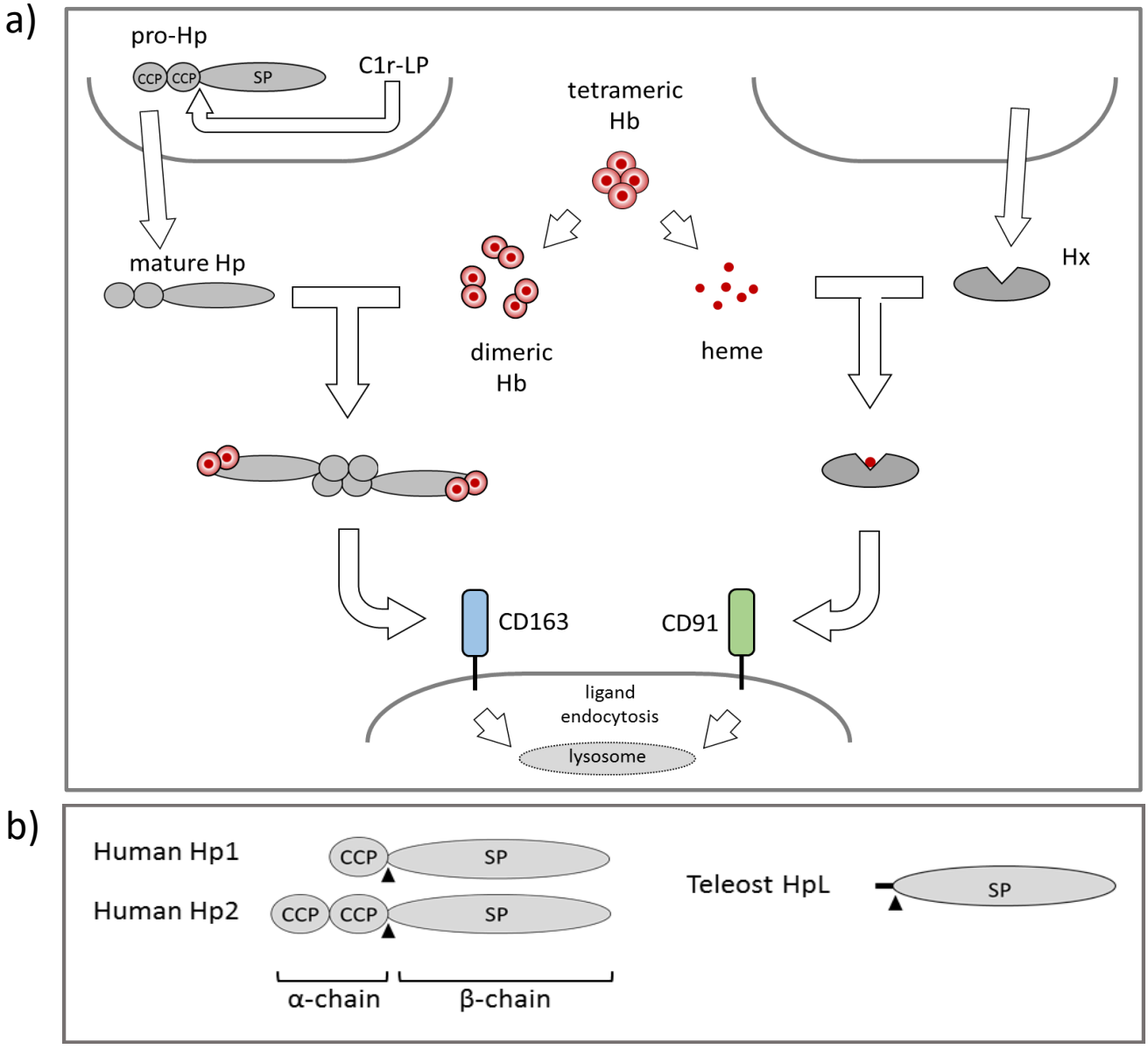
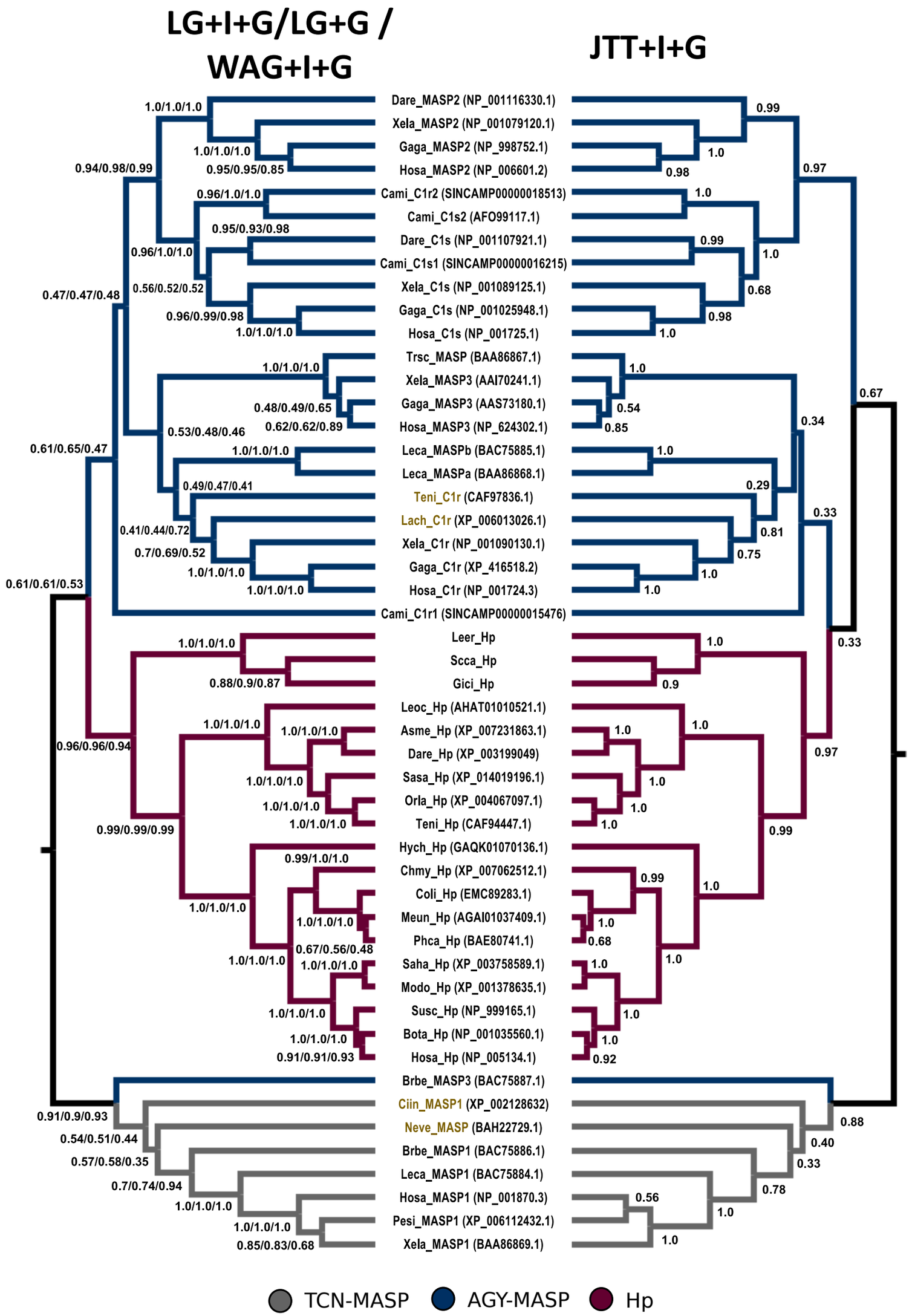
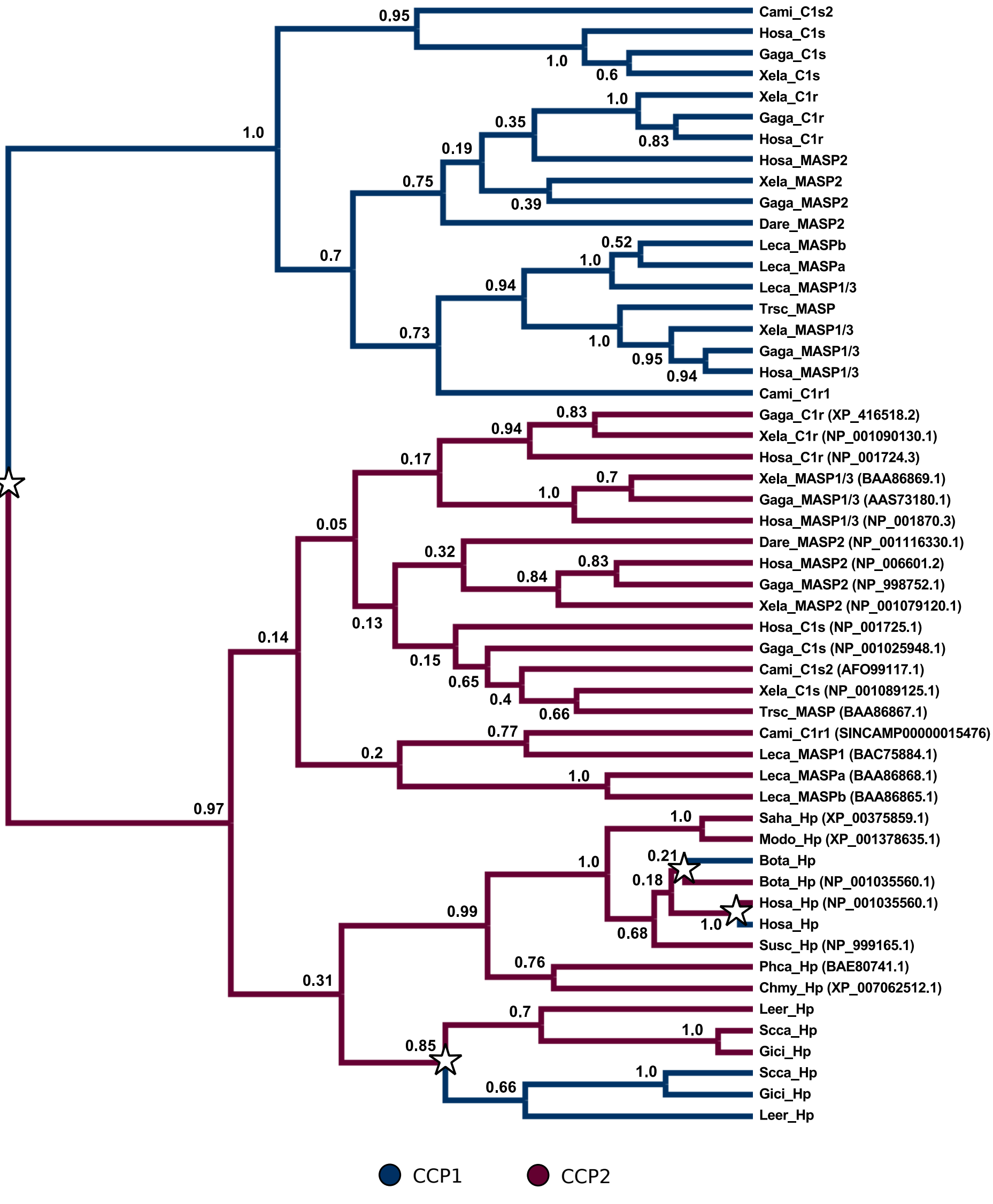


Supplemental figure 1: (a) In mammals the acute-phase plasma proteins haptoglobin (Hp) and hemopexin (Hx) have complementary functions to prevent heme-related oxidative damage. Hp is produced as a proprotein by hepatocytes and cleaved in the endoplasmic reticulum by C1r-like protein (C1r-LP). The resultant 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.

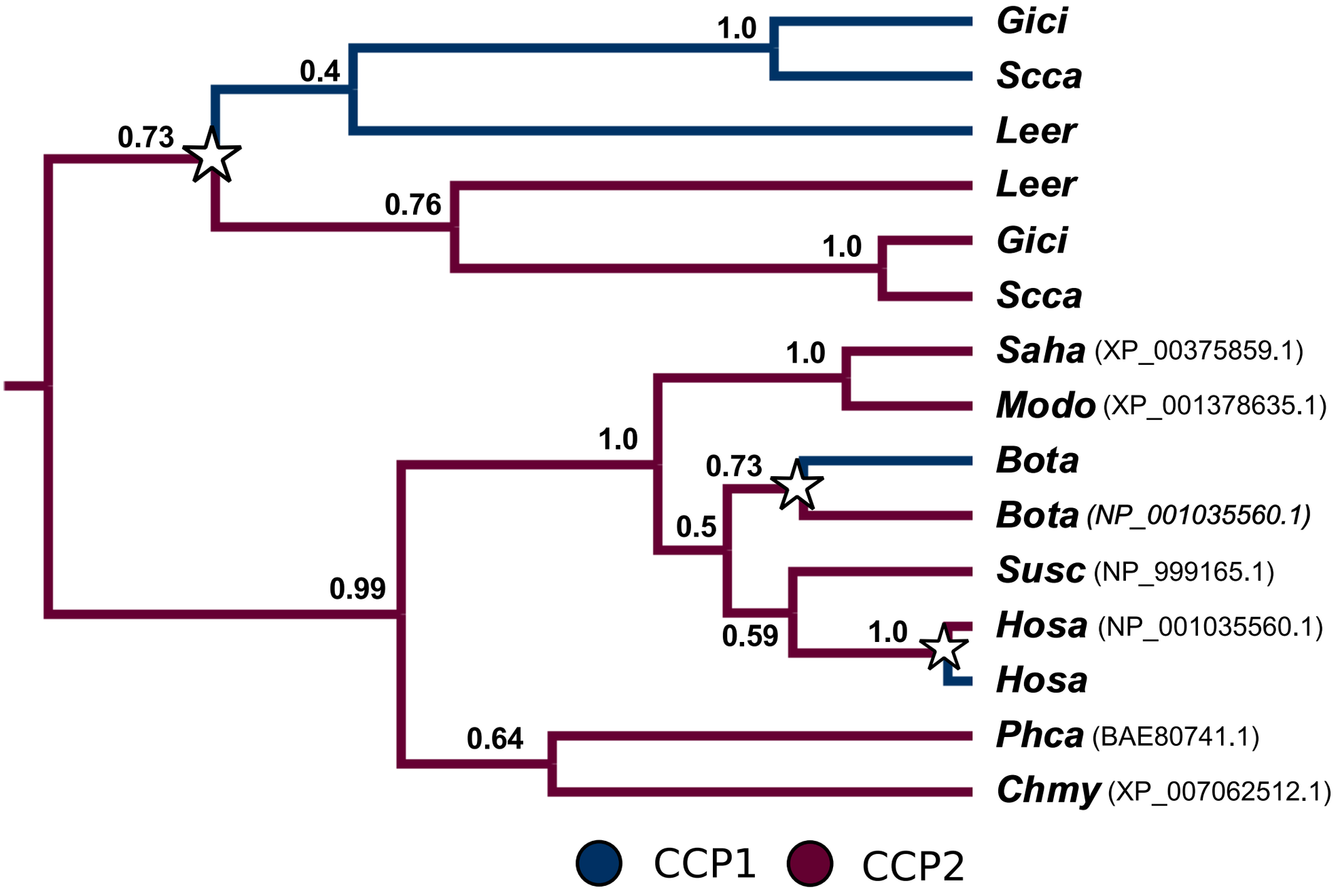
Supplemental figure 2a: MASP and HP SP domain phylogeny



Supplemental figure 2b: MASP and HP CCP domain phylogeny

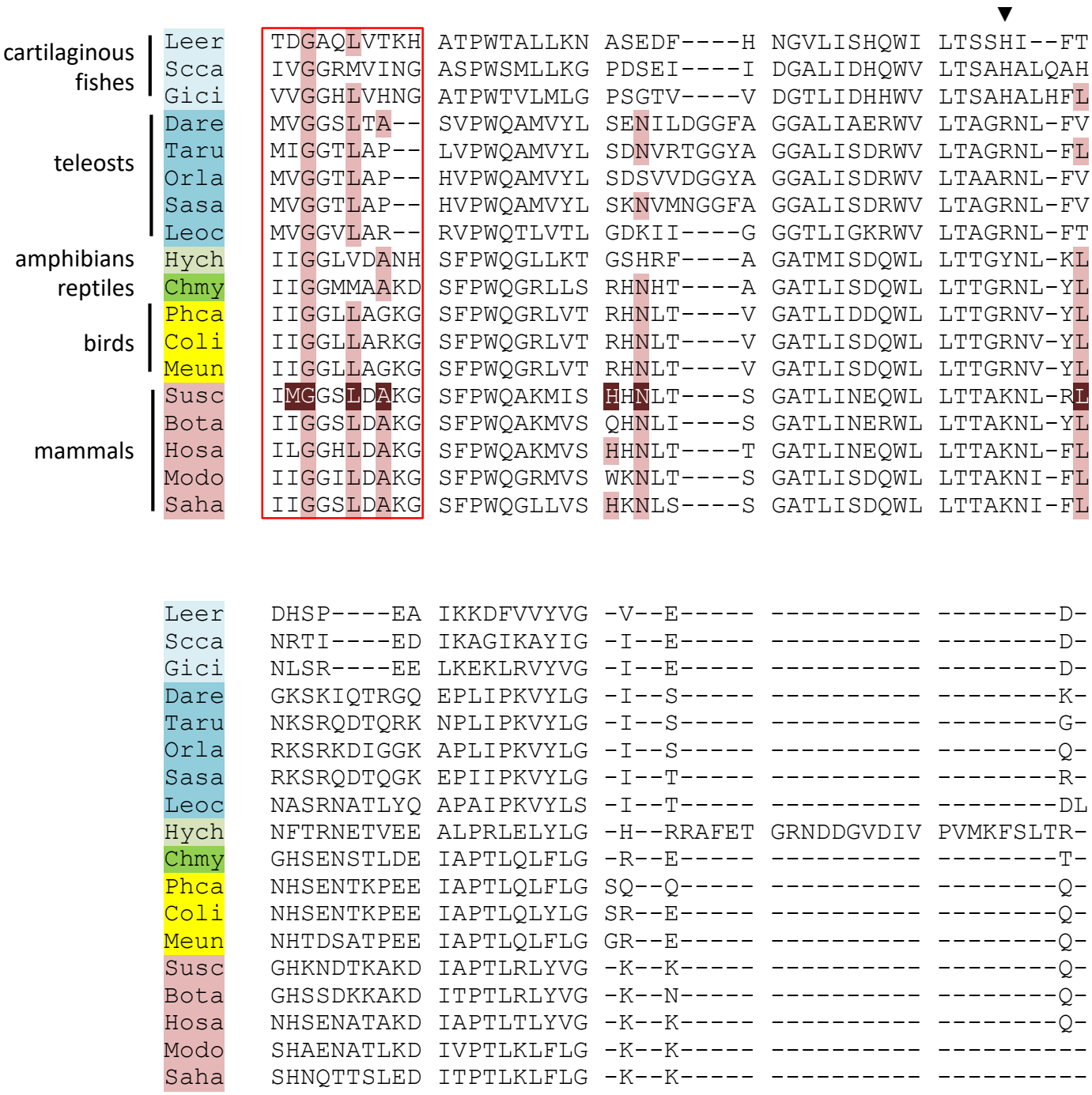


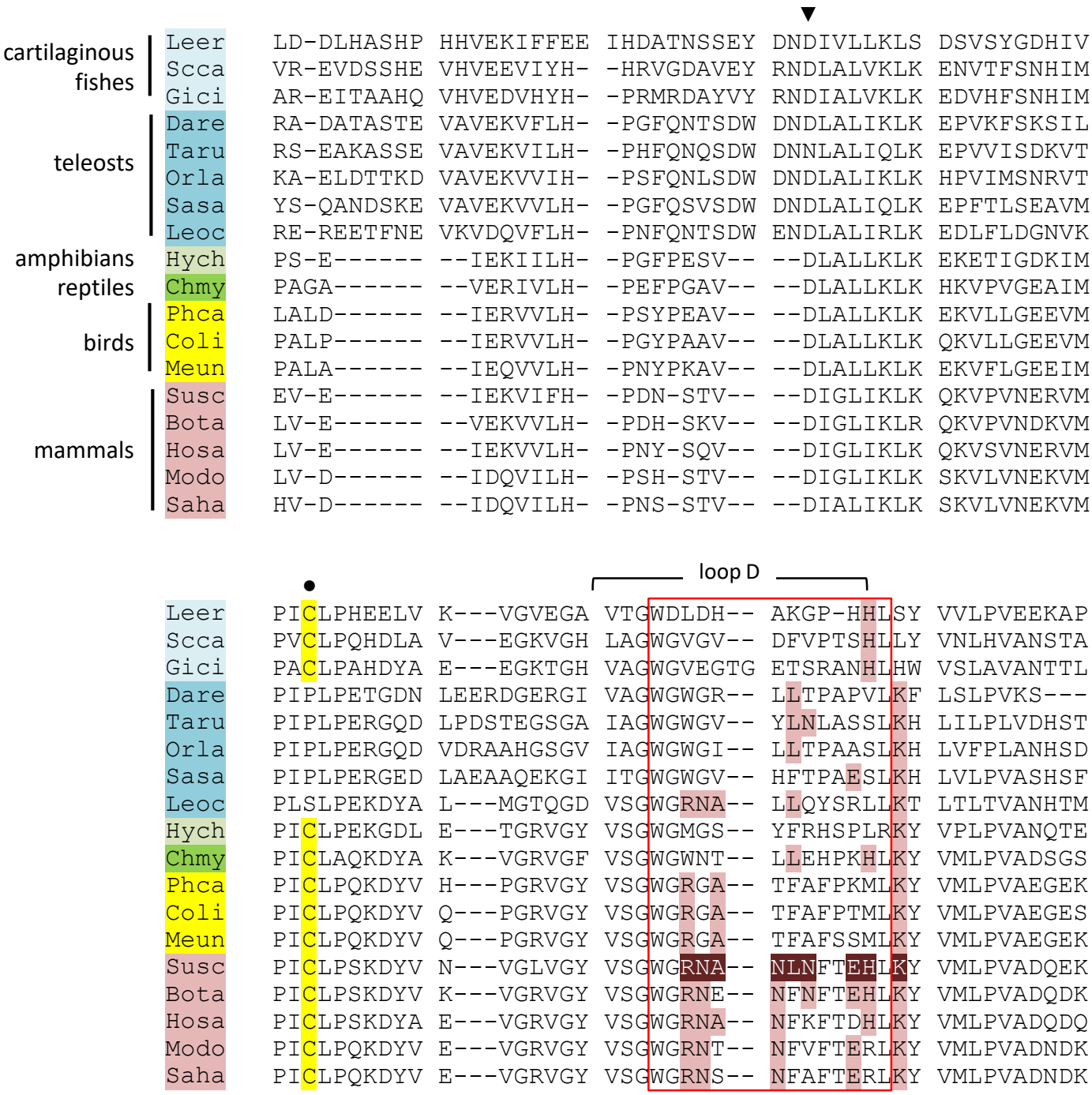
Supplemental figure 2c: HP CCP domain phylogeny

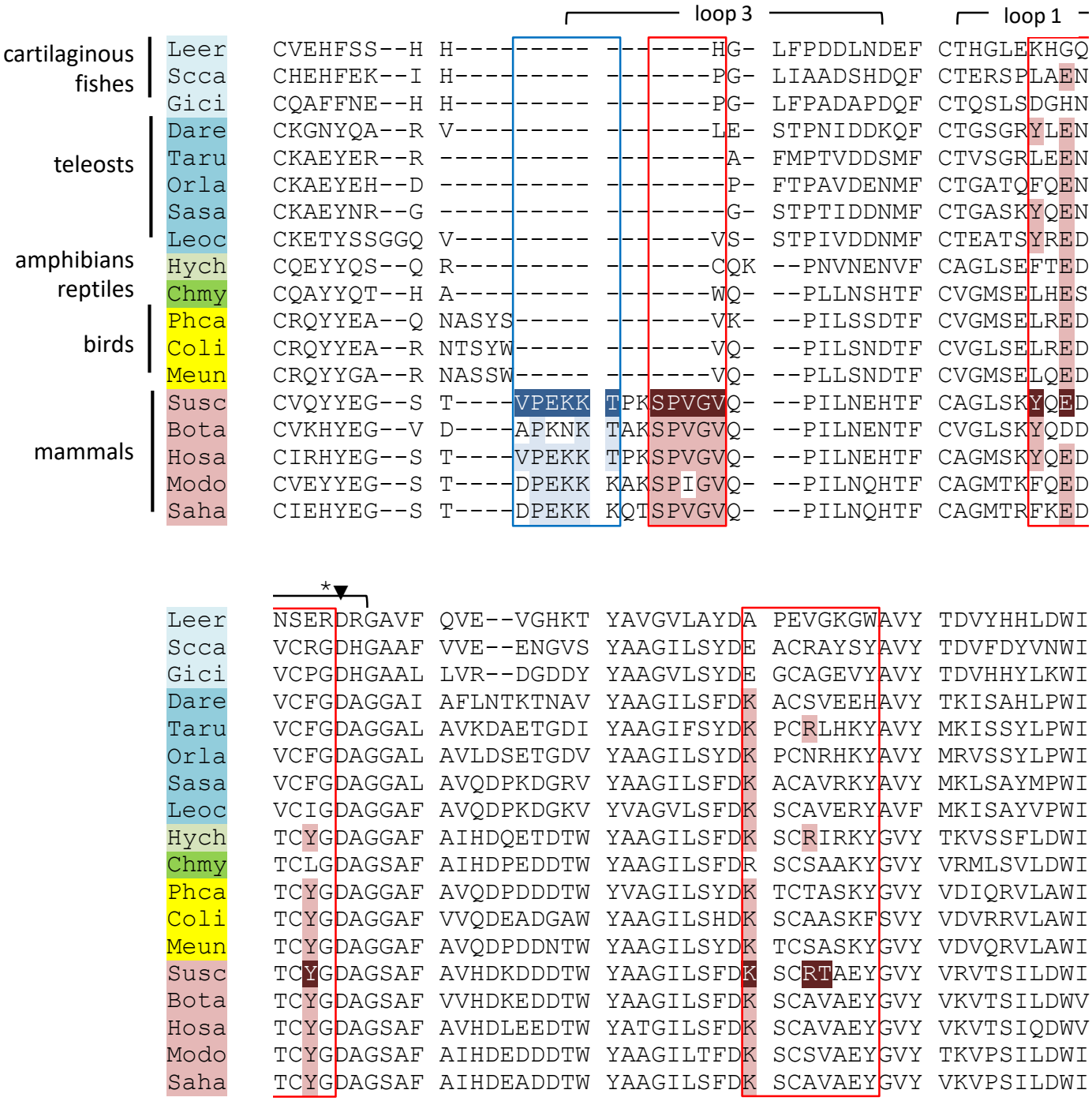


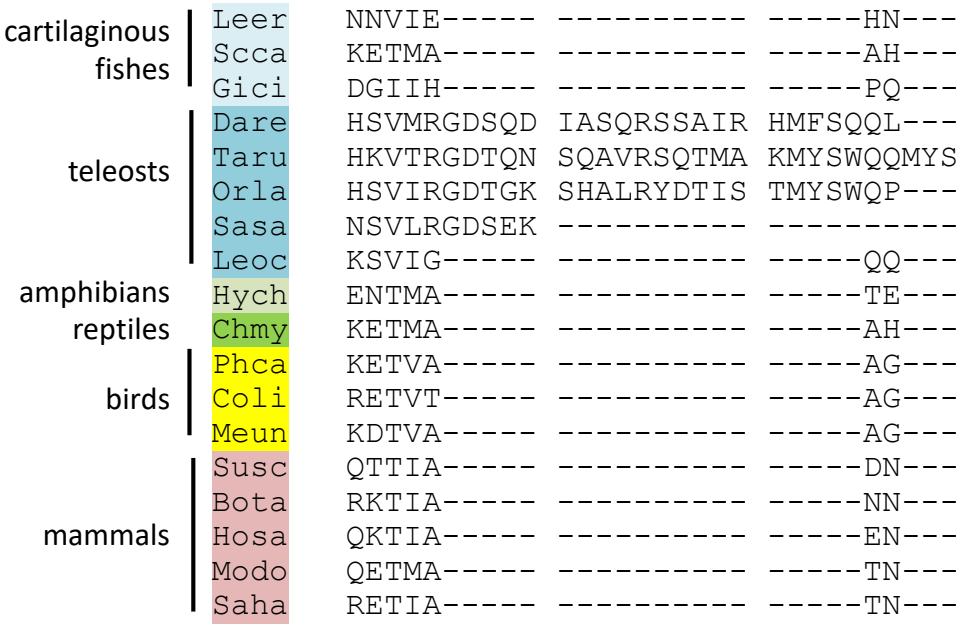
Supplemental figure 2: (a) Relaxed-clock rooted Bayesian phylogenetic analyses of Hp and the MASP family using multiple amino acid substitution models. The topology shown on the left is the maximum clade credibility tree for the best fitting model; LG+I+G. Posterior probabilities for multiple models 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



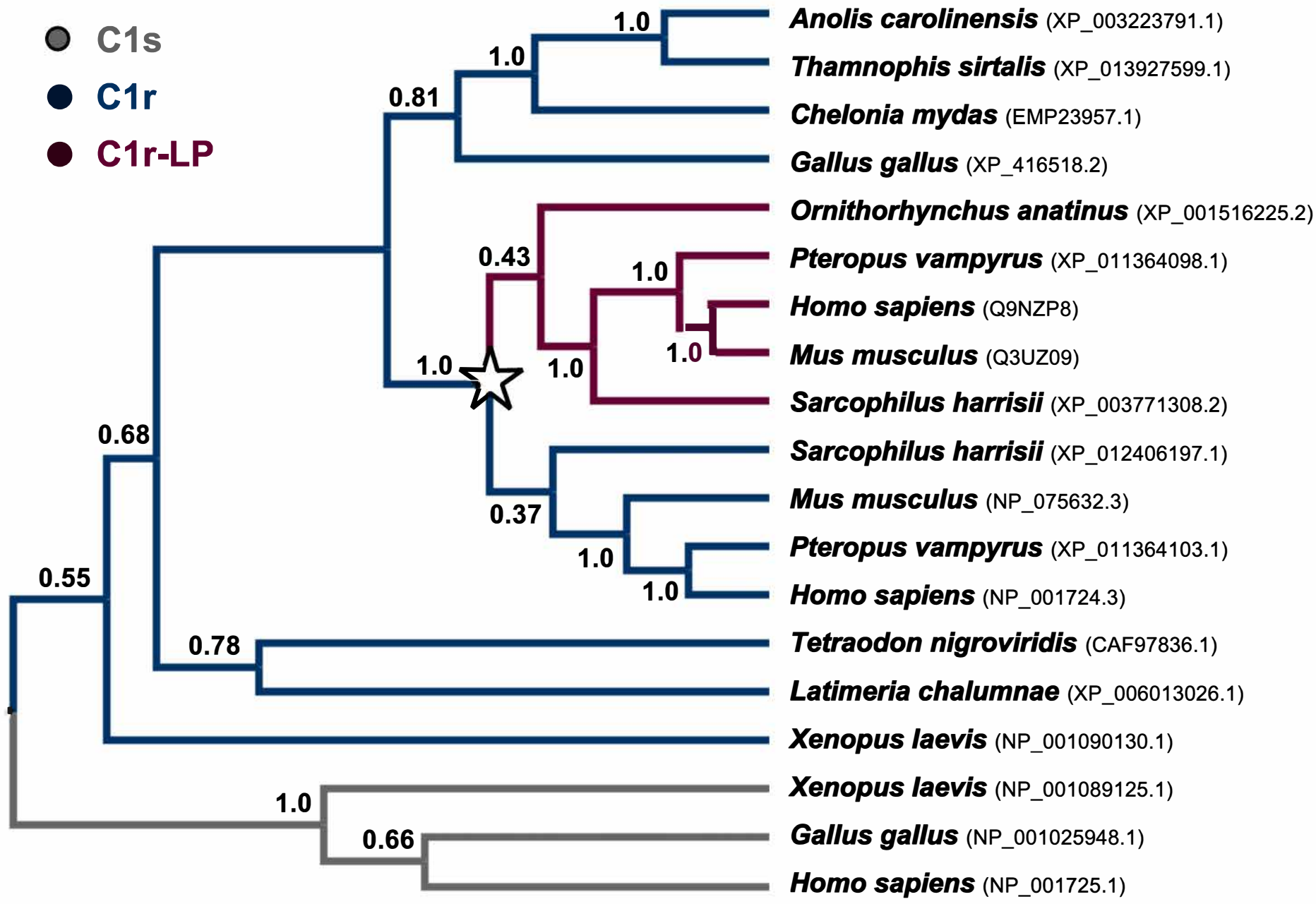

cartilaginous fishes
teleosts
amphibians reptiles
birds
mammals
Leer LD-DLHASHP HHVEKIFFEE IHDATNSSEY DNDIVLLKLS DSVSYGDHIV
Scca VR-EVDSSHE VHVEEVIYH- -HRVGDAVEY RNDLALVKLK ENVTFSNHIM
Gici AR-EITAAHQ VHVEDVHYH- -PRMRDAYVY RNDIALVKLK EDVHFSNHIM
Dare RA-DATASTE VAVEKVFLH- -PGFQNTSDW DNDLALIKLK EPVKFSKSIL
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.