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

| | | | | |
|---|---|---|---|---|
| | | *Pongo abelii* | Orangutan | PPYG2 |
| | Hylobatidae | *Nomascus leucogenys* | Gibbon | Nleu1.0 |
| | Galagidae | *Otolemur garnettii* | Bushbaby | OtoGar3 |

## Supplementary table S11

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

| **Species name** | **Read length (bp)** | **# reads** | **# nucleotides** | **Q20 %** | **GC %** | **# transcripts** | **Average length** | **# cut transcripts** | **Average cut length** |
|---|---|---|---|---|---|---|---|---|---|
| *Bathyergus suillus* | 2 x 100 | 18,175,296 | 3,635,059,200 | 95.52 | 48.97 | 176,294 | 614 | 177,332 | 604 |
| *Cryptomys h. mahali* | 2 x 100 | 17,253,040 | 3,450,608,000 | 95.46 | 48.65 | 318,549 | 574 | 320,282 | 564 |
| *Cryptomys h. natalensis* | 2 x 100 | 17,290,951 | 3,458,190,200 | 95.63 | 47.37 | 259,886 | 494 | 260,393 | 488 |
| *Cryptomys h. pretoriae* | 2 x 100 | 17,833,375 | 3,566,675,000 | 95.38 | 49.48 | 226,694 | 567 | 227,740 | 559 |
| *Fukomys damarensis* | 2 x 100 | 18,220,998 | 3,644,199,600 | 95.74 | 48.85 | 244,256 | 602 | 245,682 | 589 |
| *Georychus capensis* | 2 x 90 | 20,038,942 | 3,607,009,560 | 97.37 | 46.09 | 228,786 | 416 | 229,151 | 414 |
| *Heliophobius emini* | 2 x 100 | 17,410,109 | 3,482,021,800 | 95.79 | 41.42 | 105,595 | 367 | 105,685 | 366 |
| *Tachyoryctes splendens* | 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 | *ANXA7* | annexin A7 | 11 (8) |
| ENSG00000198931 | *APRT* | adenine phosphoribosyltransferase | 11 (8) |
| ENSG00000101200 | *AVP* | arginine vasopressin | 4 (2) |
| ENSG00000166148 | *AVPR1A* | arginine vasopressin receptor 1A | 9 (7) |
| ENSG00000140488 | *CELF6* | CUGBP, Elav-like family member 6 | 4 (2) |
| ENSG00000160716 | *CHRNB2* | cholinergic receptor, nicotinic, beta 2 (neuronal) | 4 (2) |
| ENSG00000182372 | *CLN8* | ceroid-lipofuscinosis, neuronal 8 (epilepsy, progressive with mental retardation) | 11 (8) |
| ENSG00000174469 | *CNTNAP2* | contactin associated protein-like 2 | 8 (6) |
| ENSG00000152910 | *CNTNAP4* | contactin associated protein-like 4 | 7 (5) |
| ENSG00000040531 | *CTNS* | cystinosin, lysosomal cystine transporter | 9 (6) |
| ENSG00000123454 | *DBH* | dopamine beta-hydroxylase (dopamine beta-monooxygenase) | 5 (2) |
| ENSG00000203797 | *DDO* | D-aspartate oxidase | 10 (7) |
| ENSG00000132535 | *DLG4* | discs, large homolog 4 (Drosophila) | 10 (7) |
| ENSG00000213551 | *DNAJC9* | DnaJ (Hsp40) homolog, subfamily C, member 9 | 11 (8) |
| ENSG00000184845 | *DRD1* | dopamine receptor D1 | 7 (4) |
| ENSG00000149295 | *DRD2* | dopamine receptor D2 | 7 (4) |
| ENSG00000151577 | *DRD3* | dopamine receptor D3 | 5 (2) |
| ENSG00000069696 | *DRD4* | dopamine receptor D4 | 5 (2) |
| ENSG00000107404 | *DVL1* | dishevelled, dsh homolog 1 (Drosophila) | 9 (6) |
| ENSG00000185838 | *GNB1L* | guanine nucleotide binding protein (G protein), beta polypeptide 1-like | 9 (6) |
| ENSG00000182771 | *GRID1* | glutamate receptor, ionotropic, delta 1 | 8 (5) |
| ENSG00000176884 | *GRIN1* | glutamate receptor, ionotropic, N-methyl D-aspartate 1 | 10 (7) |
| ENSG00000126010 | *GRPR* | gastrin-releasing peptide receptor | 6 (3) |
| ENSG00000120068 | *HOXB8* | homeobox B8 | <4 |
| ENSG00000165704 | *HPRT1* | hypoxanthine phosphoribosyltransferase 1 | 11 (8) |
| ENSG00000174775 | *HRAS* | v-Ha-ras Harvey rat sarcoma viral oncogene homolog | 10 (7) |
| ENSG00000197386 | *HTT* | huntingtin | 11 (8) |
| ENSG00000125538 | *IL1B* | interleukin 1, beta | 10 (7) |
| ENSG00000268944 | *IRAK1* | Interleukin-1 receptor-associated kinase 1 | 5 (2) |
| ENSG00000133703 | *KRAS* | v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog | 6 (3) |

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

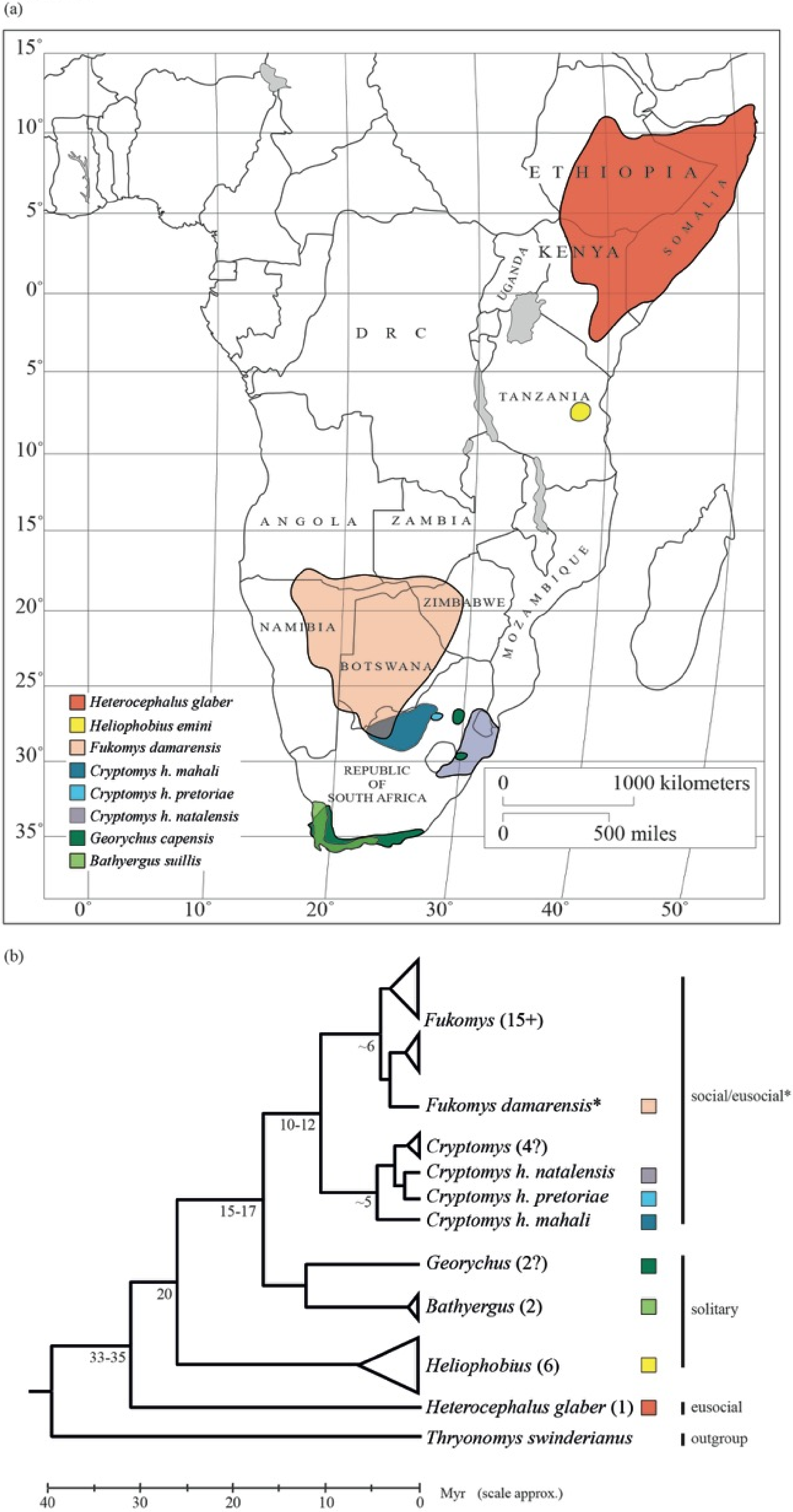

(a)
15°
10°
5°
0°
5°
10°
15°
20°
25°
30°
35°
ETHIOPIA
SOMALIA
KENYA
UGANDA
DRC
TANZANIA
ANGOLA
ZAMBIA
MOZAMBIQUE
ZIMBABWE
NAMIBIA
BOTSWANA
REPUBLIC OF SOUTH AFRICA
Heterocephalus glaber
Heliophobius emini
Fukomys damarensis
Cryptomys h. mahali
Cryptomys h. pretoriae
Cryptomys h. natalensis
Georychus capensis
Bathyergus suillis
0 1000 kilometers
0 500 miles
0°
10°
20°
30°
40°
50°
(b)
Fukomys (15+)
~6
Fukomys damarensis*
10-12
Cryptomys (4?)
Cryptomys h. natalensis
Cryptomys h. pretoriae
~5
Cryptomys h. mahali
15-17
Georychus (2?)
Bathyergus (2)
20
Heliophobius (6)
33-35
Heterocephalus glaber (1)
Thryonomys swinderianus
social/eusocial*
solitary
eusocial
outgroup
40
30
20
10
0
Myr (scale approx.)

Figure S2a

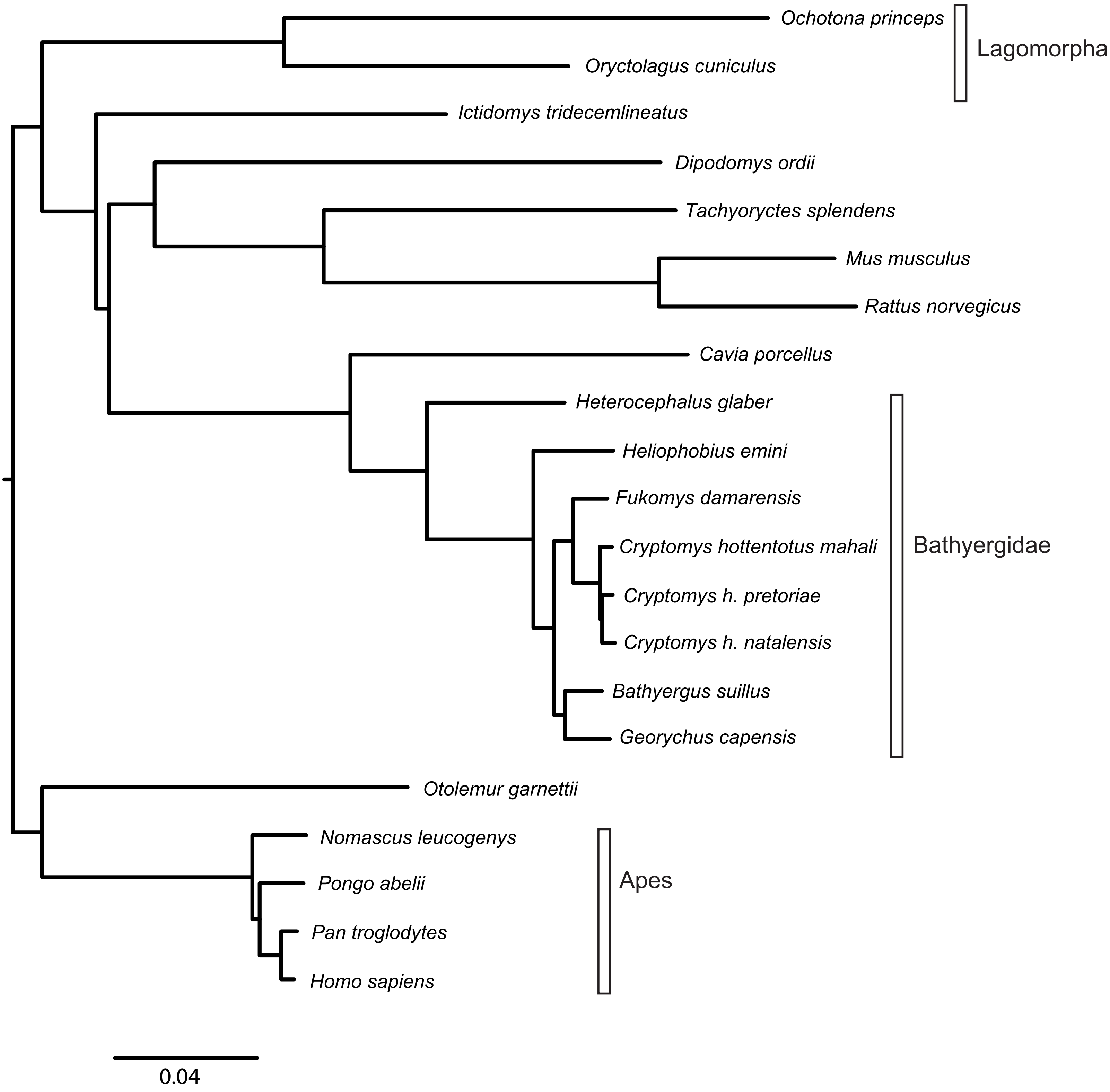

Ochotona princeps
Oryctolagus cuniculus
Lagomorpha
Ictidomys tridecemlineatus
Dipodomys ordii
Tachyoryctes splendens
Mus musculus
Rattus norvegicus
Cavia porcellus
Heterocephalus glaber
Heliophobius emini
Fukomys damarensis
Cryptomys hottentotus mahali
Cryptomys h. pretoriae
Cryptomys h. natalensis
Bathyergus suillus
Georychus capensis
Bathyergidae
Otolemur garnettii
Nomascus leucogenys
Pongo abelii
Pan troglodytes
Homo sapiens
Apes
0.04

Figure S2b

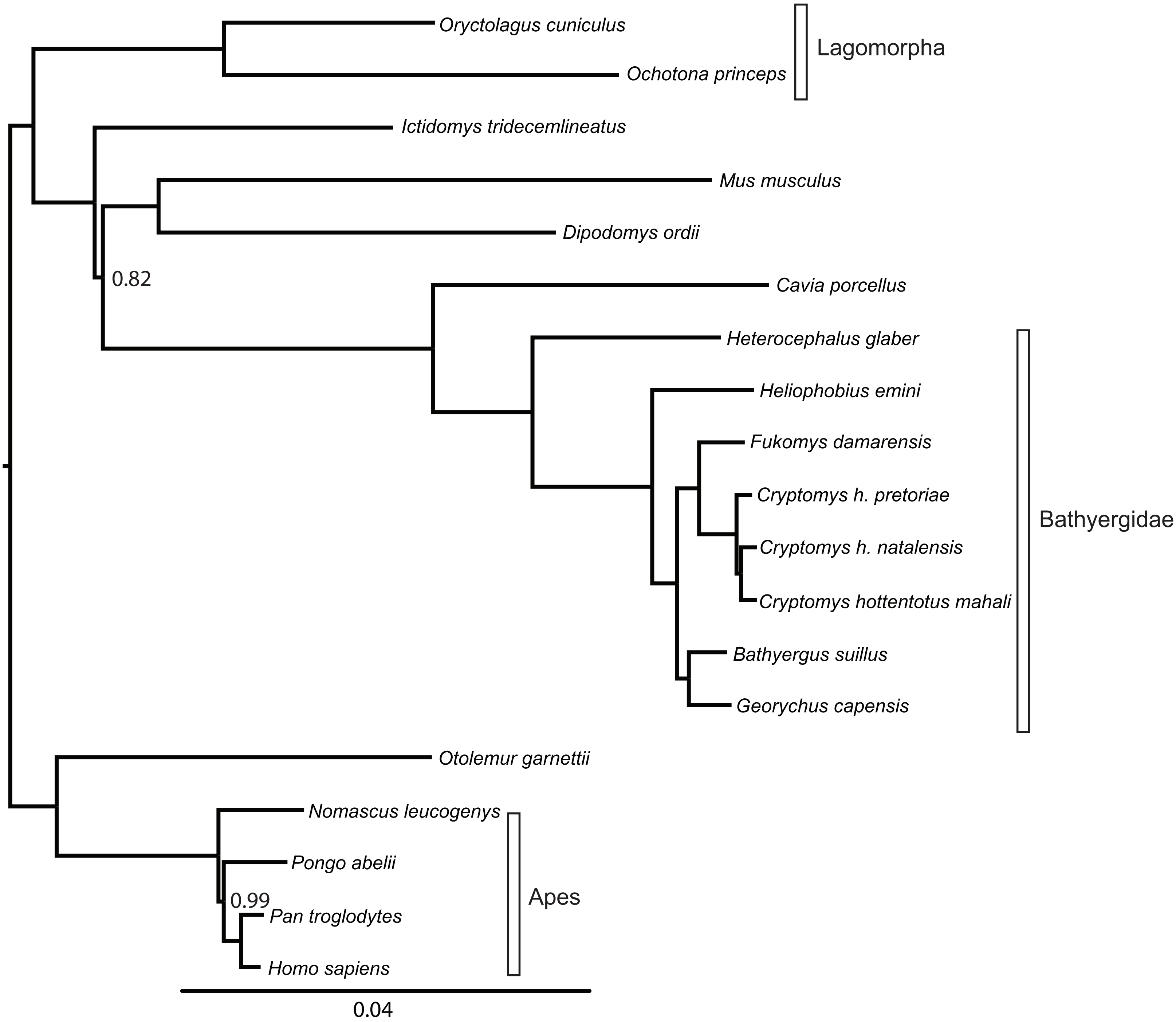
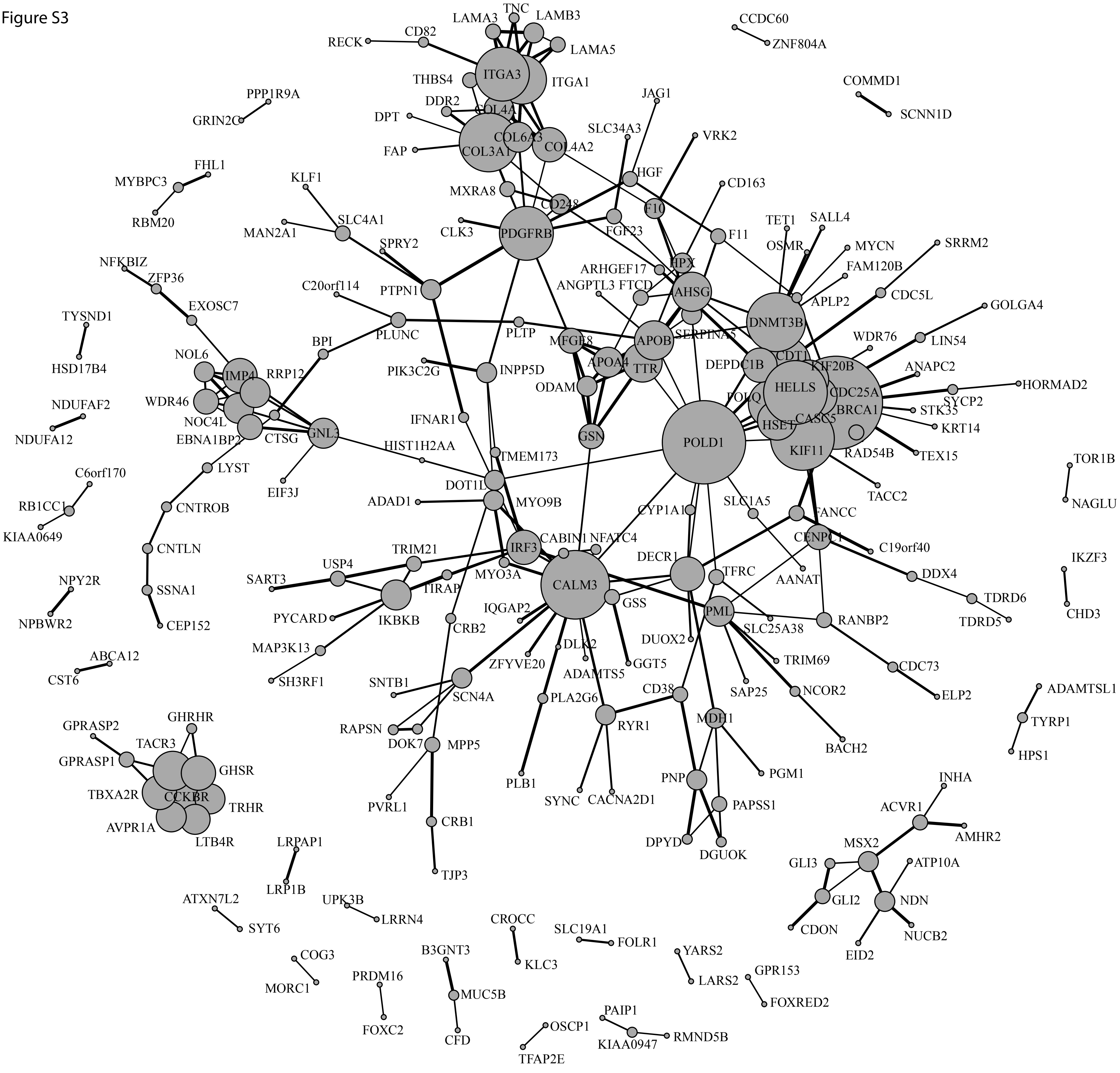

Oryctolagus cuniculus
Ochotona princeps
Lagomorpha
Ictidomys tridecemlineatus
Mus musculus
Dipodomys ordii
0.82
Cavia porcellus
Heterocephalus glaber
Heliophobius emini
Fukomys damarensis
Cryptomys h. pretoriae
Cryptomys h. natalensis
Cryptomys hottentotus mahali
Bathyergus suillus
Georychus capensis
Bathyergidae
Otolemur garnettii
Nomascus leucogenys
Pongo abelii
0.99
Pan troglodytes
Homo sapiens
Apes
0.04

Figure S3



TNC
LAMA3
LAMB3
CD82
RECK
LAMA5
CCDC60
ZNF804A
ITGA3
ITGA1
THBS4
PPP1R9A
GRIN2C
COMMD1
SCNN1D
DDR2
COL4A
JAG1
DPT
SLC34A3
VRK2
COL6A3
COL4A2
FAP
COL3A1
FHL1
MYBPC3
RBM20
HGF
CD163
KLF1
MXRA8
CD248
F10
MAN2A1
SLC4A1
CLK3
PDGFRB
FGF23
F11
TET1
SALL4
SPRY2
OSMR
MYCN
SRRM2
NFKBIZ
HPX
ARHGEF17
FAM120B
ZFP36
C20orf114
ANGPTL3
FTCD
AHSG
PTPN1
APLP2
CDC5L
EXOSC7
GOLGA4
TYSND1
DNMT3B
BPI
PLUNC
PLTP
APOB
SERPINA5
HSD17B4
NOL6
MFGE8
CDT1
WDR76
LIN54
IMP4
RRP12
KIF20B
PIK3C2G
INPP5D
APOA4
DEPDC1B
ANAPC2
WDR46
TTR
ODAM
HELLS
CDC25A
SYCP2
HORMAD2
NDUFAF2
NOC4L
POLQ
BRCA1
STK35
NDUFA12
EBNA1BP2
CTSG
GNL3
IFNAR1
GSN
HSET
CASC5
KRT14
POLD1
HIST1H2AA
TMEM173
KIF11
RAD54B
TEX15
TOR1B
C6orf170
LYST
EIF3J
NAGLU
DOT1L
RB1CC1
ADAD1
MYO9B
TACC2
CNTROB
KIAA0649
CYP1A1
SLC1A5
FANCC
CABIN1
NFATC4
CENPC1
C19orf40
CNTLN
IRF3
TRIM21
DECR1
IKZF3
USP4
DDX4
MYO3A
TFRC
CHD3
NPY2R
SART3
TIRAP
CALM3
AANAT
TDRD6
SSNA1
GSS
NPBWR2
PYCARD
IKBKB
IQGAP2
PML
RANBP2
TDRD5
CEP152
CRB2
DUOX2
SLC25A38
MAP3K13
DLK2
ZFYVE20
GGT5
TRIM69
CDC73
ABCA12
ADAMTS5
CST6
SH3RF1
NCOR2
SNTB1
SCN4A
ELP2
CD38
SAP25
ADAMTSL1
PLA2G6
MDH1
TYRP1
GHRHR
RAPSN
RYR1
BACH2
GPRASP2
DOK7
MPP5
HPS1
TACR3
GPRASP1
GHSR
PGM1
CCKBR
TBXA2R
PNP
INHA
TRHR
PLB1
SYNC
CACNA2D1
PVRL1
PAPSS1
ACVR1
AVPR1A
LTB4R
CRB1
AMHR2
DPYD
DGUOK
LRPAP1
MSX2
GLI3
ATP10A
TJP3
NDN
LRP1B
ATXN7L2
UPK3B
GLI2
LRRN4
CROCC
SLC19A1
CDON
SYT6
FOLR1
NUCB2
EID2
YARS2
COG3
B3GNT3
KLC3
GPR153
LARS2
MORC1
PRDM16
MUC5B
FOXC2
FOXRED2
PAIP1
OSCP1
RMND5B
CFD
KIAA0947
TFAP2E

Figure S4

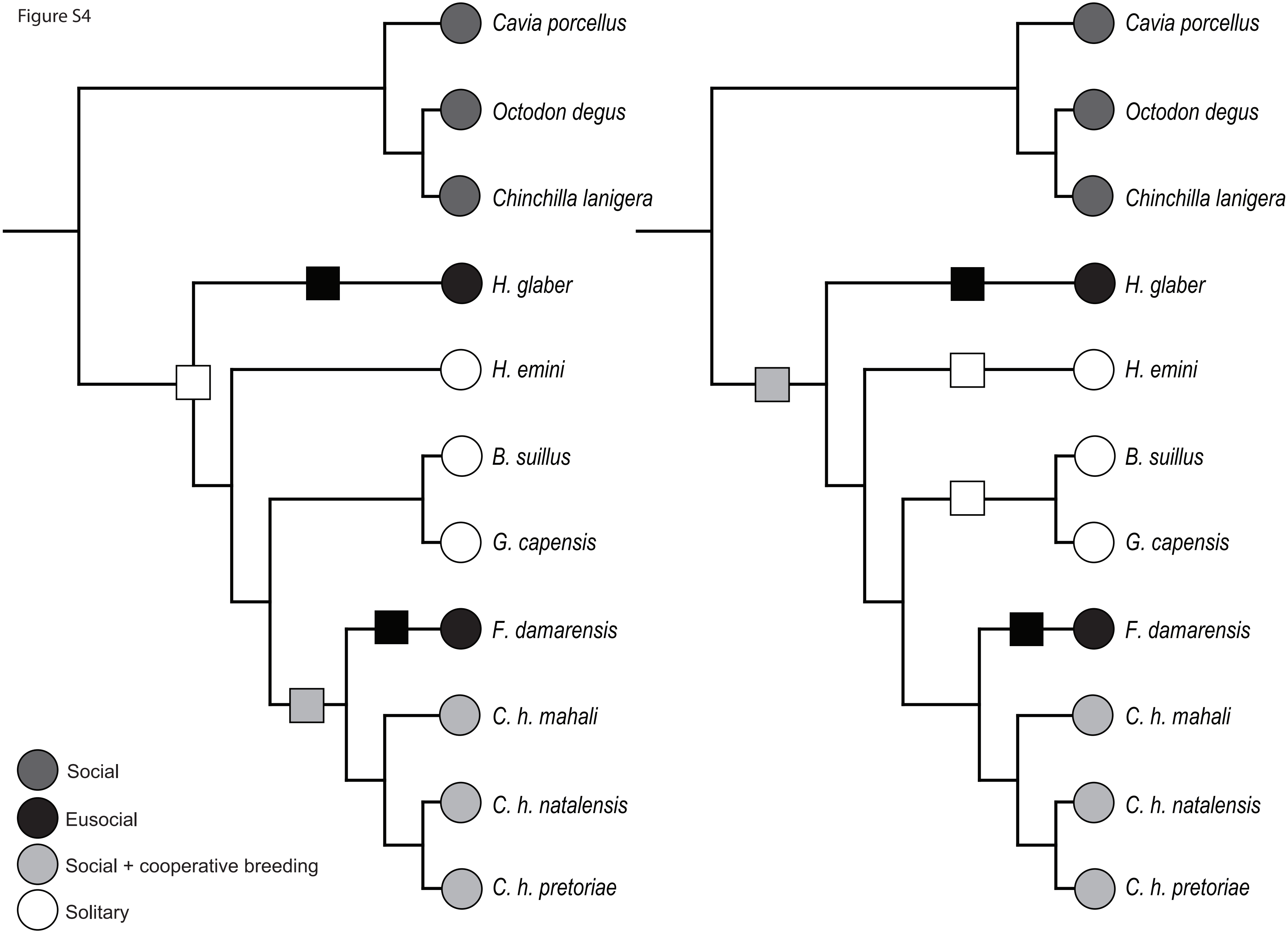
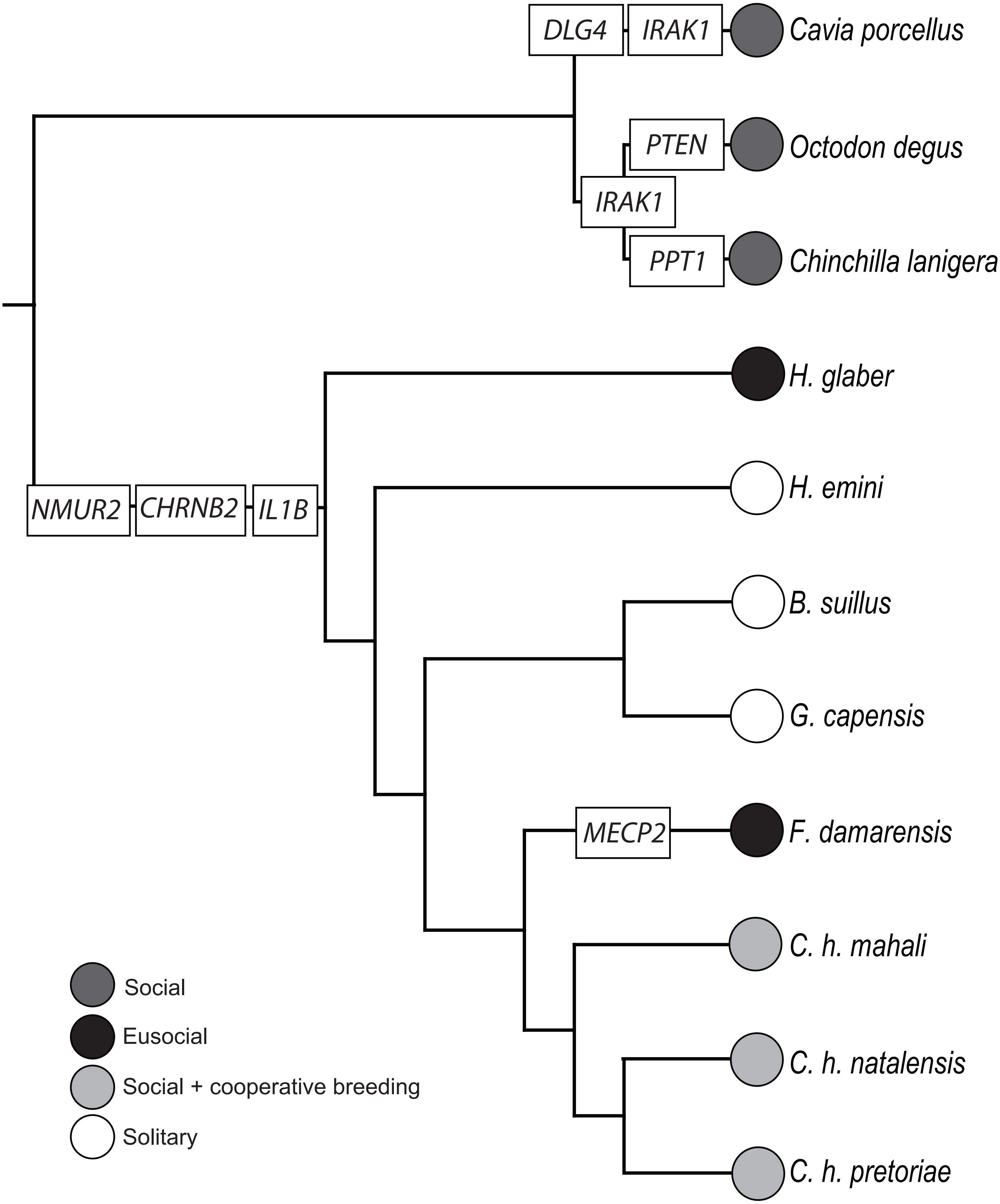

Cavia porcellus
Octodon degus
Chinchilla lanigera
H. glaber
H. emini
B. suillus
G. capensis
F. damarensis
C. h. mahali
C. h. natalensis
C. h. pretoriae
Social
Eusocial
Social + cooperative breeding
Solitary
Cavia porcellus
Octodon degus
Chinchilla lanigera
H. glaber
H. emini
B. suillus
G. capensis
F. damarensis
C. h. mahali
C. h. natalensis
C. h. pretoriae

Figure S5



DLG4
IRAK1
Cavia porcellus
PTEN
Octodon degus
IRAK1
PPT1
Chinchilla lanigera
H. glaber
NMUR2
CHRNB2
IL1B
H. emini
B. suillus
G. capensis
MECP2
F. damarensis
C. h. mahali
C. h. natalensis
C. h. pretoriae
Social
Eusocial
Social + cooperative breeding
Solitary