

## Supplementary information

### Family-wide molecular adaptations to underground life in African mole-rats revealed by phylogenomic analysis

Kalina T. J. Davies, Nigel C. Bennett, Georgia Tsagkogeorga, Stephen J. Rossiter, Christopher G. Faulkes

## Supplementary tables S1–S9

(See separate .xls files)

## Supplementary table S10

**Taxonomic classification and data source for species, and datasets, included in this study.** The majority of mole-rat data were newly generated via RNA-seq and the raw-reads have been deposited in GenBank (see below for SRA IDs). These data were combined with coding sequences from the naked mole-rat (Keane et al. 2014) and outgroup species (Flicek et al. 2014).

Order	Family	Species	Common name	Assembly/SRA ID	
Rodentia	Bathyergidae	<i>Bathyergus suillus</i>	Cape dune mole-rat	SRR2141210	
		<i>Cryptomys hottentotus mahali</i>	Lesotho mole-rat	SRR2141211	
		<i>Cryptomys h. natalensis</i>	Natal mole-rat	SRR2141212	
		<i>Cryptomys h. pretoriae</i>	Highveld mole-rat	SRR2141213	
		<i>Fukomys damarensis</i>	Damaraland mole-rat	SRR2141214	
		<i>Georchus capensis</i>	Cape mole-rat	SRR2141216	
		<i>Heliophobius emini</i>	Silvery mole-rat	SRR2141215	
		<i>Heterocephalus glaber</i>	Naked mole rat	HetGla_female_1.0	
		Caviidae	<i>Cavia porcellus</i>	Guinea pig	cavPor3
		Muridae	<i>Mus musculus</i>	Mouse	GRCm38
<i>Rattus norvegicus</i>	Rat			Rnor_5.0	
Spalacidae	<i>Tachyoryctes splendens</i>		East African root rat	SRR2141217	
Heteromyidae	<i>Dipodomys ordii</i>		Kangaroo rat	dipOrd1	
Sciuridae	<i>Ictidomys tridecemlineatus</i>	Thirteen-lined ground squirrel	spetri2		
Lagomorpha	Ochotonidae	<i>Ochotona princeps</i>	Pika	pika	
	Leporidae	<i>Oryctolagus cuniculus</i>	Rabbit	OryCun2.0	
Primates	Hominidae	<i>Homo sapiens</i>	Human	GRCh37	
		<i>Pan troglodytes</i>	Chimpanzee	CHIMP2.1.4	

		<i>Pongo abelii</i>	Orangutan	PPYG2
	Hylobatidae	<i>Nomascus leucogenys</i>	Gibbon	Nleu1.0
	Galagidae	<i>Otolemur garnettii</i>	Bushbaby	OtoGar3

**Supplementary table S11**

**Sequencing and assembly statistics for the eight rodent transcriptomes produced by this study.** Abbreviations: GC % – proportion of guanine-cytosine.

<b>Species name</b>	<b>Read length (bp)</b>	<b># reads</b>	<b># nucleotides</b>	<b>Q20 %</b>	<b>GC %</b>	<b># transcripts</b>	<b>Average length</b>	<b># cut transcripts</b>	<b>Average cut length</b>
<i>Bathyergus suillus</i>	2 x 100	18,175,296	3,635,059,200	95.52	48.97	176,294	614	177,332	604
<i>Cryptomys h. mahali</i>	2 x 100	17,253,040	3,450,608,000	95.46	48.65	318,549	574	320,282	564
<i>Cryptomys h. natalensis</i>	2 x 100	17,290,951	3,458,190,200	95.63	47.37	259,886	494	260,393	488
<i>Cryptomys h. pretoriae</i>	2 x 100	17,833,375	3,566,675,000	95.38	49.48	226,694	567	227,740	559
<i>Fukomys damarensis</i>	2 x 100	18,220,998	3,644,199,600	95.74	48.85	244,256	602	245,682	589
<i>Georychus capensis</i>	2 x 90	20,038,942	3,607,009,560	97.37	46.09	228,786	416	229,151	414
<i>Heliophobius emini</i>	2 x 100	17,410,109	3,482,021,800	95.79	41.42	105,595	367	105,685	366
<i>Tachyoryctes splendens</i>	2 x 100	18,135,605	3,627,121,000	94.49	49.26	237,399	630	238,843	619

**Supplementary table S12**

**Ensembl gene IDs and gene symbols for the 64 candidate ‘social’ genes.** Each gene is associated with at least one of the following GO terms: ‘social behavior’ (GO:0035176); ‘grooming behavior’ (GO:0007625); ‘vocalization behavior’ (GO:0071625) and ‘regulation of grooming behavior’ (GO:2000821). The total number of species per alignments and the number of mole-rat species included are listed. Alignments containing <4 taxa were not studied further.

Ensembl Gene ID	Gene name	Description	#Species (#mole-rats)
ENSG00000138279	<i>ANXA7</i>	annexin A7	11 (8)
ENSG00000198931	<i>APRT</i>	adenine phosphoribosyltransferase	11 (8)
ENSG00000101200	<i>AVP</i>	arginine vasopressin	4 (2)
ENSG00000166148	<i>AVPR1A</i>	arginine vasopressin receptor 1A	9 (7)
ENSG00000140488	<i>CELF6</i>	CUGBP, Elav-like family member 6	4 (2)
ENSG00000160716	<i>CHRN2</i>	cholinergic receptor, nicotinic, beta 2 (neuronal)	4 (2)
ENSG00000182372	<i>CLN8</i>	ceroid-lipofuscinosis, neuronal 8 (epilepsy, progressive with mental retardation)	11 (8)
ENSG00000174469	<i>CNTNAP2</i>	contactin associated protein-like 2	8 (6)
ENSG00000152910	<i>CNTNAP4</i>	contactin associated protein-like 4	7 (5)
ENSG00000040531	<i>CTNS</i>	cystinosin, lysosomal cystine transporter	9 (6)
ENSG00000123454	<i>DBH</i>	dopamine beta-hydroxylase (dopamine beta-monooxygenase)	5 (2)
ENSG00000203797	<i>DDO</i>	D-aspartate oxidase	10 (7)
ENSG00000132535	<i>DLG4</i>	discs, large homolog 4 (Drosophila)	10 (7)
ENSG00000213551	<i>DNAJC9</i>	DnaJ (Hsp40) homolog, subfamily C, member 9	11 (8)
ENSG00000184845	<i>DRD1</i>	dopamine receptor D1	7 (4)
ENSG00000149295	<i>DRD2</i>	dopamine receptor D2	7 (4)
ENSG00000151577	<i>DRD3</i>	dopamine receptor D3	5 (2)
ENSG00000069696	<i>DRD4</i>	dopamine receptor D4	5 (2)
ENSG00000107404	<i>DVLI</i>	dishevelled, dsh homolog 1 (Drosophila)	9 (6)
ENSG00000185838	<i>GNBIL</i>	guanine nucleotide binding protein (G protein), beta polypeptide 1-like	9 (6)
ENSG00000182771	<i>GRID1</i>	glutamate receptor, ionotropic, delta 1	8 (5)
ENSG00000176884	<i>GRIN1</i>	glutamate receptor, ionotropic, N-methyl D-aspartate 1	10 (7)
ENSG00000126010	<i>GRPR</i>	gastrin-releasing peptide receptor	6 (3)
ENSG00000120068	<i>HOXB8</i>	homeobox B8	<4
ENSG00000165704	<i>HPRT1</i>	hypoxanthine phosphoribosyltransferase 1	11 (8)
ENSG00000174775	<i>HRAS</i>	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	10 (7)
ENSG00000197386	<i>HTT</i>	huntingtin	11 (8)
ENSG00000125538	<i>IL1B</i>	interleukin 1, beta	10 (7)
ENSG00000268944	<i>IRAK1</i>	Interleukin-1 receptor-associated kinase 1	5 (2)
ENSG00000133703	<i>KRAS</i>	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	6 (3)

ENSG00000008735	<i>MAPK8IP2</i>	mitogen-activated protein kinase 8 interacting protein 2	9 (6)
ENSG00000166603	<i>MC4R</i>	melanocortin 4 receptor	7 (4)
ENSG00000268563	<i>MECP2</i>	methyl CpG binding protein 2 (Rett syndrome)	5 (2)
ENSG00000125863	<i>MKKS</i>	McKusick-Kaufman syndrome	11 (8)
ENSG00000166343	<i>MSS51</i>	MSS51 mitochondrial translational activator	8 (5)
ENSG00000105357	<i>MYH14</i>	myosin, heavy chain 14, non-muscle	6 (3)
ENSG00000169992	<i>NLGN2</i>	neuroligin 2	9 (7)
ENSG00000196338	<i>NLGN3</i>	neuroligin 3	7 (4)
ENSG00000146938	<i>NLGN4X</i>	neuroligin 4, X-linked	<4
ENSG00000279292	<i>NLGN4Y</i>	neuroligin 4, Y-linked	0
ENSG00000132911	<i>NMUR2</i>	neuromedin U receptor 2	4 (2)
ENSG00000112333	<i>NR2E1</i>	nuclear receptor subfamily 2, group E, member 1	8 (5)
ENSG00000179915	<i>NRXN1</i>	neurexin 1	8 (7)
ENSG00000110076	<i>NRXN2</i>	neurexin 2	9 (6)
ENSG00000021645	<i>NRXN3</i>	neurexin 3	6 (3)
ENSG00000101405	<i>OXT</i>	oxytocin/neurophysin I prepropeptide	<4
ENSG00000180914	<i>OXTR</i>	oxytocin receptor	6 (3)
ENSG00000141744	<i>PNMT</i>	phenylethanolamine N-methyltransferase	5 (2)
ENSG00000107758	<i>PPP3CB</i>	protein phosphatase 3, catalytic subunit, beta isozyme	7 (4)
ENSG00000131238	<i>PPT1</i>	palmitoyl-protein thioesterase 1	11 (8)
ENSG00000171862	<i>PTEN</i>	phosphatase and tensin homolog	11 (8)
ENSG00000188710	<i>QRFP</i>	pyroglutamylated RFamide peptide	4 (2)
ENSG00000189056	<i>RELN</i>	reelin	4 (1)
ENSG00000161681	<i>SHANK1</i>	SH3 and multiple ankyrin repeat domains 1	7 (6)
ENSG00000162105	<i>SHANK2</i>	SH3 and multiple ankyrin repeat domains 2	<4
ENSG00000251322	<i>SHANK3</i>	SH3 and multiple ankyrin repeat domains 3	6 (5)
ENSG00000108576	<i>SLC6A4</i>	solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	8 (5)
ENSG00000165300	<i>SLITRK5</i>	SLIT and NTRK-like family, member 5	8 (5)
ENSG00000102359	<i>SRPX2</i>	sushi-repeat containing protein, X-linked 2	7 (4)
ENSG00000135569	<i>TAAR5</i>	trace amine associated receptor 5	4 (2)
ENSG00000184058	<i>TBX1</i>	T-box 1	9 (6)
ENSG00000180176	<i>TH</i>	tyrosine hydroxylase	6 (3)
ENSG00000163794	<i>UCN</i>	urocortin	<4
ENSG00000197969	<i>VPS13A</i>	vacuolar protein sorting 13 homolog A (S. cerevisiae)	8 (6)

## Supplementary figure legends

### Supplementary figure S1

**Geographical distribution and divergence time between the mole-rat species included in this study.** (a) Map showing the approximate ranges of bathyergid species in this study [adapted from (Bennett and Faulkes 2000; Bennett and Faulkes unpublished data)]. (b) Simplified phylogeny for the Bathyergidae indicating the six main clades/genera, together with the closest extant outgroup, the cane rat *Thryonomys swinderianus*, and based on mitochondrial 12S rRNA and *cyt-b* sequence data. Numbers on internal nodes and scale bar represent divergence times in millions of years ago (Myr) estimated using a molecular clock approach, and using the Bathyergid fossil *Proheliophobius* for calibration of genetic distances. Numbers in parentheses indicate current estimates of species numbers in each genus [data and figure adapted from (Faulkes et al. 1997; Faulkes et al. 2004; Ingram et al. 2004; Faulkes et al. 2010; Faulkes et al. 2011; Faulkes and Bennett 2013)].

### Supplementary figure S2

**Phylogenetic relationships between the Bathyergidae species under study, based on:** (a) Maximum likelihood analysis of a partitioned dataset of 3,999 concatenated genes (1,754,367 codons), all nodes were recovered with 100% bootstrap support, and branch lengths correspond to the substitution rate. (b) Bayesian analysis of the 172 most informative gene partitions (60,783 amino acids) subsampled from the above dataset, all nodes unless otherwise shown, were recovered with a Bayesian posterior probability of 1.00.

### Supplementary figure S3

**Protein-protein interaction network between products encoded by genes identified as being under positive selection in mole-rats by branch-site models.** Node size is scaled proportionally to the number of connections, and connection thickness corresponds to the confidence of the interaction, with a thicker line indicating stronger evidence. Evidence is based on combined interaction scores of the following lines of evidence: neighborhood, fusion, co-occurrence, homology, co-expression, experimental knowledge and text mining.

### Supplementary figure S4

**Distribution of Bathyergidae sociality across the phylogeny:** the documented social states (Faulkes and Bennett 2013) are mapped onto tips (shaded circles). Inferred hypothetical evolutionary scenarios explaining the distribution of social structure across the family are shown along branches (shaded squares). The different forms of sociality are represented by the following colors: white – solitary; light grey – social + cooperative breeding; dark grey – social and black – eusocial.

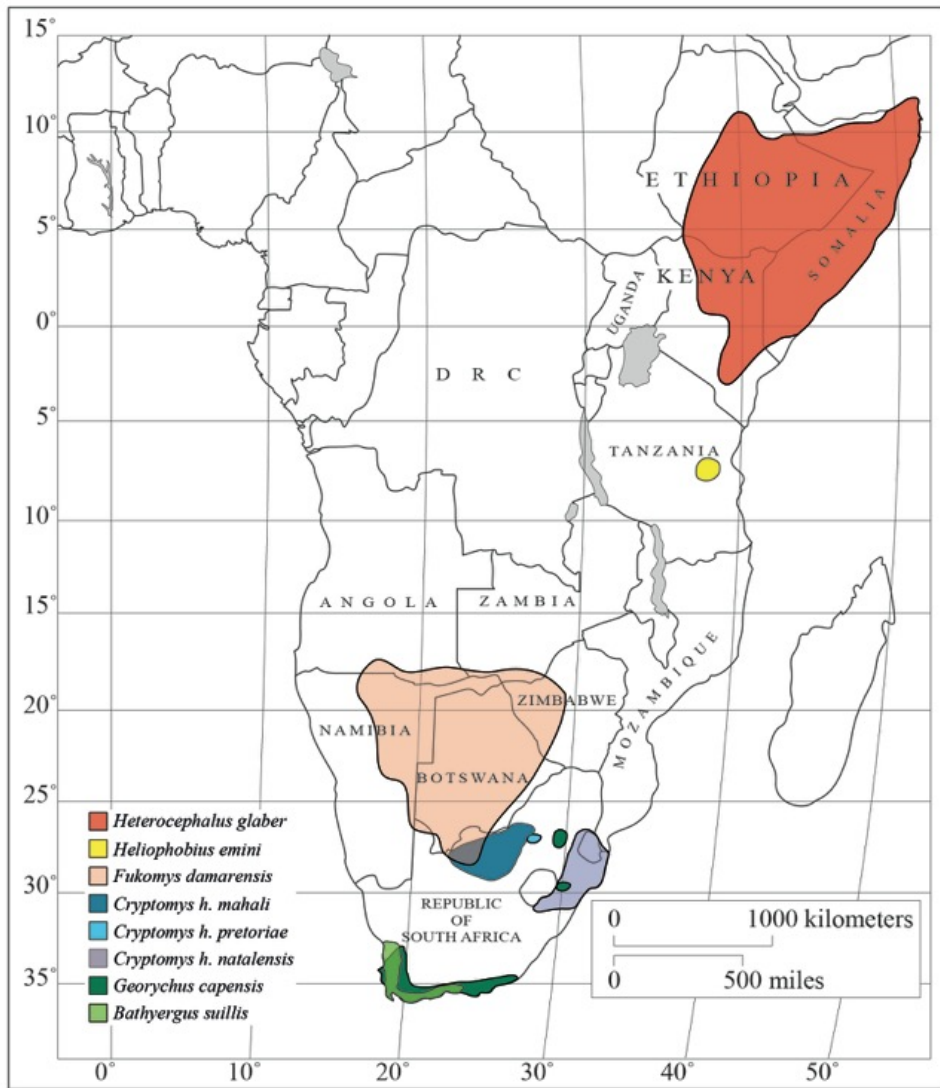
### Supplementary figure S5

**Candidate ‘social’ genes found to be under positive selection in mole-rats and the guinea pig, using the Branch-site REL model.** Documented social states (Faulkes and Bennett 2013) are mapped onto tips (shaded circles). The different forms of sociality are represented by the following colors: white – solitary; light grey – social + cooperative breeding; dark grey – social and black – eusocial.

**Supplementary references:**

- Bennett NC, Faulkes CG. 2000. African mole-rats: ecology and eusociality: Cambridge University Press.**
- Bennett NC, Faulkes CG. unpublished data.**
- Faulkes CG, Bennett NC. 2013. Plasticity and constraints on social evolution in African mole-rats: ultimate and proximate factors. *Phil. Trans. R. Soc. B* 368:20120347.**
- Faulkes CG, Bennett NC, Bruford MW, O'Brien HP, Aguilar GH, et al. 1997. Ecological constraints drive social evolution in the African mole-rats. *Proc. R. Soc. Biol. Sci. B*. 264:1619–1627.**
- Faulkes CG, Bennett NC, Cotterill FPD, Stanley W, Mgone GF, et al. 2011. Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat, genus *Heliophobius* (Family: Bathyergidae). *J. Zool.* 285:324–338.**
- Faulkes CG, Mgone GF, Le Comber SC, Bennett NC. 2010. Cladogenesis and endemism in Tanzanian mole-rats, genus *Fukomys* (Family: Bathyergidae): A role for tectonics? *Biol. J. Linn. Soc.* 100:337–352.**
- Faulkes CG, Verheyen E, Verheyen W, Jarvis JUM, Bennett NC. 2004. Phylogeographical patterns of genetic divergence and speciation in African mole-rats (Family: Bathyergidae). *Mol. Ecol.* 13:613–629.**
- Flicek P, Amode MR, Barrell D, Beal K, Billis K, et al. 2014. Ensembl 2014. *Nucleic Acids Res.* 42:D749–D755.**
- Ingram CM, Burda H, Honeycutt RL. 2004. Molecular phylogenetics and taxonomy of the African mole-rats, genus *Cryptomys* and the new genus *Coetomys* Gray, 1864. *Mol. Phylogenet. Evol.* 31:997–1014.**
- Keane M, Craig T, Alföldi J, Berlin AM, Johnson J, et al. 2014. The Naked Mole Rat Genome Resource: facilitating analyses of cancer and longevity-related adaptations. *Bioinformatics* 30:3558–3560.**

Figure S1  
(a)



(b)

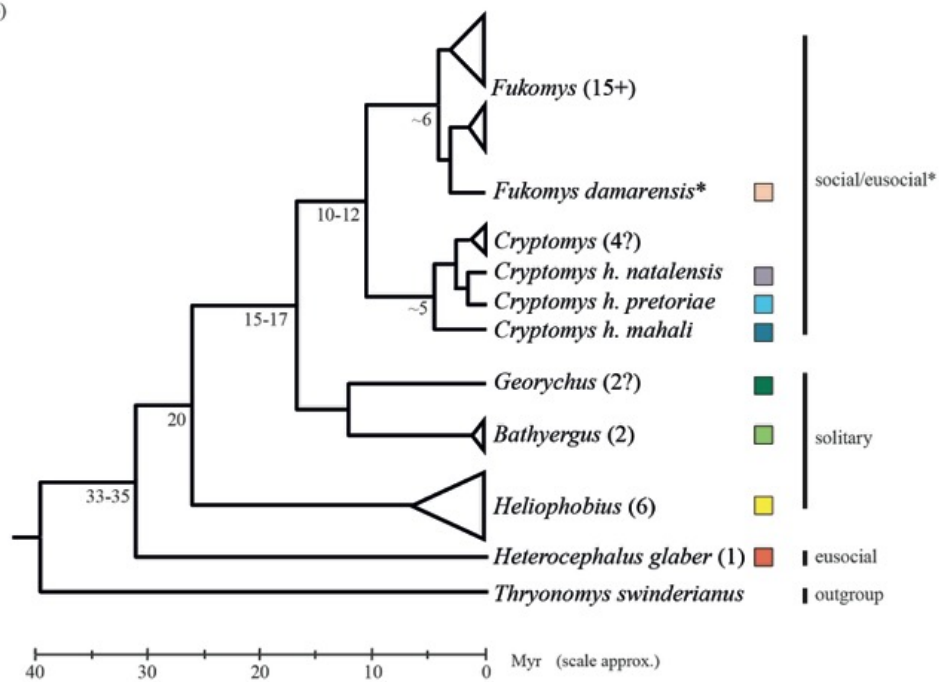




Figure S2a

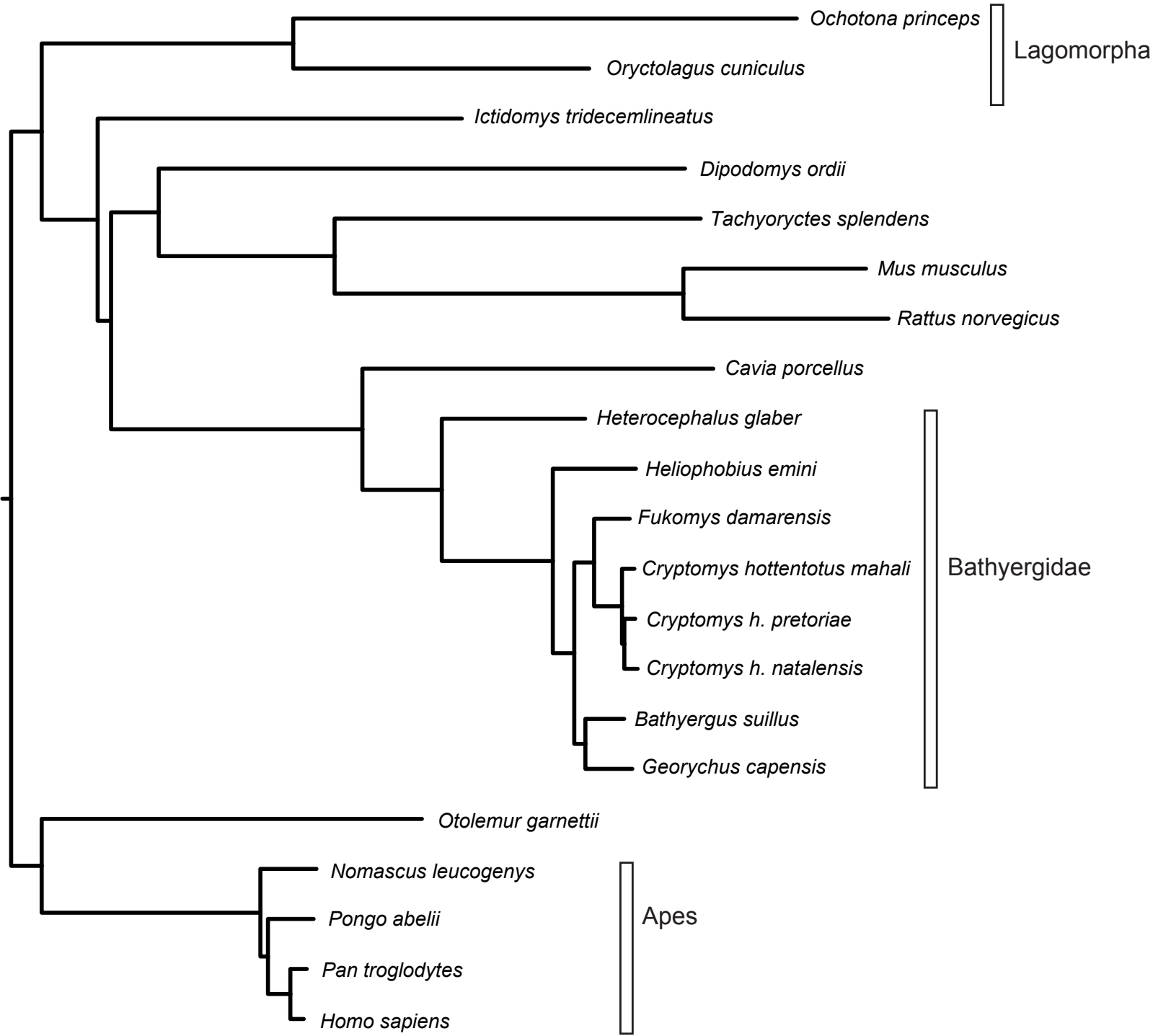


Figure S2b

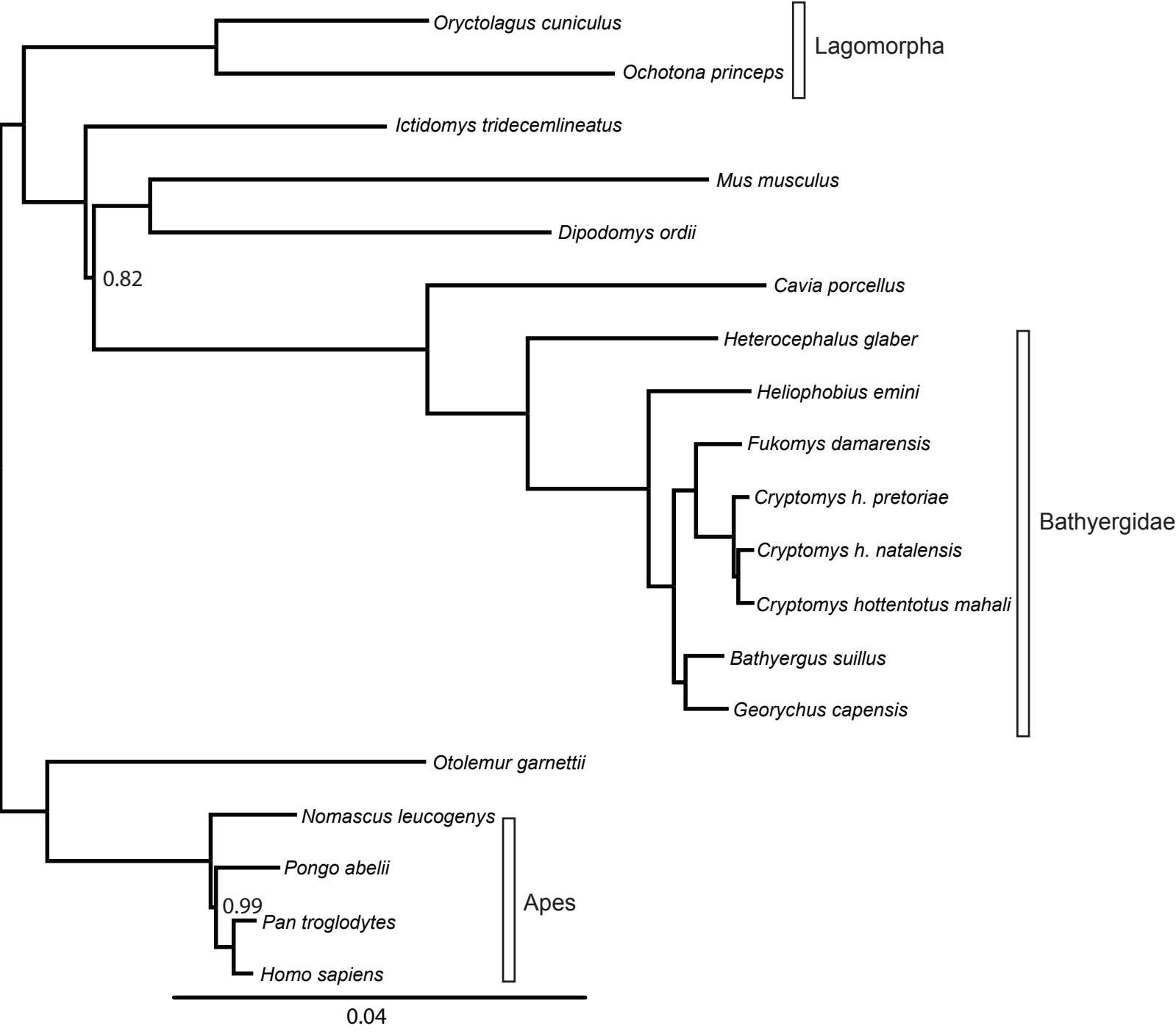


Figure S3

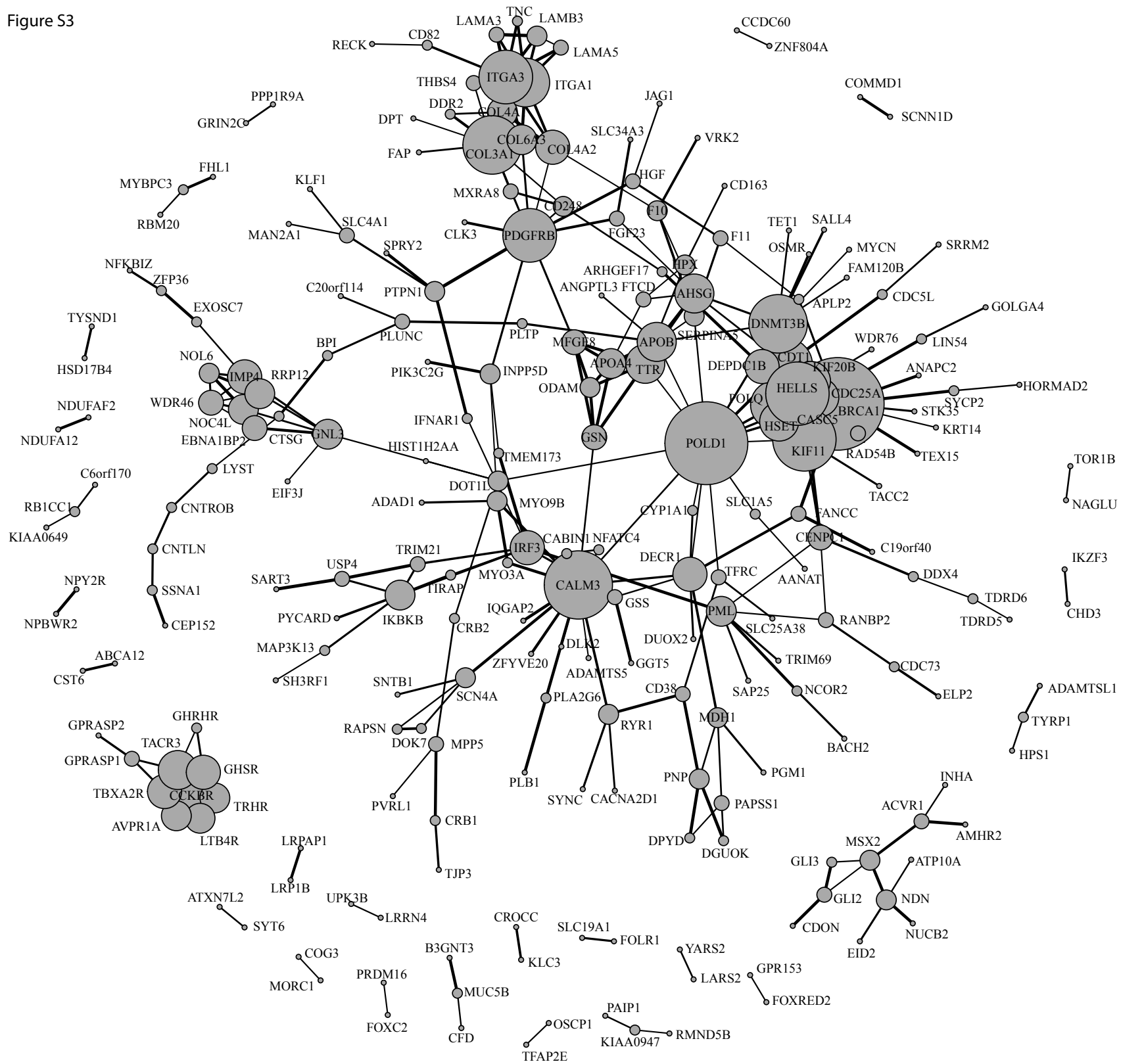


Figure S4

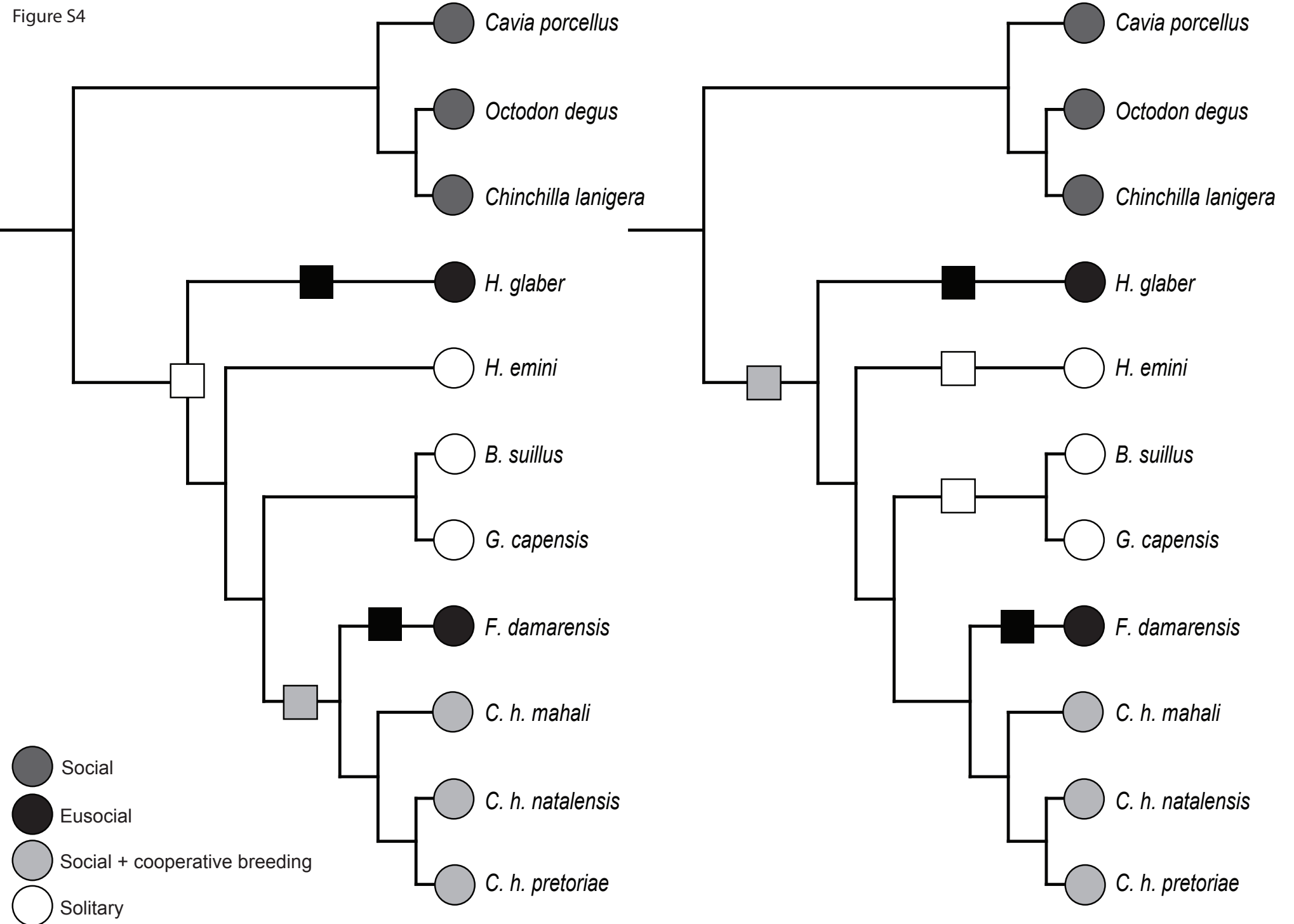


Figure S5

