal 14	VVCGRPIVPLEQhrq
Internal 16	DAYVYR
Internal 17	VVCGRPv/iWLEQH
Internal 23	LWEDVHFSNHIMPAclpah
Internal 28	VYVGIEDAR
Internal 34	NTDLGYEFPTxexv
Internal 36	wIDGIIHpk

Lowercase letters denote the probable amino acid at that position; x denotes that the amino acid could not be determined. v/i denotes that a mixture of these amino acids was found at that position. Bold denotes a probable N-linked glycosylation site.

Table II. Sequences of primers used in this work

Primer	Sequence
GcP1-degenF	5′-GTNCAYAAYGGiGCiACiCC-3′
GcP2-degenF	5′-AARGTNCAYTGYGGiGTiCCiG-3′
GcP2-F2	5′-GCACAGACGATAATCAGTGG-3′
GcP2-R1	5′-GGCCATGGTCCACGGCAG-3′
GcP2-R2	5′-CCACTGATTATCGTCTGTGC-3′
GcP1/2-F1	5′-CCCTCTCCCTCCAGCTGGTAC-3′
GcP1/2-R1	5′-CACTGCGGGTGAATGATGCCGTCG-3′
ScHp-F1	5′-ATGCTTCTCACAAAGATGTTCACTGTGG-3′
ScHp-R1	5′-ACTCAATGTGCGGCCATGGTTTCC-3′
LeHp-F1	5′-ATGTGGTTCCTCGTGTTAAACG-3′
LeHp-R1	5′-TCAGTTATGTTCTATGACGTTGTTGATCC-3′

Standard nucleotide ambiguity codes are used; i indicates the presence of inosine.

small-spotted catshark (*S. canicula*) in our in-house transcriptome (A.K. Redmond, D.J. Macqueen, and H. Dooley, submitted for publication); both transcripts were amplified by RT-PCR from liver RNA from their respective species and confirmed by Sanger sequencing.

Phylogenetic analysis of Hp and the MASP family revealed the presence of a monophyletic Hp clade in jawed vertebrates that included our cartilaginous fish sequences, with branching patterns matching closely to expected species phylogeny (Fig. 1C). Critically, unlike previous studies (9, 14), we probabilistically determined the best-fitting amino acid substitution model and incorporated a relaxed molecular clock model to infer the root of the tree (25), as an appropriate outgroup is not known. The root fell between a MASP-1 clade (TCN) and an Hp + AGY-type MASP clade (Fig. 1C; Supplemental Fig. 2A). This strongly indicates Hp is a MASP-family member that arose in the common ancestor of jawed vertebrates. Analyses using less well-fitting amino acid substitution models corroborated this finding, but the relationship of the AGY-type MASPs with Hp varied by model (Fig. 1C).

Characterization of cartilaginous fish Hp

All identified cartilaginous fish Hp orthologs have two CCP domains and an SP domain with an incomplete catalytic triad (H-D-R/H instead of H-D-S; Fig. 2A). Although CCP1 and CCP2 of classical MASPs grouped into separate clades (Fig. 2B, Supplemental Fig. 2B), both CCPs from human and cow Hp fell within the CCP2 clade (Fig. 2C, Supplemental Fig. 2C), fitting with their known evolutionary histories (26). Further, both cartilaginous fish CCPs fell within the CCP2 clade, indicating an independent CCP domain duplication event after their split with bony vertebrates, but before the emergence of elasmobranchs. Thus, Hp of the jawed vertebrate ancestor likely had a single CCP domain (CCP2) that has independently duplicated in several vertebrate lineages.

With two CCP domains, cartilaginous fish Hp has the same structure as human Hp2 (12, 27) and, according to its SEC profile, likewise forms multimers in plasma. However, as this interaction is disrupted under nonreducing conditions, oligomerization must occur via noncovalent forces rather than disulphide bonding, as in human Hp. Sequence comparison indicates that the unpaired Cys found at the N terminus of each CCP domain in human Hp (highlighted in yellow in Fig. 2A) are paired in cartilaginous fish Hps (and classical MASPs; dotted lines on Fig. 2A), making it unavailable for intermolecular bonding. In human Hp, electrostatic docking occurs between the inter-CCP linker and tip of the N-terminal CCP domain of the two molecules before disulfide bond formation (key residues marked with +/- in Fig. 2A) (27); although these residues differs between the species, from our SEC data it appears that salt bonds maintain the quaternary structure of shark Hp, despite the harsh (high-salt, high-urea) conditions in vivo.

Both nurse shark and catshark Hp have a proprotein convertase site akin to mammalian Hp (and not a subtilisin-like site as in teleost fish Hp), followed by an Ile-(Val/Leu/Ile)-Gly-Gly motif which is thought to drive a conformational change in the SP domains of trypsin-like SPs through interaction with a conserved Asp (marked with a star below the alignment in Fig. 2A) at the base of the active-site cavity (27) postcleavage. Neither the cleavage site nor I-(V/L/I)-G-G motif were found in little skate Hp, suggesting the proprotein is not processed in the same manner (or at all) in this species. However, the Cys pair required to link the cleaved chains in the mature protein (marked with dots in Fig. 2A) is present in all species.

Recurrent losses of Hp during vertebrate evolution

To better understand the evolution of Hp, we screened an array of sequence datasets for orthologs (see Supplemental Table I for full species list) and identified none within jawless fishes or nonvertebrate chordates. Although we could not identify Hp in the elephant shark (a chimera) genome, its presence in sharks and skates indicates this to be a secondary loss. Further, the presence of Hp across vertebrate phylogeny was sporadic; for example, although it is conserved in ray-finned fish, including holosteans and teleosts, we did not identify Hp in coelacanth. Additionally, although Hp was missing from multiple anuran (frog) transcriptomes, it was identified in a salamander transcriptome. In reptiles, we identified Hp in turtle and crocodile, but not lizard or snake (squamate), datasets. Our data also indicate that the previously reported loss of Hp in birds is likely limited to the Galloanserae (land and water fowl). Although Hp is the primary route for Hb sequestration and recycling in humans, the high frequency of loss events in other vertebrate lineages led us to question the role of Hp outside mammals.

Hp–Hb binding is species dependent

To test the Hb-binding ability of cartilaginous fish Hp, we obtained blood from three species: nurse shark, small-spotted catshark, and thornback ray (*R. clavata*). Red cell pellets were lysed in hypotonic buffer, and the released Hb was covalently coupled to activated cyanogen bromide Sepharose, which was then used as a matrix for precipitation of proteins from species-matched blood plasma. Following 1, 5, or 10 washes, bound proteins were eluted by boiling the Sepharose in reducing LSB prior to SDS-PAGE; low intensity bands, representing leached Hb, were observed in all no-plasma controls (Fig. 3, −ve lane). As it was shown previously that shark Hp could be enriched from plasma by IMAC (23, 28), we also

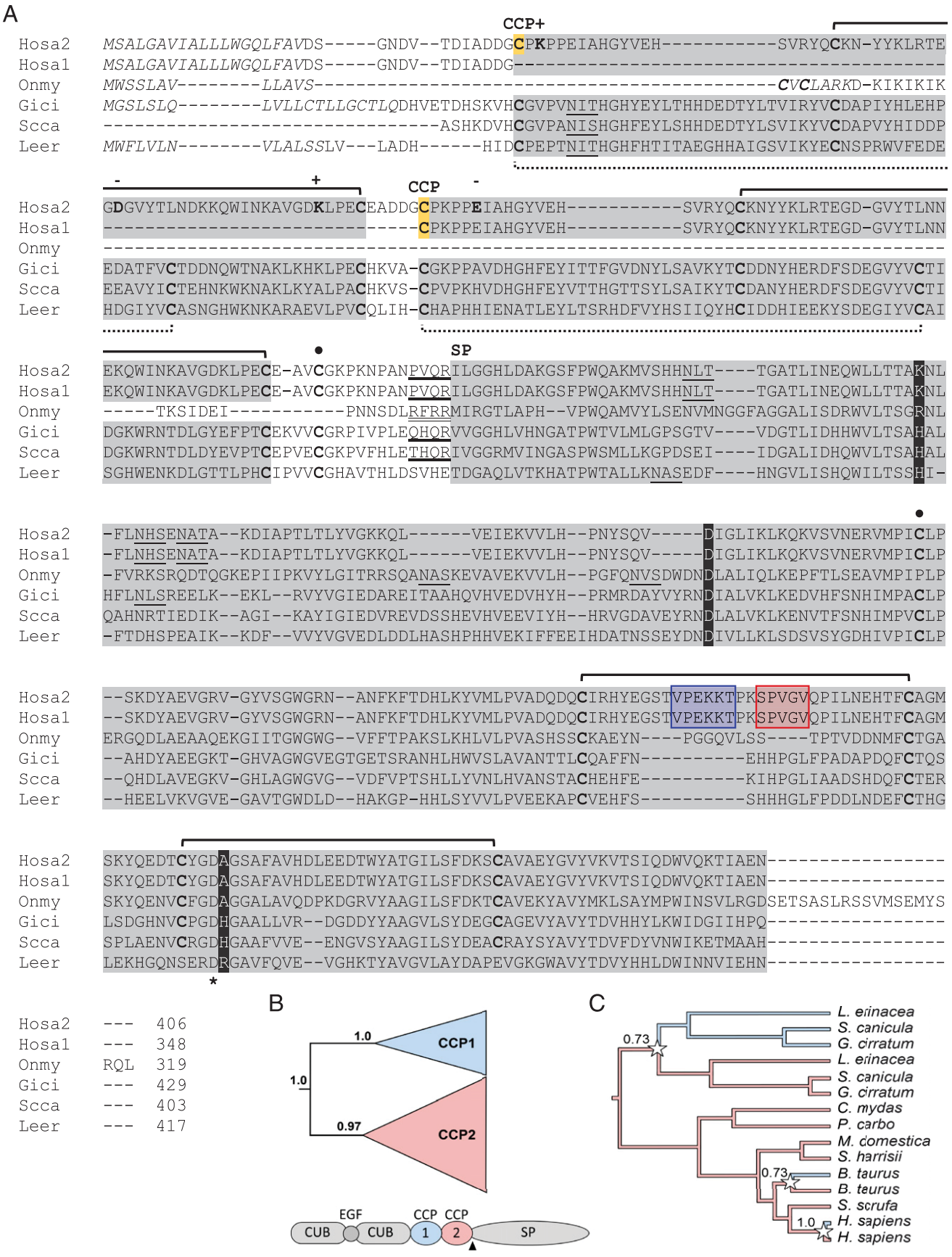


FIGURE 2. Sequence features of cartilaginous fish Hp. (**A**) MSA of nurse shark (Gici), little skate (Leer), and small-spotted catshark (Scca) Hp with the two human (Hosa) Hp alleles (GenBank accession numbers; Hp1 NP_005134.1; Hp2 NP_001119574.1) and trout (Onmy) Hp-like (sequence from nine); CCP and SP domains are shaded and the pro-Hp cleavage motif are underlined. The cysteines responsible for interchain bonding are marked by dots above the sequence, and the unpaired cysteines that facilitate Hp oligomerization in mammals are highlighted in yellow. Cysteine pairs that form disulphide bonds in mammalian Hp are indicated by solid lines above the alignment, whereas additional disulphide bonds predicted in cartilaginous fish Hp are indicated by dashed lines below the alignment. The charged residues important for the electrostatic docking of human Hp oligomers are *(Figure legend continues)*

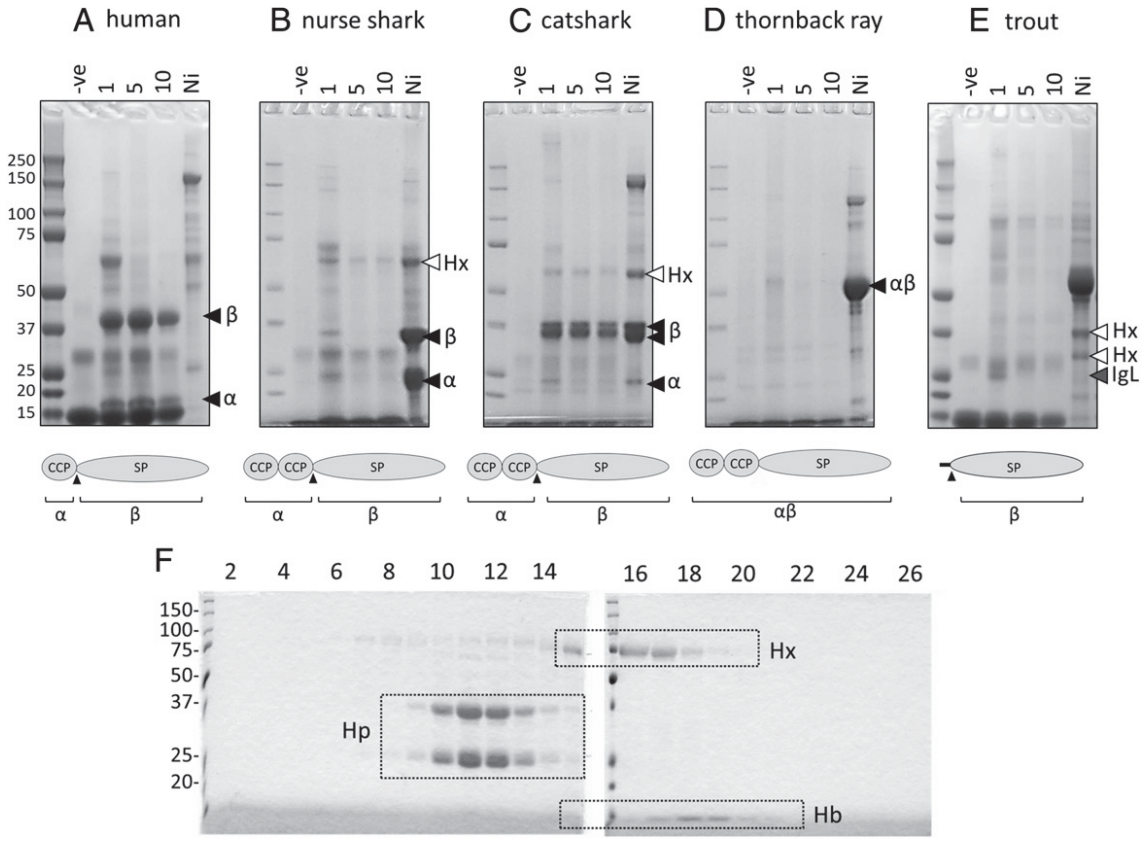


FIGURE 3. SDS-PAGE of Hb-precipitated proteins obtained from (**A**) human, (**B**) nurse shark, (**C**) thornback ray, (**D**) small-spotted catshark, and (**E**) rainbow trout plasma. Proteins were eluted from species matched Hb–Sepharose incubated without plasma (−ve) or with plasma following 1, 5, or 10 washes with species-matched PBST and run under reducing conditions on 4–12% gradient gels. The structure of Hp in each species is illustrated under each SDS-PAGE image, and their expected products upon reduction is indicated underneath. Plasma incubated with Ni^{2+}–Sepharose (Ni) proves that Hp (black arrowheads indicate the SP domain–containing β-chain and CCP-containing α-chain) is present in all cartilaginous fish plasma samples; however, only human and catshark Hp are significantly enriched on species-matched Hb. White arrowheads indicate full-length or partially degraded forms of the heme-binding protein Hx, and the gray arrowhead identifies dissociated trout Ig L chain (IgL). All band identities were confirmed by LC-MS/MS. (**F**) SEC confirms that nurse shark Hp does not bind species-matched Hb. Hp and Hx were IMAC purified from nurse shark plasma and mixed with a limited amount of species-matched Hb. The mixture was passed over an S300 SEC column, and the resulting fractions were subject to reducing SDS-PAGE to assess their protein content. Fraction numbers are indicated above the lanes, and boxes indicate the presence of Hb (which also gave the fraction a red color), Hp, or Hx in each fraction.

incubated plasma from each species with Ni^{2+}–Sepharose and washed it 10 times before elution as above. Human Hp did not bind to Ni^{2+}–Sepharose, but two strong bands representing the dissociated Hp α- and β-chains were observed following incubation with species-matched Hb–Sepharose, regardless of pre-elution wash stringency (Fig. 3A). Given that the ability of an interaction to withstand dissociation by elution correlates well with affinity measured by other means [e.g., equilibrium dialysis or BIAcore (29, 30)], this result was consistent with the previously reported high affinity interaction of human Hp with dimeric Hb (10, 31).

Confirming previous observations (28), three bands were precipitated from nurse shark plasma with Ni^{2+}–Sepharose, with those at 25 and 35 kDa being confirmed by LC-MS/MS as Hp α- and β-chains, respectively (black arrowheads, Fig. 3B). Neither of these bands was enriched on Hb–Sepharose except under the lowest stringency wash conditions in which faint bands were observed. The band at ∼70 kDa (white arrowhead Fig. 3B) is Hx (28), indicating that our pull-down was successful but that nurse shark Hp binds Hb with low affinity. Only a single band of ∼50 kDa was observed for Hp following Ni^{2+}–Sepharose pull-down of thornback ray plasma, suggesting that like little skate Hp, this species also lacks a convertase cleavage motif and so is present in plasma in an unprocessed form. Regardless, there was no significant enrichment of this band on species-matched Hb–Sepharose, even under the lowest stringency wash conditions (Fig. 3D). To ensure that the lack of Hp binding was not a consequence of the Hb–Sepharose coupling procedure, we mixed IMAC-enriched nurse shark plasma proteins with a limited amount of species-matched Hb, and the resultant mix was passed over an S300 SEC column. Fractions were then subject to

indicated by +/- above the alignment. The residues that form the active site in other SP domains are indicated by white lettering on black, and the conserved Asp located at the base of the active-site cavity is marked with an asterisk. Residues identified as important in Hp–Hb complex formation (32) in mammals are shaded red, whereas those important for CD163 binding (34) are shaded blue. (**B**) Phylogenetic analysis of MASP CCP domains; regardless of the MASP family member or species from which they are derived, the CCP1 and CCP2 domains (shown in blue and red, respectively, on the structural schematic for a typical MASP family member) segregate into distinct clades. (**C**) However, both CCP domains in human, cow, shark, and skate Hp fall within the CCP2 group indicating independent domain duplication events in these lineages (marked by stars). Full tree topologies, posterior probabilities, species names, and accession numbers for (A) and (B) are presented in Supplemental Fig. 2B, 2C, respectively.

SDS-PAGE to assess their protein content. Confirming our pull-down results, nurse shark Hb was found in the same fractions as Hx, not Hp (Fig. 3F). Contrasting with nurse shark and thornback ray, Hp from small-spotted catshark did bind Hb–Sepharose (Fig. 3C). Several bands were enriched on both Ni^{2+}–Sepharose and Hb–Sepharose; LC-MS/MS revealed that the band at ~60 kDa (white arrowhead) is a degradation product of catshark Hx, whereas those below (black arrowheads) are the α- and β-chains of catshark Hp.

Information on Hb binding by Hp in other nonmammalian lineages is sparse; however, Wicher and Fries (9) previously reported successful precipitation of pufferfish (*Takifugu rubripes*) Hp on immobilized Hb from common carp (*Cyprinus carpio*). Conversely, we found that even under our least stringent wash conditions, rainbow trout (*O. mykiss*) Hp was not enriched on species-matched Hb (Fig. 3E; lacking CCP domains, the expected molecular mass of trout pro-Hp is ~35 kDa, and mature Hp is ~30 kDa). However, the only bands approximating this size were determined by LC-MS/MS to be degradation products of trout Hx (also known as WAP65; white arrowheads) and Ig L chains (gray arrowhead).

Known Hb-interacting residues are poorly conserved outside mammals

Structural analysis of the mammalian Hp–Hb complex identified a number of residues important for Hb binding (10, 32). Specifically, residues within loop D, at the end of loop 3 and within loop 1 [using the nomenclature of Perona and Craik (33)], form extensive interactions with Hb, contributing significantly to the high affinity binding of the two molecules. MSA of Hp SP domains from across phylogeny (Supplemental Fig. 3) indicates these Hb-interacting residues are poorly preserved between mammals and other vertebrates. This implies that distinct molecular interactions are used for complex formation in nonmammalian species in which Hp does bind Hb. Interestingly, we observed that cartilaginous fish Hp is enriched for histidine (mean ± SD: 9.1 ± 1.7% His across the SP domain, n3 sequences) when compared with mammalian (2.9 ± 0.5%, n6), reptile (4.4 ± 1.0%, n3), bird (1.7 ± 0.3%, n6), and teleost (2.1 ± 0.9%, n7) orthologs, as well as other cartilaginous fish MASP family members (2.8 ± 0.8%, n14; see Supplemental Table I for further details). His enrichment likely accounts for the ability of cartilaginous fish Hp to bind Ni^{2+}–Sepharose (a property not possessed by mammalian or teleost Hp; Fig. 3A, 3E) and suggests Hb binding by catshark Hp may occur via His coordination of Hb's Fe^{2+} ion.

The CD163 binding region of Hp arose in mammals

Loop 3 of human Hp is extended when compared with all other SPs (34, 35), forming a protrusion that binds the scavenger receptor CD163, allowing the removal of circulating Hp-Hb complexes (34). The MSA of vertebrate Hp SP domains, generated using PRANK to accurately predict indels (19), revealed that loop 3 is very short or missing in all vertebrate lineages besides mammals (Fig. 4), being more like that of MASPs and other SPs. This suggests that the entire loop 3 extension was acquired as a single insertion during early mammalian evolution (Fig. 5) and was present in the common therian ancestor. Importantly, as mutation of individual residues within loop 3 of human Hp abolishes CD163 binding (34), it is unlikely that Hb-complexed Hp proteins lacking loop 3 are removed from the blood by the same mechanism.

Discussion

Mammalian Hp binds to Hb with extremely high affinity, allowing the resultant complex to be removed via CD163, thus limiting oxidative damage following intravascular hemolysis. In this study, we provide multiple lines of evidence that Hp acquired this role late in its evolutionary history.

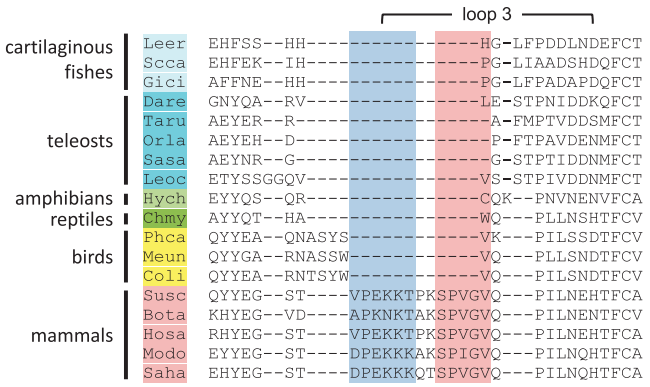


FIGURE 4. Comparison of the Hp loop 3 region across phylogeny. MSA of the loop 3 region across vertebrate phylogeny. Four-letter abbreviations for genus and species are as detailed in Supplemental Table I. Residues identified as important in Hp–Hb complex formation (32) in mammals are shaded red, whereas those important for CD163 binding (34) are shaded blue.

Our analyses indicate that the proto-Hp (CCP-SP structure) arose sometime before the emergence of extant jawed vertebrate lineages, via partial duplication of an AGY-type MASP family member (Fig. 5). Mutation of the active site serine would have rendered the proto-Hp unable to cleave other MASP-family members, whereas the absence of CUB and EGF domains would likely have excluded its incorporation into C1 or MASP complexes (36). However, the maintenance of the proprotein convertase site (and lack of C1r-LP demonstrated in this study) suggests the new molecule was still cleaved by complement system initiators, most likely C1r. Extant cartilaginous fishes have extremely high levels (>20 mg/ml) of IgM in their plasma (37); perhaps the equally high levels of Hp expressed by members of this lineage (regardless of whether Hb binding is evident or not) help prevent complement overactivation through the Ab-mediated, C1r-initiated, classical pathway.

Following their split from a common ancestor with bony vertebrates, the Hp of cartilaginous fish underwent CCP domain duplication (to regain a CCP–CCP–SP structure) and became markedly enriched for the amino acid histidine. The selective force behind this His enrichment is unknown but, intriguingly, in the case of catshark Hp may confer the ability to bind Hb. We predict that in this instance binding occurs through coordination of Hb's Fe^{2+} ion, rather than the noncovalent interactions that dictate mammalian Hp–Hb binding. A similar cross-reactivity has been observed for other His-rich proteins; for example, mammalian HRG (His-rich glycoprotein) binds a wide-range of targets that include divalent metal ions and heme (38), with binding to both being negated by chemical modification of the His residues (39). It is noteworthy that, in our hands, trout Hp did not show Hb (or Ni^{2+}) binding. This contradicts the observation of Wicher and Fries (9) with pufferfish Hp and suggests that Hb binding is also species specific in bony fishes. How Hb binding is mediated in teleost species, lacking many of the Hb-interacting residues of mammalian Hp or the His enrichment of cartilaginous fish Hp, requires further investigation. Further, the mechanism used to safely detoxify any Hp–Hb complex formed in vivo in nonmammals remains unclear, given that it likely cannot be removed from the circulation by CD163. Some of these questions will hopefully be answered as crystal structures of Hp from different nonmammalian vertebrates become available.

It is apparent, however, that Hp underwent significant changes during early mammalian evolution, with further gene duplications

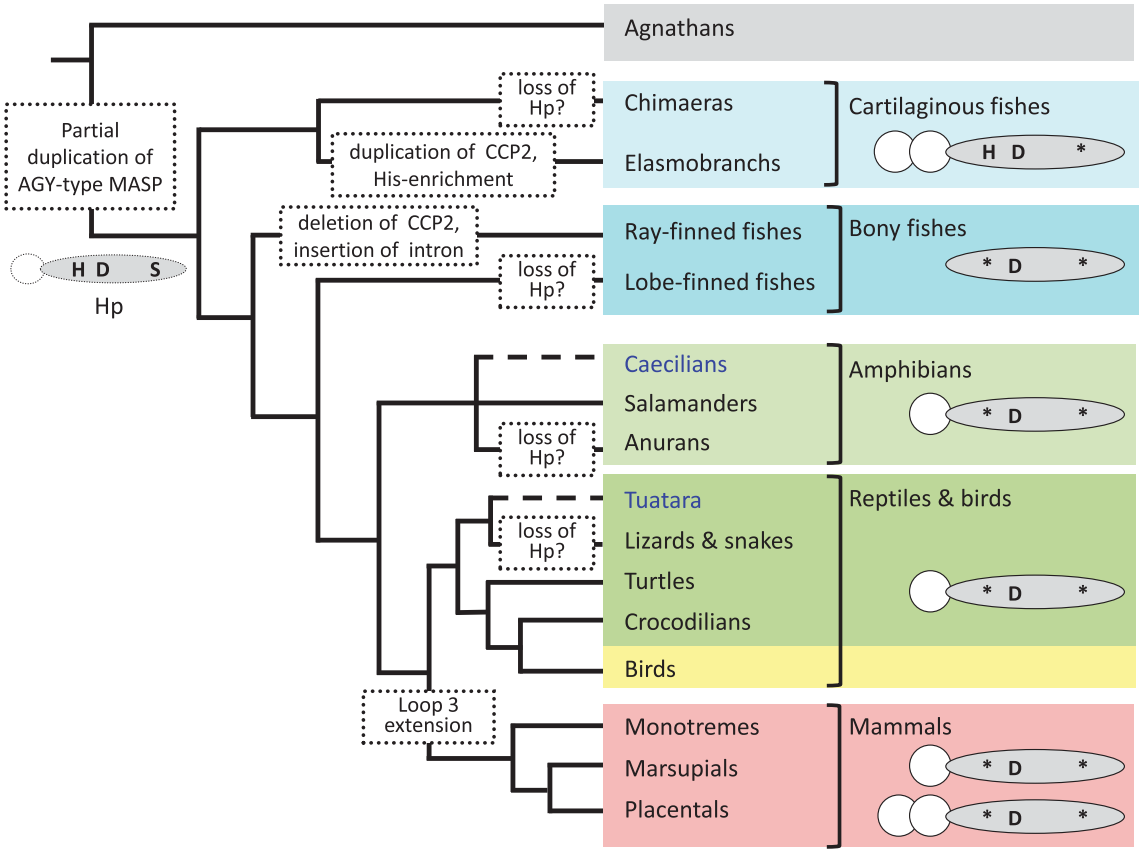


FIGURE 5. Phylogeny and proposed evolution of Hp. From the phylogeny of Hp, we propose that a partial gene duplication of an AGY-type MASP gave rise to the proto-Hp, composed of one SP domain (shaded oval) and a single CCP domain (CCP2; white circle), prior to the emergence of the jawed vertebrates. This "proto-Hp" likely carried a complete catalytic triad (H-D-S) in its SP domain; however, mutation (*) of the serine (S) residue soon after would have rendered the molecule proteolytically inactive. In cartilaginous fishes, the CCP domain was duplicated to give a CCP–CCP–SP structure, enabling formation of higher-order Hp oligomers, and became enriched for histidine. Bony fishes, in contrast, lost their CCP domain, meaning their Hp cannot oligomerize and is found as a monomer in their blood. Although some groups (Caecilians and Tuatara; highlighted in blue text) could not be assessed because of a lack of genomic data, from datasets currently available, it appears Hp has been lost in several lineages (notably the anuran amphibians, scaled reptiles, and possibly also chimera and lobe-finned fishes). Additional, independent, CCP duplication events occurred in several mammalian lineages (e.g., human, cow, and deer). The extended loop 3 region required for high affinity binding to both Hb and the scavenger receptor CD163 is only found in mammals; this implies that even if Hp in other vertebrate groups can bind free Hb, it is unlikely that the resultant complex is removed by the same mechanism.

[giving rise to the primate Hpr proteins (40)], the emergence of C1r-LP (Supplemental Fig. 4) allowing cleavage of pro-Hp prior to secretion (8), and extension of the loop 3 region of the SP domain (Supplemental Fig. 3). The loop 3 extension enabled Hp to bind Hb with high affinity by increasing the electrostatic pairing and surface area of interaction between the two molecules, as well as enabling the removal of the resultant Hp–Hb complex by the scavenger receptor CD163 (41) that had also newly emerged (42). Intriguingly, our data suggest the Hb–Hp–CD163 axis arose around the time RBCs became enucleated. Although loss of their nuclei allowed RBCs to increase intracellular Hb levels, and hence aerobic capacity (43), it also prevented them from producing proteins for the maintenance and repair of their cell membranes. Thus, they rupture in large numbers even under normal physiological conditions (44). The need to compensate for increased levels of free Hb would provide a strong selective force for the emergence of a more efficient binding and clearance mechanism: the role appropriated by Hp in mammals. In contrast, ectothermic vertebrates, with lower counts of less fragile, nucleated RBCs (45, 46) within lower pressure circulatory systems, likely have much lower rates of RBC lysis. Further, as the subunits of other vertebrate Hbs are bound together more strongly than those of mammals (47, 48), we would expect less dissociation following hemolysis, and consequently only small amounts of dimeric Hb available for capture by Hp (49). In such species, the heme scavenger Hx, alone or in combination with other, more passive, systems [e.g., membrane-anchored scavenger receptors (50)], should be sufficient to cope with the expected lower levels of free Hb. Indeed, contrasting the high frequency of Hp loss events, Hx has been almost universally retained across vertebrate phylogeny (Supplemental Table I), supporting the idea that it is the main route of Hb sequestration and detoxification in many nonmammalian species.

Although the role(s) performed by Hp in nonmammal vertebrates awaits further investigation, given Hp's origin as a member of the MASP family, we anticipate they will be immune related. Indeed, this hypothesis is supported by the fact that even mammalian Hp has been ascribed additional, immunomodulatory functions, beyond the removal of Hb [e.g., regulation of endotoxin-induced inflammation (3) and suppression of T cell proliferation and Th2 cytokine release (4)]. Our work suggests these could be evolutionary more ancient functions, which were retained after Hb removal and became the primary role of Hp in mammals. Future work should aim to better understand the extent of Hp's functional repertoire, especially if off-target effects are to be minimized upon the clinical application of human Hp [e.g., to limit organ damage following blood transfusion or sickle cell crisis (reviewed in Ref. 51)].

Acknowledgments

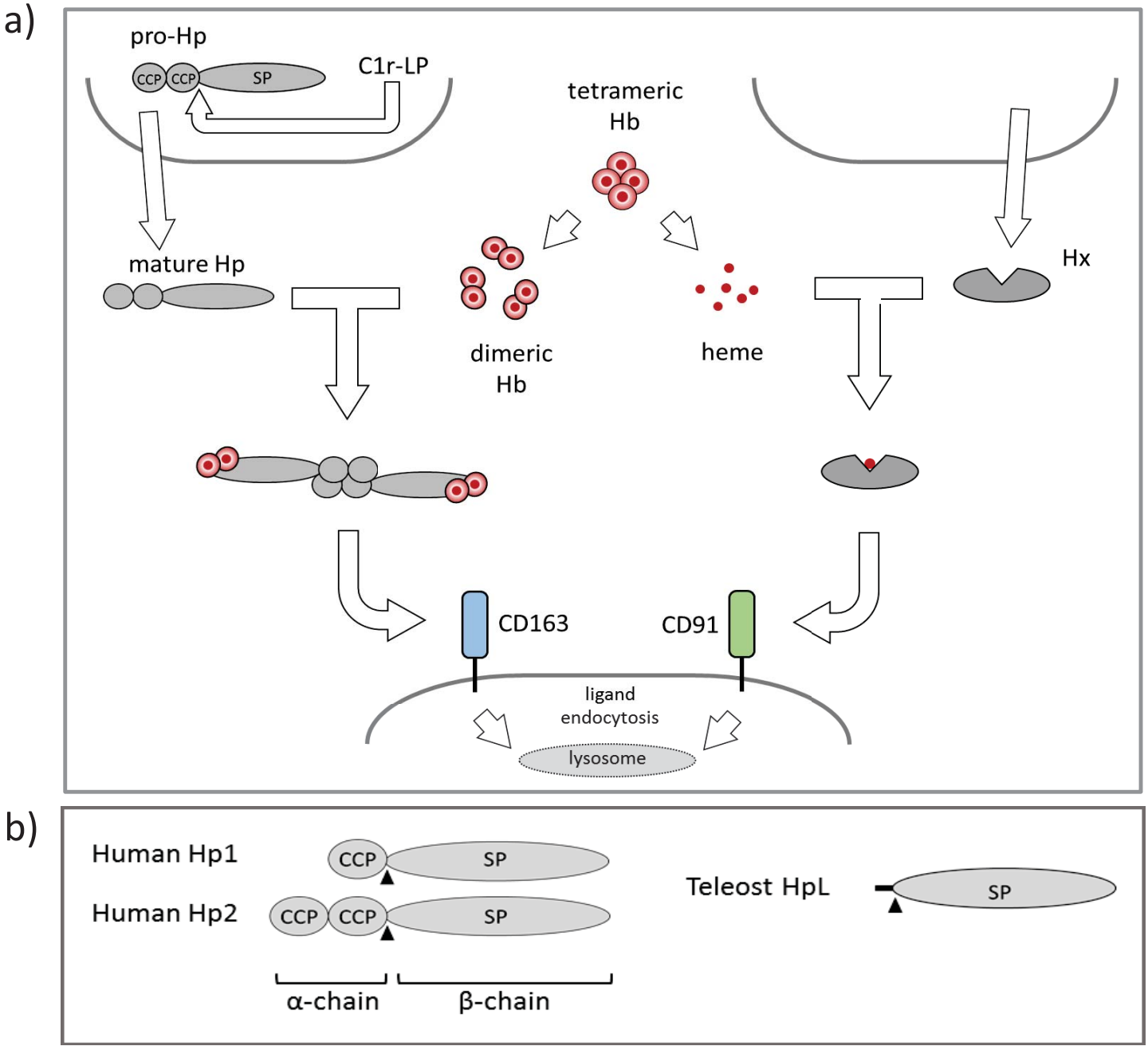
We thank E. Bryan Buckingham at the University of Maryland School of Medicine and the staff at the University of Aberdeen Proteomics Facility for technical assistance.

Disclosures

The authors have no financial conflicts of interest.

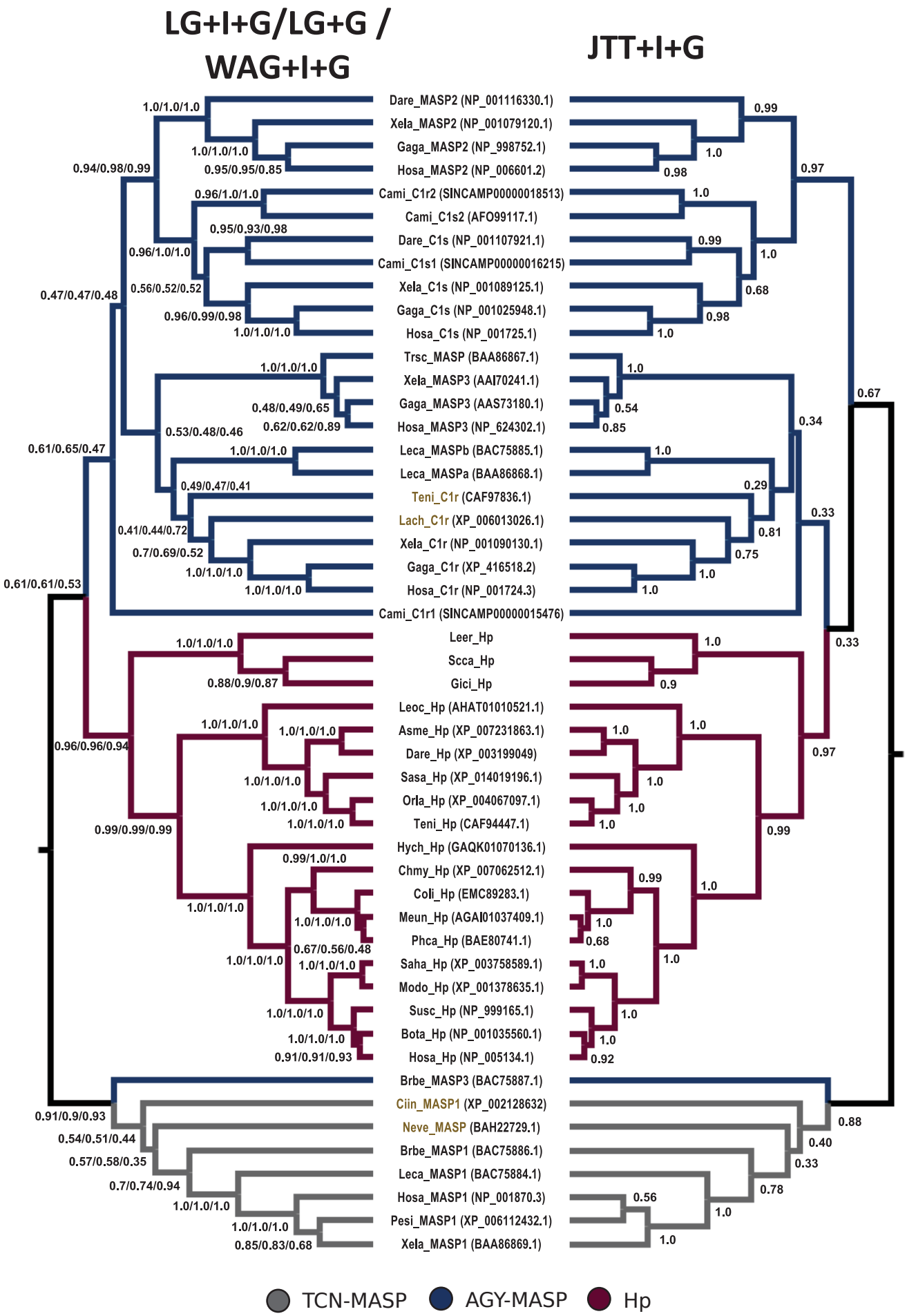
References

1. Polonovski, M., and M. Jayle. 1938. Existence in the blood plasma of a substance hastening the peroxydasic activity of the hemoglobin. *Cr Soc Biol* 129: 457–460.
2. Huntoon, K. M., Y. Wang, C. A. Eppolito, K. W. Barbour, F. G. Berger, P. A. Shrikant, and H. Baumann. 2008. The acute phase protein haptoglobin regulates host immunity. *J. Leukoc. Biol.* 84: 170–181.
3. Arredouani, M. S., A. Kasran, J. A. Vanoirbeek, F. G. Berger, H. Baumann, and J. L. Ceuppens. 2005. Haptoglobin dampens endotoxin-induced inflammatory effects both in vitro and in vivo. *Immunology* 114: 263–271.
4. Arredouani, M., P. Matthijs, E. Van Hoeyveld, A. Kasran, H. Baumann, J. L. Ceuppens, and E. Stevens. 2003. Haptoglobin directly affects T cells and suppresses T helper cell type 2 cytokine release. *Immunology* 108: 144–151.
5. Levy, A. P., R. Asleh, S. Blum, N. S. Levy, R. Miller-Lotan, S. Kalet-Litman, Y. Anbinder, O. Lache, F. M. Nakhoul, R. Asaf, et al. 2010. Haptoglobin: basic and clinical aspects. *Antioxid. Redox Signal.* 12: 293–304.
6. Kristiansen, M., J. H. Graversen, C. Jacobsen, O. Sonne, H. J. Hoffman, S. K. Law, and S. K. Moestrup. 2001. Identification of the haemoglobin scavenger receptor. *Nature* 409: 198–201.
7. Smith, A., and R. J. McCulloh. 2015. Hemopexin and haptoglobin: allies against heme toxicity from hemoglobin not contenders. *Front. Physiol.* 6: 187.
8. Wicher, K. B., and E. Fries. 2004. Prohaptoglobin is proteolytically cleaved in the endoplasmic reticulum by the complement C1r-like protein. *Proc. Natl. Acad. Sci. USA* 101: 14390–14395.
9. Wicher, K. B., and E. Fries. 2006. Haptoglobin, a hemoglobin-binding plasma protein, is present in bony fish and mammals but not in frog and chicken. *Proc. Natl. Acad. Sci. USA* 103: 4168–4173.
10. Andersen, C. B., M. Torvund-Jensen, M. J. Nielsen, C. L. de Oliveira, H. P. Hersleth, N. H. Andersen, J. S. Pedersen, G. R. Andersen, and S. K. Moestrup. 2012. Structure of the haptoglobin-haemoglobin complex. *Nature* 489: 456–459.
11. Maeda, N., F. Yang, D. R. Barnett, B. H. Bowman, and O. Smithies. 1984. Duplication within the haptoglobin Hp2 gene. *Nature* 309: 131–135.
12. Wejman, J. C., D. Hovsepian, J. S. Wall, J. F. Hainfeld, and J. Greer. 1984. Structure and assembly of haptoglobin polymers by electron microscopy. *J. Mol. Biol.* 174: 343–368.
13. Tosi, M., C. Duponchel, T. Meo, and E. Couture-Tosi. 1989. Complement genes C1r and C1s feature an intronless serine protease domain closely related to haptoglobin. *J. Mol. Biol.* 208: 709–714.
14. Wicher, K. B., and E. Fries. 2010. Evolutionary aspects of hemoglobin scavengers. *Antioxid. Redox Signal.* 12: 249–259.
15. Endo, Y., M. Takahashi, M. Nakao, H. Saiga, H. Sekine, M. Matsushita, M. Nonaka, and T. Fujita. 1998. Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates. *J. Immunol.* 161: 4924–4930.
16. Liang, C. C. 1957. The formation of complexes between haemoglobins and plasma proteins in a variety of animals. *Biochem. J.* 66: 552–558.
17. Mumm, D. P., D. H. Atha, and A. Riggs. 1978. The hemoglobin of the common sting-ray, Dasyatis sabina: structural and functional properties. *Comp. Biochem. Physiol. B* 60: 189–193.
18. Dooley, H., and M. F. Flajnik. 2005. Shark immunity bites back: affinity maturation and memory response in the nurse shark, Ginglymostoma cirratum. *Eur. J. Immunol.* 35: 936–945.
19. Löytynoja, A., and N. Goldman. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320: 1632–1635.
20. Chang, J. M., P. Di Tommaso, V. Lefort, O. Gascuel, and C. Notredame. 2015. TCS: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction. *Nucleic Acids Res.* 43(W1): W3–W6.
21. Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32: 268–274.
22. Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969–1973.
23. Vazquez-Moreno, L., J. Porath, S. F. Schluter, and J. J. Marchalonis. 1992. Purification of a novel heterodimer from shark (Carcharhinus plumbeus) serum by gel-immobilized metal chromatography. *Comp. Biochem. Physiol. B* 103: 563–568.
24. Wyffels, J., B. L. King, J. Vincent, C. Chen, C. H. Wu, and S. W. Polson. 2014. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. *F1000 Res.* 3: 191.
25. Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4: e88.
26. Wicher, K. B., and E. Fries. 2007. Convergent evolution of human and bovine haptoglobin: partial duplication of the genes. *J. Mol. Evol.* 65: 373–379.
27. Polticelli, F., A. Bocedi, G. Minervini, and P. Ascenzi. 2008. Human haptoglobin structure and function--a molecular modelling study. *FEBS J.* 275: 5648–5656.
28. Dooley, H., E. B. Buckingham, M. F. Criscitiello, and M. F. Flajnik. 2010. Emergence of the acute-phase protein hemopexin in jawed vertebrates. *Mol. Immunol.* 48: 147–152.
29. Macdonald, R. A., C. S. Hosking, and C. L. Jones. 1988. The measurement of relative antibody affinity by ELISA using thiocyanate elution. *J. Immunol. Methods* 106: 191–194.
30. McCloskey, N., M. W. Turner, and T. D. Goldblatt. 1997. Correlation between the avidity of mouse-human chimeric IgG subclass monoclonal antibodies measured by solid-phase elution ELISA and biospecific interaction analysis (BIA). *J. Immunol. Methods* 205: 67–72.
31. Hwang, P. K., and J. Greer. 1980. Interaction between hemoglobin subunits in the hemoglobin. haptoglobin complex. *J. Biol. Chem.* 255: 3038–3041.
32. Nantasenamat, C., V. Prachayasittikul, and L. Bulow. 2013. Molecular modeling of the human hemoglobin-haptoglobin complex sheds light on the protective mechanisms of haptoglobin. *PLoS One* 8: e62996.
33. Perona, J. J., and C. S. Craik. 1997. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem.* 272: 29987–29990.
34. Nielsen, M. J., S. V. Petersen, C. Jacobsen, S. Thirup, J. J. Enghild, J. H. Graversen, and S. K. Moestrup. 2007. A unique loop extension in the serine protease domain of haptoglobin is essential for CD163 recognition of the haptoglobin-hemoglobin complex. *J. Biol. Chem.* 282: 1072–1079.
35. Nielsen, M. J., and S. K. Moestrup. 2009. Receptor targeting of hemoglobin mediated by the haptoglobins: roles beyond heme scavenging. *Blood* 114: 764–771.
36. Venkatraman Girija, U., A. R. Gingras, J. E. Marshall, R. Panchal, M. A. Sheikh, J. A. Harper, P. Gál, W. J. Schwaeble, D. A. Mitchell, P. C. Moody, and R. Wallis. 2013. Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation. [Published erratum appears in 2016 *Proc. Natl. Acad. Sci. USA* 113: E6722–E6723.] *Proc. Natl. Acad. Sci. USA* 110: 13916–13920.
37. Pettinello, R., and H. Dooley. 2014. The immunoglobulins of cold-blooded vertebrates. *Biomolecules* 4: 1045–1069.
38. Poon, I. K., K. K. Patel, D. S. Davis, C. R. Parish, and M. D. Hulett. 2011. Histidine-rich glycoprotein: the swiss army knife of mammalian plasma. *Blood* 117: 2093–2101.
39. Morgan, W. T. 1981. Interactions of the histidine-rich glycoprotein of serum with metals. *Biochemistry* 20: 1054–1061.
40. McEvoy, S. M., and N. Maeda. 1988. Complex events in the evolution of the haptoglobin gene cluster in primates. *J. Biol. Chem.* 263: 15740–15747.
41. Nielsen, M. J., C. B. Andersen, and S. K. Moestrup. 2013. CD163 binding to haptoglobin-hemoglobin complexes involves a dual-point electrostatic receptor-ligand pairing. *J. Biol. Chem.* 288: 18834–18841.
42. Herzig, C. T., R. W. Waters, C. L. Baldwin, and J. C. Telfer. 2010. Evolution of the CD163 family and its relationship to the bovine gamma delta T cell co-receptor WC1. *BMC Evol. Biol.* 10: 181.
43. Snyder, G. K., and B. A. Sheafor. 1999. Red blood cells: centerpiece in the evolution of the vertebrate circulatory system. *Integr. Comp. Biol.* 39: 189–198.
44. Thomsen, J. H., A. Etzerodt, P. Svendsen, and S. K. Moestrup. 2013. The haptoglobin-CD163-heme oxygenase-1 pathway for hemoglobin scavenging. *Oxid. Med. Cell. Longev.* 2013: 523652.
45. Aldrich, K. J., D. K. Saunders, L. M. Sievert, and G. Sievert. 2006. Comparison of erythrocyte osmotic fragility among amphibians, reptiles, birds and mammals. *Trans. Kans. Acad. Sci.* 109: 149–158.
46. Arnold, J. E. 2005. Hematology of the sandbar shark, Carcharhinus plumbeus: standardization of complete blood count techniques for elasmobranchs. *Vet. Clin. Pathol.* 34: 115–123.
47. Edelstein, S. J., B. McEwen, and Q. H. Gibson. 1976. Subunit dissociation in fish hemoglobins. *J. Biol. Chem.* 251: 7632–7637.
48. Chiancone, E., P. Vecchini, L. Forlani, E. Antonini, and J. Wyman. 1966. Dissociation of hemoglobin from different animal species into subunits. *Biochim. Biophys. Acta* 127: 549–552.
49. Nagel, R. L., and Q. H. Gibson. 1971. The binding of hemoglobin to haptoglobin and its relation to subunit dissociation of hemoglobin. *J. Biol. Chem.* 246: 69–73.
50. Schaer, D. J., C. A. Schaer, P. W. Buehler, R. A. Boykins, G. Schoedon, A. I. Alayash, and A. Schaffner. 2006. CD163 is the macrophage scavenger receptor for native and chemically modified hemoglobins in the absence of haptoglobin. *Blood* 107: 373–380.
51. Andersen, C. B. F., K. Stødkilde, K. L. Sæderup, A. Kuhlee, S. Raunser, J. H. Graversen, and S. K. Moestrup. 2017. Haptoglobin. *Antioxid. Redox Signal.* 26: 814–831.

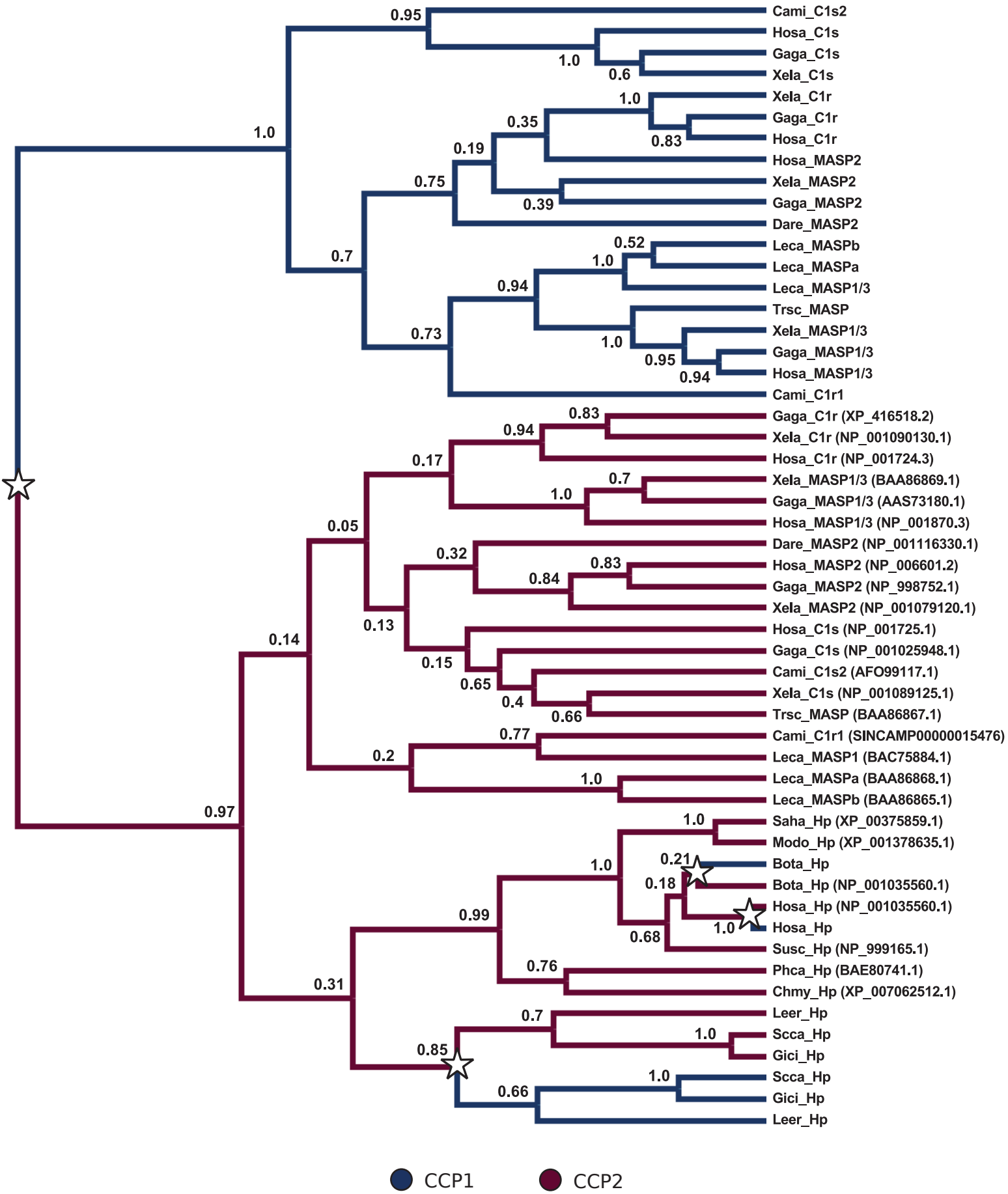


Supplemental figure 1: (a) In mammals the acute-phase plasma proteins haptoglobin (Hp) and hemopexin (Hx) have complementary functions to prevent heme-related oxidative damage. Hp is produced as a proprotein by hepatocytes and cleaved in the endoplasmic reticulum by C1r-like protein (C1r-LP). The resultant α- and β-chains disulphide-bond to generate the mature Hp protein that is secreted into the bloodstream. Hx, is similarly produced by hepatocytes, however requires no proteolytic processing before secretion. Following red blood cell (RBC) lysis the released haemoglobin (Hb) tetramers begin to degrade, dissociating into highly reactive Hb-dimers and releasing heme. Mature Hp binds with high affinity to dimeric Hb and, once complexed, binds to the scavenger receptor CD163 present on the surface of monocytes and macrophages and is internalized for lysosomal degradation and thus detoxification. Hx, in contrast, binds to free heme with extremely high affinity and is taken up via the scavenger receptor CD91 for detoxification. In humans the Hp pathway appears to be the primary protector against Hb-induced toxicity, with Hx providing backup when Hp is depleted [1]. **(b) Hp structure differs between species/individuals of the same species.** Hp is generally produced as a pro-protein containing one or two complement control protein (CCP) domains and an enzymatically-inactive serine protease (SP) domain. The SP domain mediates binding to both Hb and CD163, while the CCP domains dictate the oligomerization state of Hp in the blood. The human Hp gene exists in two major allelic forms, designated Hp1 and Hp2; the Hp1 allele has a lone CCP and forms disulphide-bonded dimers, while the Hp2 allele has two CCP domains and can form higher-order multimers. Heterozygous individuals (Hp1-2) exhibit a combination of Hp oligoforms [2]. In contrast, teleost fish Hp (HpL) has a short (20 aa) peptide instead of CCP domains and circulates as a monomer.

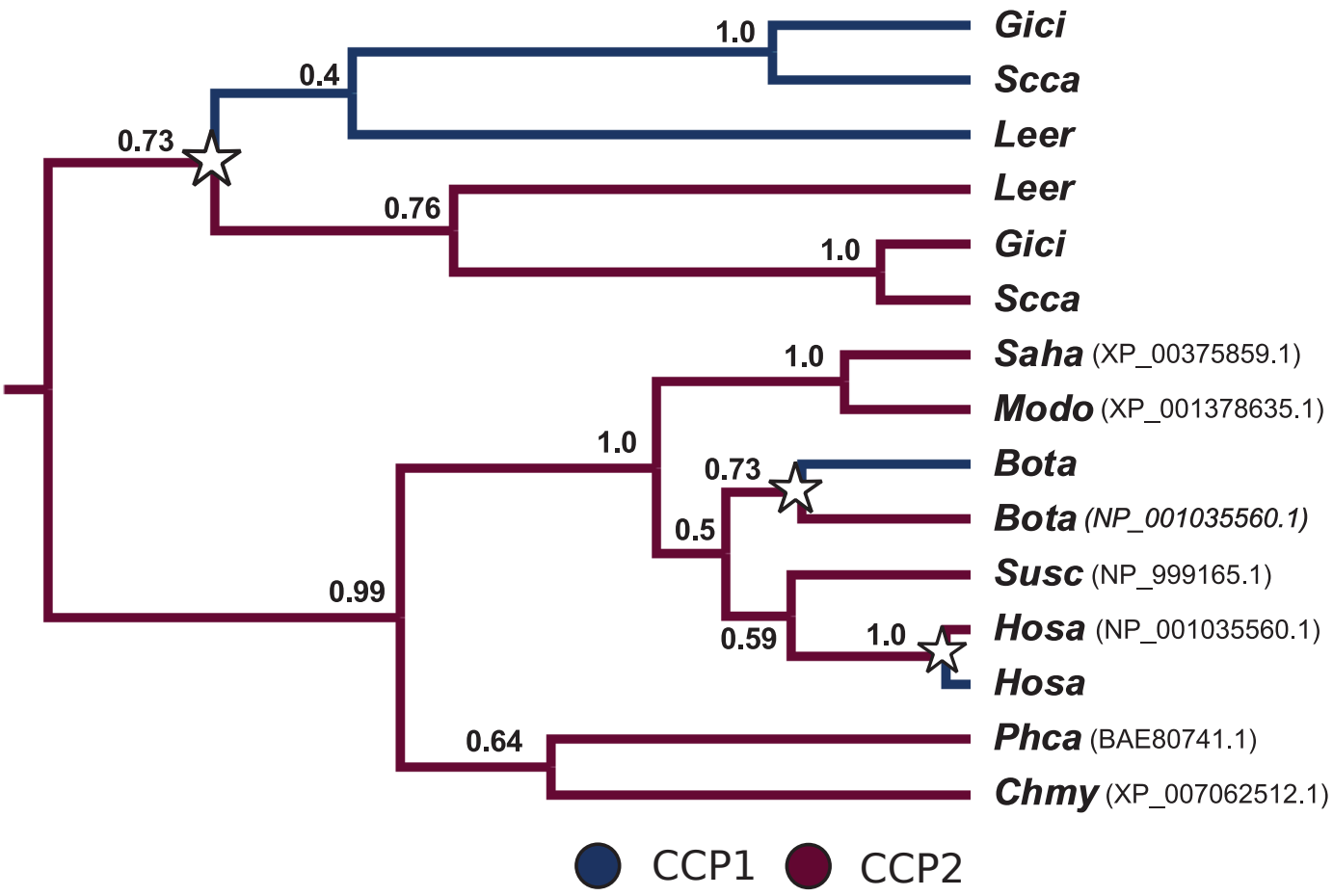
Supplemental figure 2a: MASP and HP SP domain phylogeny



Supplemental figure 2b: MASP and HP CCP domain phylogeny

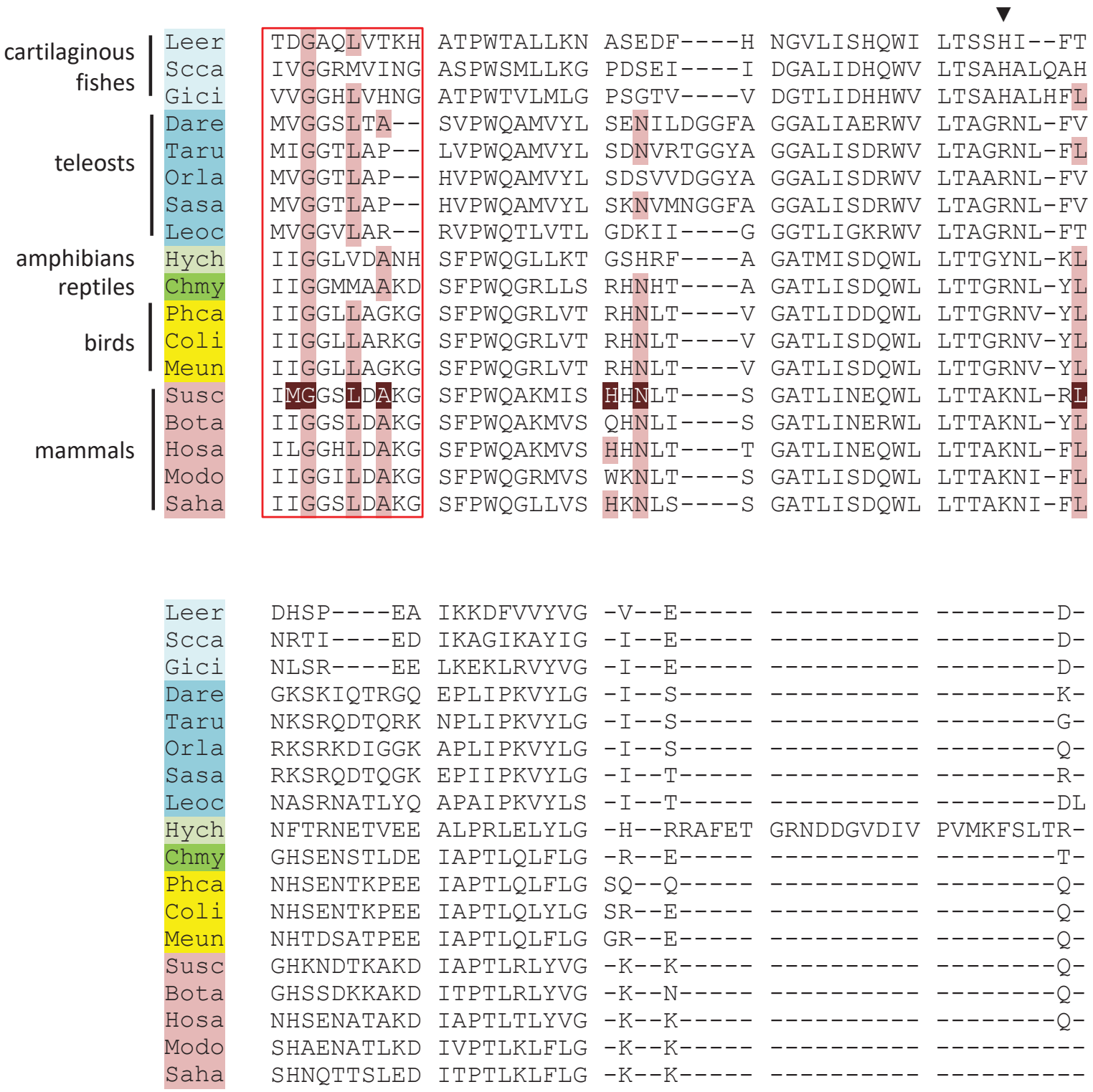


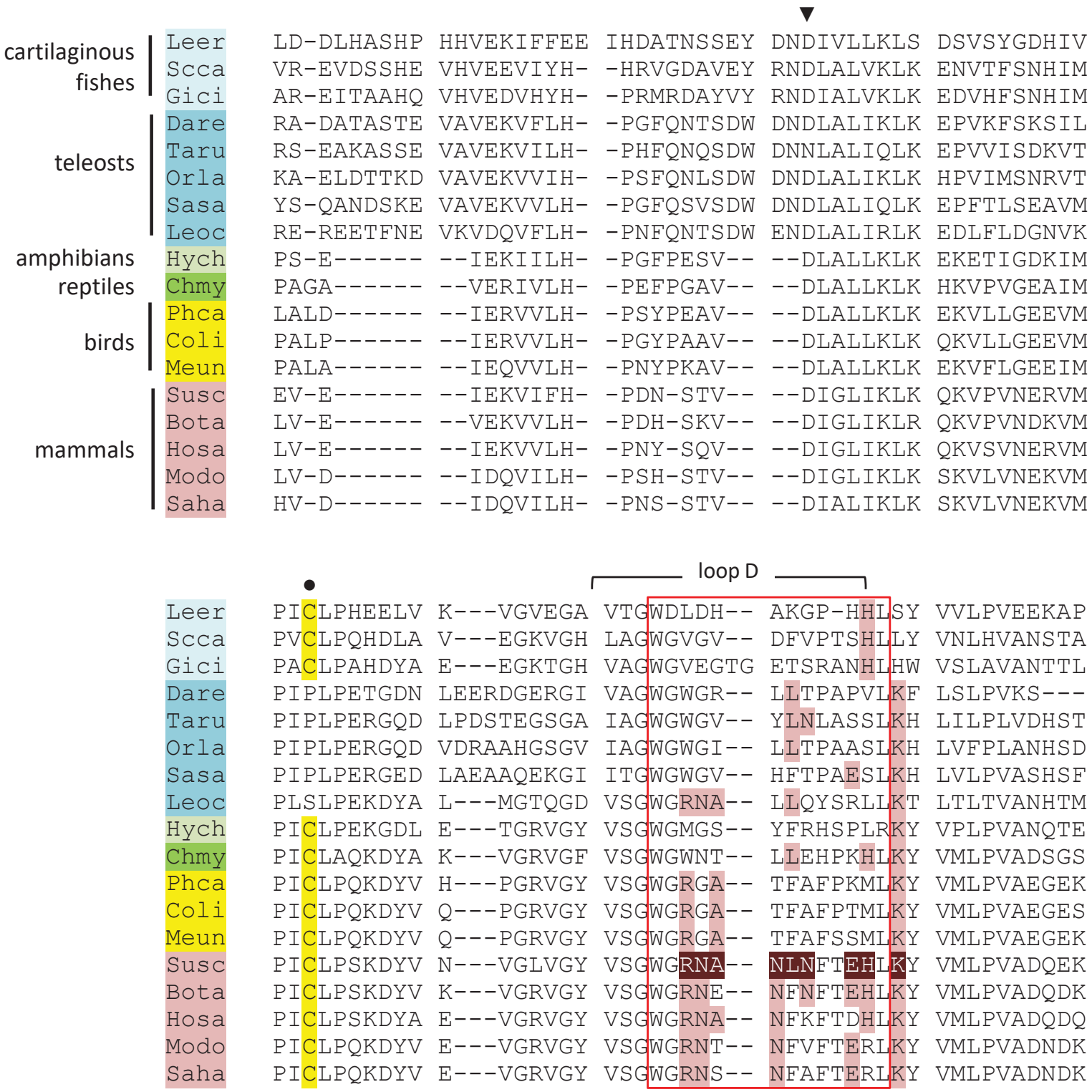
Supplemental figure 2c: HP CCP domain phylogeny

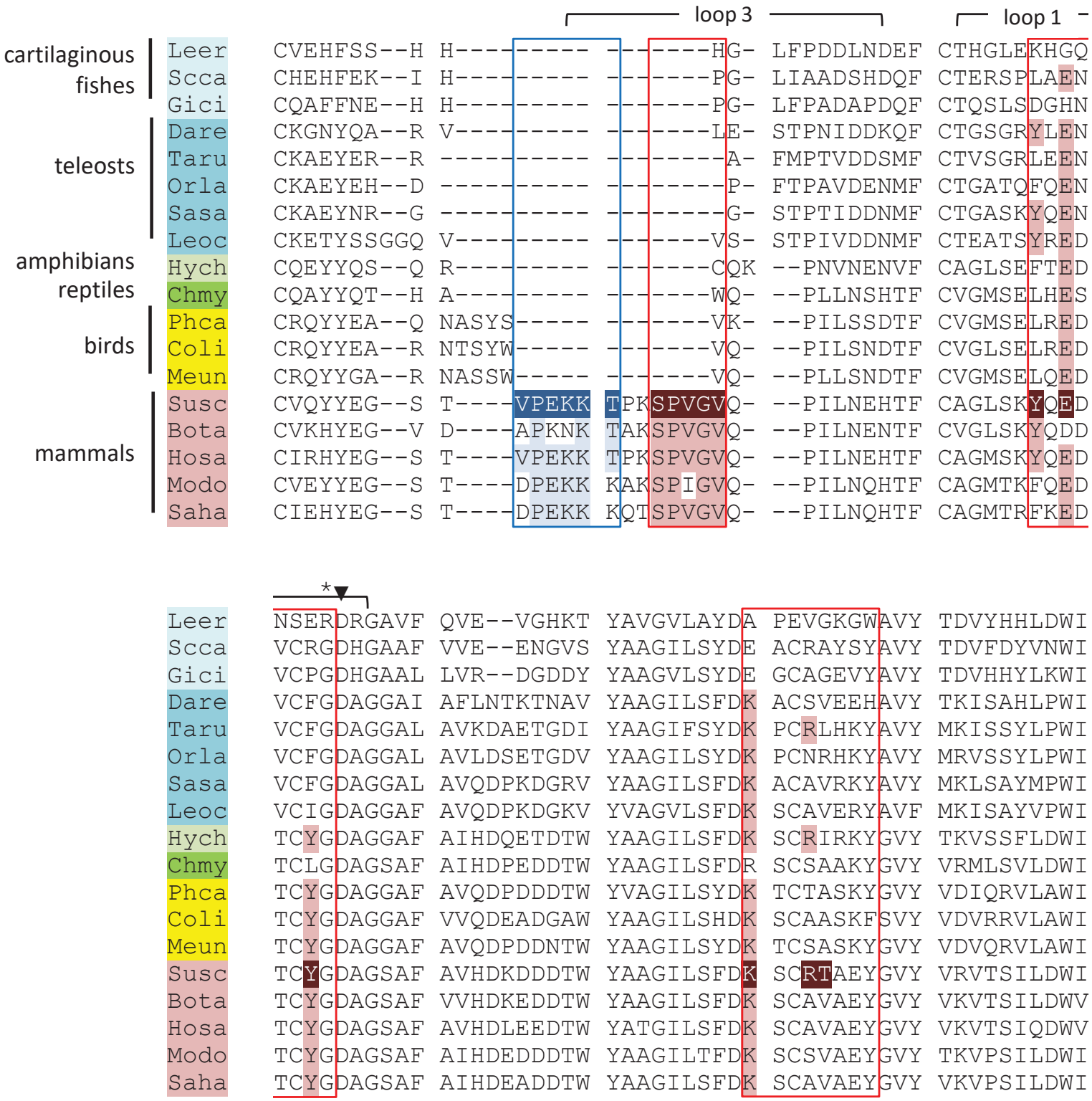


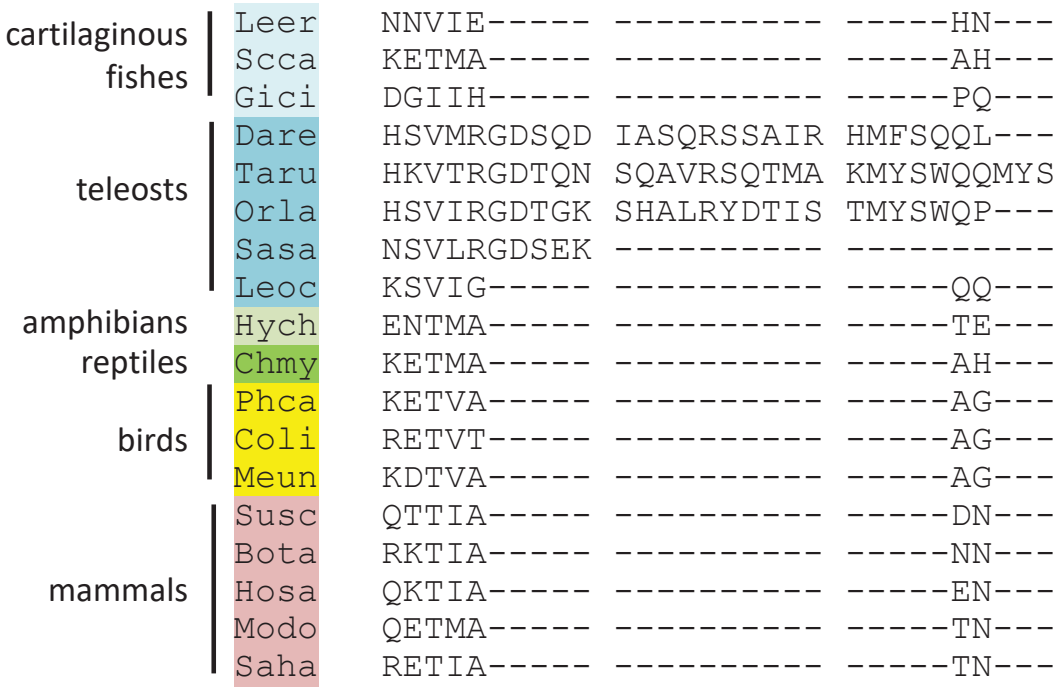
Supplemental figure 2: (a) Relaxed-clock rooted Bayesian phylogenetic analyses of Hp and the MASP family using multiple amino acid substitution models. The topology shown on the left is the maximum clade credibility tree for the best fitting model; LG+I+G. Posterior probabilities for multiple models are displayed for each clade in the form: LG+I+G/LG+G/WAG+I+G. The maximum clade credibility tree generated under the JTT +I+G model is displayed on the right, along with associated posterior probabilities. For pairs of sequence titles shown in gold the branching order is reversed for all models except LG+I+G. AGY-type MASP branches are displayed in blue, TCN-type in grey, and Hp in red. **(b) Relaxed-clock rooted Bayesian maximum clade credibility tree of CCP domains of the MASP family including Hp and (c) of the Hp CCP domains alone.** CCP1 domain branches are displayed in blue, while those of CCP2 are displayed in red. White stars represent domain duplication events. Posterior probabilities are shown for each clade. Four letter abbreviations for genus and species are used as detailed in supplemental table 1.

Supplemental figure 3



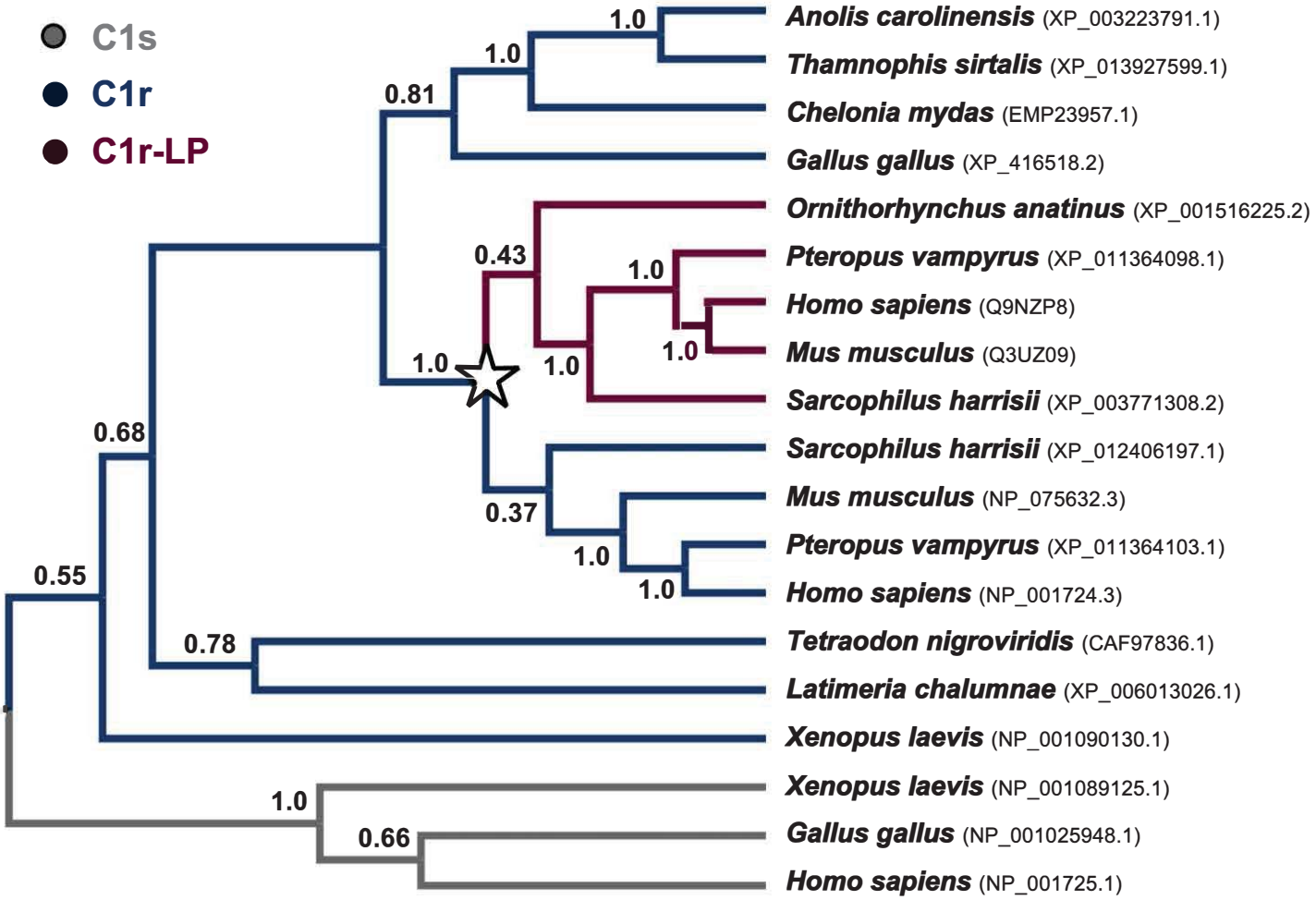

cartilaginous fishes
teleosts
amphibians reptiles
birds
mammals
Leer LD-DLHASHP HHVEKIFFEE IHDATNSSEY DNDIVLLKLS DSVSYGDHIV
Scca VR-EVDSSHE VHVEEVIYH- -HRVGDAVEY RNDLALVKLK ENVTFSNHIM
Gici AR-EITAAHQ VHVEDVHYH- -PRMRDAYVY RNDIALVKLK EDVHFSNHIM
Dare RA-DATASTE VAVEKVFLH- -PGFQNTSDW DNDLALIKLK EPVKFSKSIL
Taru RS-EAKASSE VAVEKVILH- -PHFQNQSDW DNNLALIQLK EPVVISDKVT
Orla KA-ELDTTKD VAVEKVVIH- -PSFQNLSDW DNDLALIKLK HPVIMSNRVT
Sasa YS-QANDSKE VAVEKVVLH- -PGFQSVSDW DNDLALIQLK EPFTLSEAVM
Leoc RE-REETFNE VKVDQVFLH- -PNFQNTSDW ENDLALIRLK EDLFLDGNVK
Hych PS-E------ --IEKIILH- -PGFPESV-- --DLALLKLK EKETIGDKIM
Chmy PAGA------ --VERIVLH- -PEFPGAV-- --DLALLKLK HKVPVGEAIM
Phca LALD------ --IERVVLH- -PSYPEAV-- --DLALLKLK EKVLLGEEVM
Coli PALP------ --IERVVLH- -PGYPAAV-- --DLALLKLK QKVLLGEEVM
Meun PALA------ --IEQVVLH- -PNYPKAV-- --DLALLKLK EKVFLGEEIM
Susc EV-E------ --IEKVIFH- -PDN-STV-- --DIGLIKLK QKVPVNERVM
Bota LV-E------ --VEKVVLH- -PDH-SKV-- --DIGLIKLR QKVPVNDKVM
Hosa LV-E------ --IEKVVLH- -PNY-SQV-- --DIGLIKLK QKVSVNERVM
Modo LV-D------ --IDQVILH- -PSH-STV-- --DIGLIKLK SKVLVNEKVM
Saha HV-D------ --IDQVILH- -PNS-STV-- --DIALIKLK SKVLVNEKVM
loop D
Leer PICLPHEELV K---VGVEGA VTGWDLDH-- AKGP-HHLSY VVLPVEEKAP
Scca PVCLPQHDLA V---EGKVGH LAGWGVGV-- DFVPTSHLLY VNLHVANSTA
Gici PACLPAHDYA E---EGKTGH VAGWGVEGTG ETSRANHLHW VSLAVANTTL
Dare PIPLPETGDN LEERDGERGI VAGWGWGR-- LLTPAPVLKF LSLPVKS---
Taru PIPLPERGQD LPDSTEGSGA IAGWGWGV-- YLNLASSLKH LILPLVDHST
Orla PIPLPERGQD VDRAAHGSGV IAGWGWGI-- LLTPAASLKH LVFPLANHSD
Sasa PIPLPERGED LAEAAQEKGI ITGWGWGV-- HFTPAESLKH LVLPVASHSF
Leoc PLSLPEKDYA L---MGTQGD VSGWGRNA-- LLQYSRLLKT LTLTVANHTM
Hych PICLPEKGDL E---TGRVGY VSGWGMGS-- YFRHSPLRKY VPLPVANQTE
Chmy PICLAQKDYA K---VGRVGF VSGWGWNT-- LLEHPKHLKY VMLPVADSGS
Phca PICLPQKDYV H---PGRVGY VSGWGRGA-- TFAFPKMLKY VMLPVAEGEK
Coli PICLPQKDYV Q---PGRVGY VSGWGRGA-- TFAFPTMLKY VMLPVAEGES
Meun PICLPQKDYV Q---PGRVGY VSGWGRGA-- TFAFSSMLKY VMLPVAEGEK
Susc PICLPSKDYV N---VGLVGY VSGWGRNA-- NLNFTEHLKY VMLPVADQEK
Bota PICLPSKDYV K---VGRVGY VSGWGRNE-- NFNFTEHLKY VMLPVADQDK
Hosa PICLPSKDYA E---VGRVGY VSGWGRNA-- NFKFTDHLKY VMLPVADQDQ
Modo PICLPQKDYV E---VGRVGY VSGWGRNT-- NFVFTERLKY VMLPVADNDK
Saha PICLPQKDYV E---VGRVGY VSGWGRNS-- NFAFTERLKY VMLPVADNDK


loop 3
loop 1
cartilaginous fishes
teleosts
amphibians reptiles
birds
mammals
Leer CVEHFSS--H H---------- --------HG- LFPDDLNDEF CTHGLEKHGQ
Scca CHEHFEK--I H---------- --------PG- LIAADSHDQF CTERSPLAEN
Gici CQAFFNE--H H---------- --------PG- LFPADAPDQF CTQSLSDGHN
Dare CKGNYQA--R V---------- --------LE- STPNIDDKQF CTGSGRYLEN
Taru CKAEYER--R ---------- ---------A- FMPTVDDSMF CTVSGRLEEN
Orla CKAEYEH--D ---------- ---------P- FTPAVDENMF CTGATQFQEN
Sasa CKAEYNR--G ---------- ---------G- STPTIDDNMF CTGASKYQEN
Leoc CKETYSSGGQ V---------- --------VS- STPIVDDNMF CTEATSYRED
Hych CQEYYQS--Q R---------- --------CQK --PNVNENVF CAGLSEFTED
Chmy CQAYYQT--H A---------- --------WQ- --PLLNSHTF CVGMSELHES
Phca CRQYYEA--Q NASYS------ --------VK- --PILSSDTF CVGMSELRED
Coli CRQYYEA--R NTSYW------ --------VQ- --PILSNDTF CVGLSELRED
Meun CRQYYGA--R NASSW------ --------VQ- --PLLSNDTF CVGMSELQED
Susc CVQYYEG--S T----VPEKK TPKSPVGVQ- --PILNEHTF CAGLSKYQED
Bota CVKHYEG--V D----APKNK TAKSPVGVQ- --PILNENTF CVGLSKYQDD
Hosa CIRHYEG--S T----VPEKK TPKSPVGVQ- --PILNEHTF CAGMSKYQED
Modo CVEYYEG--S T----DPEKK KAKSPIGVQ- --PILNQHTF CAGMTKFQED
Saha CIEHYEG--S T----DPEKK KQTSPVGVQ- --PILNQHTF CAGMTRFKED
*
Leer NSERDRGAVF QVE--VGHKT YAVGVLAYDA PEVGKGWAVY TDVYHHLDWI
Scca VCRGDHGAAF VVE--ENGVS YAAGILSYDE ACRAYSYAVY TDVFDYVNWI
Gici VCPGDHGAAL LVR--DGDDY YAAGVLSYDE GCAGEVYAVY TDVHHYLKWI
Dare VCFGDAGGAI AFLNTKTNAV YAAGILSFDK ACSVEEHAVY TKISAHLPWI
Taru VCFGDAGGAL AVKDAETGDI YAAGIFSYDK PCRLHKYAVY MKISSYLPWI
Orla VCFGDAGGAL AVLDSETGDV YAAGILSYDK PCNRHKYAVY MRVSSYLPWI
Sasa VCFGDAGGAL AVQDPKDGRV YAAGILSFDK ACAVRKYAVY MKLSAYMPWI
Leoc VCIGDAGGAF AVQDPKDGKV YVAGVLSFDK SCAVERYAVF MKISAYVPWI
Hych TCYGDAGGAF AIHDQETDTW YAAGILSFDK SCRIRKYGVY TKVSSFLDWI
Chmy TCLGDAGSAF AIHDPEDDTW YAAGILSFDR SCSAAKYGVY VRMLSVLDWI
Phca TCYGDAGGAF AVQDPDDDTW YVAGILSYDK TCTASKYGVY VDIQRVLAWI
Coli TCYGDAGGAF VVQDEADGAW YAAGILSHDK SCAASKFSVY VDVRRVLAWI
Meun TCYGDAGGAF AVQDPDDNTW YAAGILSYDK TCSASKYGVY VDVQRVLAWI
Susc TCYGDAGSAF AVHDKDDDTW YAAGILSFDK SCRTAEYGVY VRVTSILDWI
Bota TCYGDAGSAF VVHDKEDDTW YAAGILSFDK SCAVAEYGVY VKVTSILDWV
Hosa TCYGDAGSAF AVHDLEEDTW YATGILSFDK SCAVAEYGVY VKVTSIQDWV
Modo TCYGDAGSAF AIHDEDDDTW YAAGILTFDK SCSVAEYGVY TKVPSILDWI
Saha TCYGDAGSAF AIHDEADDTW YAAGILSFDK SCAVAEYGVY VKVPSILDWI



Supplemental figure 3: Multiple sequence alignment of Hp SP domains from across vertebrate phylogeny. Four letter abbreviations for genus and species are used (for details see supplemental table 1) and vertebrate groups are identified to the left of the alignment. Residues identified by Nantasenamat *et al*., [3] as important in Hp-Hb complex formation are boxed in red, while those confirmed as Hb-interacting by Andersen *et al*., [4] are highlighted in dark red on the pig (Susc) Hp sequence and a lighter shade of red where conserved in other species. The residues identified by Nielsen *et al*., [5] as important for CD163 binding by mammalian Hp are boxed in blue, with critical residues [6] shaded dark blue on the human sequence and a lighter shade of blue where conserved in other species. Loop designations (according to Perona & Craik [7]) are indicated above the alignment. The residues which form the catalytic triad (H-D-S; required for the proteolytic activity of the SP domain) in other MASP-family members are indicated by triangles above the alignment, the conserved Asp residue found at the base of the active-site cavity is marked with a star, while the cysteine that forms the interchain disulphide is highlighted in yellow.

Supplemental figure 4: C1r-LP phylogeny



Supplemental figure 4: Relaxed-clock rooted Bayesian maximum clade credibility tree of the C1r gene family showing the emergence of C1r-LP in the ancestor of mammals. Canonical C1r branches are displayed in blue, those of C1s in grey, and C1r-LP in red. A white star denotes the duplication of C1r giving rise to C1r-LP.

Supplemental figures reference list

1. Deuel, J.W., et al., Different target specificities of haptoglobin and hemopexin define a sequential protection system against vascular hemoglobin toxicity. Free Radical Biology and Medicine, 2015. 89: p. 931-943.
2. Langlois, M.R. and J.R. Delanghe, Biological and clinical significance of haptoglobin polymorphism in humans. Clin. Chem, 1996. 42(10): p. 1589-1600.
3. Nantasenamat, C., V. Prachayasittikul, and L. Bulow, Molecular modeling of the human hemoglobin-haptoglobin complex sheds light on the protective mechanisms of haptoglobin. PLoS One, 2013. 8(4): p. e62996.
4. Andersen, C.B., et al., Structure of the haptoglobin-haemoglobin complex. Nature, 2012. 489(7416): p. 456-9.
5. Nielsen, M.J., et al., A unique loop extension in the serine protease domain of haptoglobin is essential for CD163 recognition of the haptoglobin-hemoglobin complex. J Biol Chem, 2007. 282(2): p. 1072-9.
6. Nielsen, M.J., C.B. Andersen, and S.K. Moestrup, CD163 binding to haptoglobin-hemoglobin complexes involves a dual-point electrostatic receptor-ligand pairing. J Biol Chem, 2013. 288(26): p. 18834-41.
7. Perona, J.J. and C.S. Craik, Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. J Biol Chem, 1997. 272(48): p. 29987-90.