

# **Molecular evolution of the hyaluronan synthase 2 gene in mammals: implications for adaptations to the subterranean niche and cancer resistance**

Christopher G. Faulkes<sup>1\*</sup>, Kalina T.J. Davies<sup>1</sup>, Stephen J. Rossiter<sup>1</sup>, Nigel C. Bennett<sup>2</sup>

<sup>1</sup> School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS

<sup>2</sup> Department of Zoology & Entomology, University of Pretoria

\* Author for correspondence (c.g.faulkes@qmul.ac.uk).

Short title: adaptive evolution in the hyaluronan synthase 2 (HAS2) gene

## **Abstract**

The naked mole-rat (NMR) *Heterocephalus glaber* is a unique and fascinating mammal exhibiting many unusual adaptations to a subterranean lifestyle. The recent discovery of their resistance to cancer and exceptional longevity has opened up new and important avenues of research. Part of this resistance to cancer has been attributed to the fact that NMRs produce a modified form of hyaluronan—a key constituent of the extracellular matrix—that is thought to confer increased elasticity of the skin as an adaptation for living in narrow tunnels. This so-called high molecular mass hyaluronan (HMM-HA) stems from two apparently unique substitutions in the hyaluronan synthase 2 enzyme (HAS2). To test whether other subterranean mammals with similar selection pressures also show molecular adaptation in their *HAS2* gene, we sequenced the *HAS2* gene for 11 subterranean mammals and closely related species, and combined these with data from 57

other mammals. Comparative screening revealed that one of the two putatively important HAS2 substitutions in the NMR predicted to have a significant effect on hyaluronan synthase function was uniquely shared by all African mole-rats. Interestingly, we also identified multiple other amino acid substitutions in key domains of the HAS2 molecule, although the biological consequences of these for hyaluronan synthesis remain to be determined. Despite these results, we found evidence of strong purifying selection acting on the *HAS2* gene across all mammals, and the NMR remains unique in its particular HAS2 sequence. Our results indicate that more work is needed to determine whether the apparent cancer resistance seen in NMR is shared by other members of the African mole-rat clade.

Keywords: naked mole-rat, *Heterocephalus*, hyaluronan, Has-2, hyaluronan synthase, Bathyergidae, subterranean mammal

## **Introduction**

The naked mole-rat (*Heterocephalus glaber*) is emerging as an important “non-model” organism for the study of longevity and healthy ageing. A range of physiological and molecular/biochemical adaptations underpin the lack of senescence observed in this small Hystricomorph rodent that can live for 32 years – ten times longer than a mouse and more than five times longer than predicted for its body size (Buffenstein, 2008).

There has been considerable interest in the ability of naked mole-rats to resist cancer (Buffenstein, 2008; Delaney *et al.*, 2013), and recently, a mechanism involving the production of a high molecular mass hyaluronan (HMM-HA) has been proposed (Tian *et*

*al.*, 2013). Hyaluronan (HA) is a glycosaminoglycan and its presence in a variety of tissues and the extracellular matrix has been linked to many cellular processes, such cell division, motility and morphogenesis (Frazer *et al.*, 1997), and also implicated in the development of some cancers (Stern, 2009). HMM-HA is up to six times the mass of the largest human HA. The novel anticancer mechanism identified in the naked mole rat has been termed early contact inhibition (ECI). This is a process whereby cell growth occurring when cells come in to contact with each other, or the extracellular matrix, is arrested at much lower densities than in the mouse. Contact inhibition is lost in cancer cells, and the loss of ECI makes cells more susceptible to malignant transformation (Seluanov, *et al.*, 2009). ECI is controlled by the interaction of HA with the CD44–NF2 pathway which mediates contact inhibition (Tian *et al.*, 2013). In the naked mole-rat, it has been postulated that selection for its characteristic loose, elastic skin is an adaptation to living underground in tight tunnels and confined spaces, and that the elasticity of the skin is facilitated by HMM-HA. Thus, the cancer resistance imparted by HMM-HA may be a secondary and fortuitous consequence of the primary function of HMM-HA under selection.

Hyaluronan synthase 2 (HAS2) is one of three characterized membrane embedded HA synthases responsible for the synthesis of HA from intracellular precursors, and deposition into the extracellular matrix (Watanabe & Yamaguchi, 1996). Tian *et al.* (2013) showed that HMM-HA is produced in naked mole-rats by a uniquely modified version of the HAS2 gene. Specifically, two serine substitutions at highly conserved sites in the cytoplasmic domains of exons 2 and 3 appear to confer the protein molecule the

ability to produce HMM-HA: when naked mole-rat HAS2 was overexpressed in a human HEK293 cell culture, secretion of HMM-HA was observed (Tian *et al.*, 2013).

Given the link between naked mole-rat cancer resistance, HMM-HA and specific mutations in HAS2, together with the advantages of an elastic skin in the subterranean niche, we predict that production of HMM-HA may not be unique to naked mole-rats, and that discovery of other HAS2 mutations may be of interest to cancer research. This study therefore aims to investigate possible convergence and adaptive evolution of HAS2 in a broad range of mammals, but in particular in other species within the African mole-rat family (Bathyergidae), and in other subterranean mammals.

## **Methods**

Samples for HAS2 sequence analysis were selected from a total of 13 different subterranean mammal species, previously collected for phylogenetic studies (e.g. Faulkes *et al.* 2004; Davies *et al.*, 2014) and representing five divergent families (Table 1). These were combined with a further 57 representative mammalian HAS2 sequences from all species available at the time of the study on GenBank (Table 1). Sequences were aligned manually for analysis using Mesquite (Madison & Madison, 2014) and genetic distances calculated using MEGA 6 (Tamura *et al.*, 2013). The signature of selection acting along 519 codons of HAS2 (exons 2 and 4) was characterised across all 70 mammal species included in our study using site-wise, branch and clade models within the codeml package in PAMLv4.4 (Yang, 2007), using a mammal species tree topology based on published studies (Faulkes *et al.*, 2004; Blanga-Kanfi *et al.*, 2009; Meredith *et al.*, 2011).

**Table 1** – sample list, taxonomy and accession numbers/EMSEMBL Transcript ID for the HAS2 sequences included in for analysis; an asterisk denotes species sequenced for this study.

Scientific name	Common name	Order; Family	Accession number
<i>Homo sapiens</i>	human	Primates; Hominidae	U54804.1
<i>Pan troglodytes</i>	chimp	Primates; Hominidae	XM528222.4
<i>Pan paniscus</i>	bonobo	Primates; Hominidae	XM3820492.1
<i>Pongo abelii</i>	orangutan	Primates; Hominidae	XM3777303.1
<i>Nomascus leucogenys</i>	northern white-cheeked gibbon	Primates; Cercopithecidae	XM3256164.2
<i>Macaca fascicularis</i>	crab-eating macaque	Primates; Cercopithecidae	XM5564005.1
<i>Macaca mulatta</i>	rhesus macaque	Primates; Cercopithecidae	XM1098841.2
<i>Chlorocebus sabaeus</i>	green monkey	Primates; Cercopithecidae	XM8001453.1
<i>Saimiri boliviensis boliviensis</i>	squirrel monkey	Primates; Cebidae	XM3933098.1
<i>Callithrix jacchus</i>	common marmoset	Primates; Callitrichidae	XM2759268.2
<i>Otolemur garnettii</i>	Garnett's greater galago	Primates; Galagidae	XM_3782374.1
<i>Tarsius syrichta</i>	Philippine tarsier	Primates; Tarsiidae	XM8056607.1
<i>Tupaia chinensis</i>	Chinese tree shrew	Scandentia; Tupaiidae	XM_6157522.1
<i>Lipotes vexillifer</i>	Yangtze River dolphin	Cetacea; Lipotidae	XM_7445586.1
<i>Physeter catodon</i>	sperm whale	Cetacea; Physeteridae	XM_7106932.1
<i>Odobenus rosmarus divergens</i>	walrus	Carnivora; Odobenidae	XM_4416640.1
<i>Leptonychotes weddellii</i>	Weddell seal	Carnivora; Phocidae	XM_6739332.1
<i>Canis lupus familiaris</i>	dog	Carnivora; Canidae	XM_539153.4
<i>Panthera tigris altaica</i>	tiger	Carnivora; Felidae	XM_7075592.1
<i>Felis catus</i>	cat	Carnivora; Felidae	XM_4000089.2
<i>Ailuropoda melanoleuca</i>	panda	Carnivora; Ursidae	XM_2927908.1
<i>Mustela putorius furo</i>	ferret	Carnivora; Mustelidae	XM_4804975.1
<i>Myotis davidii</i>	David's mouse-eared bat	Chiroptera; Vespertilionidae	XM_6769087.1
<i>Myotis lucifugus</i>	little brown bat	Chiroptera; Vespertilionidae	XM_6085251.1
<i>Myotis brandtii</i>	Brandt's bat	Chiroptera; Vespertilionidae	XM_5885867.1
<i>Eptesicus fuscus</i>	big brown bat	Chiroptera; Vespertilionidae	XM_8143356.1
<i>Pteropus alecto</i>	black flying fox	Chiroptera; Pteropodidae	XM6916584.1
<i>Elephantulus edwardii</i>	Cape elephant shrew	Macroscelidea; Macroscelididae	XM_6879317.1
<i>Orycteropus afer afer</i>	aardvark	Tubulidentata; Orycteropodidae	XM_7943422.1
<i>Trichechus manatus latirostris</i>	Florida manatee	Sirenia; Trichechidae	XM_4372979.1
<i>Procavia capensis</i>	rock hyrax	Hyracoidea; Procaviidae	ENSPCAG00000005792
<i>Loxodonta africana</i>	African elephant	Proboscidea; Elephantidae	XM_003408169
<i>Echinops telfairi</i>	lesser hedgehog tenrec	Afrosoricida; Tenrecidae	XM_004697409
<i>Amblysomus hottentotus*</i>	golden mole	Afrosoricida; Chrysochloridae	
<i>Oryctolagus cuniculus</i>	European rabbit	Lagomorpha; Leporidae	AB055978.1
<i>Ochotona princeps</i>	American pika	Lagomorpha; Ochotonidae	XM_6982381.1
<i>Spermophilus tridecemlineatus</i>	thirteen-lined ground squirrel	Rodentia; Sciuridae	XM_5316195.1
<i>Mus musculus</i>	mouse	Rodentia; Muridae	U52524.2
<i>Rattus norvegicus</i>	rat	Rodentia; Muridae	AF008201.1
<i>Peromyscus maniculatus bairdii</i>	prairie deer mouse	Rodentia; Cricetidae	XM_6982381.1
<i>Cricetulus griseus</i>	Chinese hamster	Rodentia; Cricetidae	XM_7638417.1
<i>Mesocricetus auratus</i>	golden hamster	Rodentia; Cricetidae	XM_5082729.1
<i>Tachyoryctes splendens*</i>	East African root rat	Rodentia; Spalacidae	
<i>Nannospalax galili</i>	blind mole-rat	Rodentia; Spalacidae	ENSMUST00000050544
<i>Fukomys zechi*</i>	Ghana mole-rat	Rodentia; Bathyergidae	
<i>Fukomys damarensis*</i>	Damaraland mole-rat	Rodentia; Bathyergidae	
<i>Georchus capensis*</i>	Cape dune mole-rat	Rodentia; Bathyergidae	
<i>Cryptomys hottentotus*</i>	common mole-rat	Rodentia; Bathyergidae	
<i>Bathyergus janetta*</i>	Namaqua dune mole-rat	Rodentia; Bathyergidae	
<i>Bathyergus suillus*</i>	dune mole-rat	Rodentia; Bathyergidae	
<i>Heliophobius argenteocinereus*</i>	silvery mole-rat	Rodentia; Bathyergidae	
<i>Heterocephalus glaber</i>	naked mole-rat	Rodentia; Bathyergidae	XM_004883123

<i>Thryonomys swinderianus</i> *	cane rat	Rodentia; Thryonomyidae	
<i>Hystrix africaeaustralis</i> *	Cape porcupine	Rodentia; Hystricidae	
<i>Ctenomys sp</i> *	tuco tuco	Rodentia; Ctenomyidae	
<i>Cavia porcellus</i>	guinea pig	Rodentia; Caviidae	XM_003463665
<i>Chinchilla lanigera</i>	chinchilla	Rodentia; Chinchillidae	XM5410908.1
<i>Sorex araneus</i>	common shrew	Eulipotyphla; Soricidae	XM4607644.1
<i>Talpa europaea</i> *	European mole	Eulipotyphla; Talpidae	
<i>Condylura cristata</i>	Star nosed mole	Eulipotyphla; Talpidae	XM_4679612.1
<i>Erinaceus europaeus</i>	European hedgehog	Eulipotyphla; Erinaceidae	XM7520808.1
<i>Vicugna pacos</i>	alpaca	Artiodactyla; Camelidae	XM_6211418.1
<i>Bos taurus</i>	cow	Artiodactyla; Bovidae	XM_174079.2
<i>Capra hircus</i>	goat	Artiodactyla; Bovidae	XM_5688873.1
<i>Sus scrofa</i>	pig	Artiodactyla; Suidae	XM_214053.1
<i>Equus caballus</i>	horse	Perissodactyla; Equidae	XM_1081801.1
<i>Ceratotherium simum simum</i>	white rhinoceros	Perissodactyla; Rhinocerotidae	XM_4431107.1
<i>Dasyopus novemcinctus</i>	nine-banded armadillo	Cingulata; Dasypodidae	XM_4480750.1
<i>Monodelphis domestica</i>	opossum	Didelphimorphia; Didelphidae	XM_1370252.2
<i>Ornithorhynchus anatinus</i>	duck-billed platypus	Monotremata; Ornithorhynchidae	XM_1505190.2

---

## Results and discussion

Overall measures of variation among 70 taxa revealed p-distances (nucleotide) ranging from 0.08 (chimp versus bonobo) to 19.53% (Philippine tarsier versus Opossum) and from zero (e.g. human versus chimp) to 15.26% (Cape elephant shrew versus opossum) for amino acid substitutions (see Table S1 for distances and S2 for a complete amino acid alignment for all taxa). Site-wise, branch and clade models implemented in PAML did not find evidence for positive selection, but instead, site and clade models revealed significant ( $P < 0.001$ ) evidence of purifying (negative selection) with  $\omega$  values of less than one (see supplementary results). Despite this, amino acid substitutions are apparent in key parts of the HAS2 molecule, including those previously described for the NMR. Figure 1 shows a reduced dataset of 22 mammals from five clades of subterranean mammals, including non-subterranean in-group comparisons. We observed no obvious convergence in substitutions among subterranean mammals. Of particular interest are the residues at sites 178 and 301 that facilitate production of HMM-HA in NMRs. Our results show that the serine substitution at site 178 in NMRs is only present in one other mammal, the cane rat, a close outgroup to the Bathyergidae, and thus perhaps arose convergently. Interestingly, however, the neighboring site 177 has a serine residue substituted with an alanine in the cane rat and all bathyergids except the NMR. Replacement of a serine (which is readily phosphorylated and often important in the active site of enzymes) with an alanine is likely to have functional significance. The serine substitution at site 301 of the NMR is present in all bathyergid genera, but no other mammals, and is a shared derived character (synapomorphy) for the group (Figures S1-3). There are also a number of other unique substitutions in particular species of the

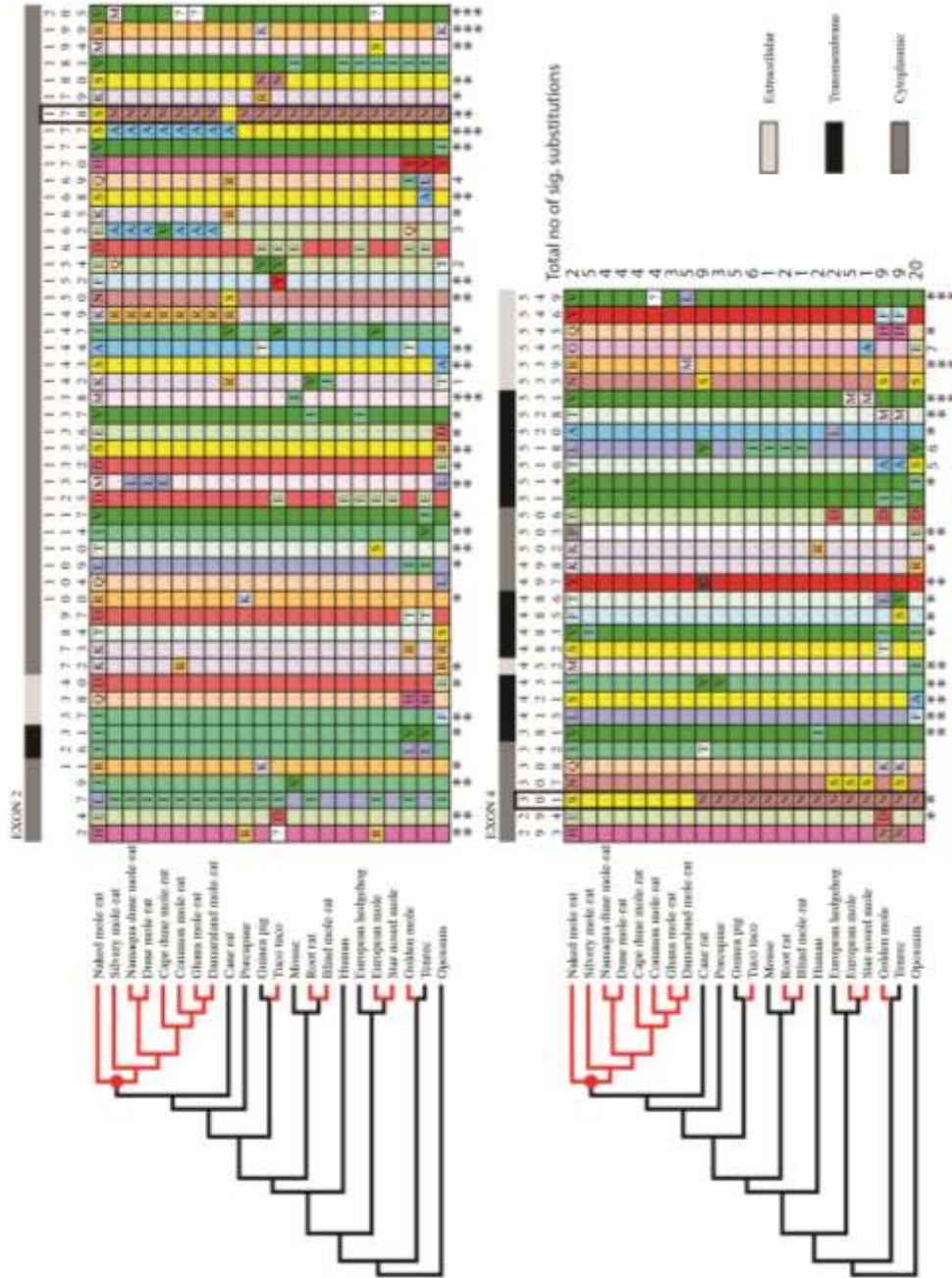


Figure 1. Phylogenetic relationships and corresponding HAS2 sequences of five clades containing subterranean mammals (red branches), including non-subterranean ingroup comparisons, the human, and a marsupial (opossum) outgroup (black branches). The Bathyergidae are the monophyletic clade denoted by the red circle. Adjacent panels show the respective variable amino acids for exons 2 and 4 (site numbers indicated above columns). Shaded bars indicate the relative locations of sites in the molecule (extracellular, transmembrane or cytoplasmic). The key amino acid residues at sites 178 and 301 - that facilitate production of high molecular mass hyaluronan in naked mole-rats - are indicated by the bold border. Asterisks below sites denote significance of substitutions estimated by MAPP analysis: \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ ; 1: R=ns, V=\*\*, I=\*, T=\*\*; 2: T=\*\*, V=\*, Q=ns; 3: V=\*\*\*, A=\*\*, Q=ns; 4: R=\*\*, I=ns, L=ns; 5: A=\*, S=ns; 6: V=\*, I=ns; 7: E=\*, A=ns (ns = not significant). Numbers at the end of the respective sequence alignments denote the number of substitutions per taxon that are predicted to have a significant impact on protein function.



Bathyergidae/cane rat (e.g. sites 149 and 162; Figure 1). Analysis of these polymorphisms using MAPP (Stone & Siddow, 2005), which implements a predictive statistical framework to score the physicochemical impact of substitutions in multiple alignments of orthologs, reveals multiple significant mutations. Among the bathyergids, the Damaraland and silvery mole-rats ranked highest with five significant substitutions, followed by four in common, Cape dune, dune and Namaqua dune mole-rats, and three in the Ghana mole-rat. The NMR had just two – the aforementioned substitutions at sites 178 and 301 (Figure 1 & Table S2). The blind mole-rat HAS2 sequence is unremarkable and similar to the mouse and root rat, despite the fact that this species has also been reported to produce HMM-HA (Tian et al., 2013). Thus the mechanism of HMM-HA production in blind mole-rats may differ to that of NMRs, and cancer resistance also reported in this species also appears to be mediated by a different mechanism (Gorbuova *et al.*, 2012). It is noteworthy that these substitutions were relatively low within the context of the entire mammal data set examined, where a maximum value of 21 substitutions predicted to have a significant affect were observed in the Cape elephant shrew (Table S2 and Supplementary Results).

These results raise interesting questions regarding the functional significance of the observed changes in HAS2 amino acid sequence, and whether there is any correlation with longevity and cancer resistance (where known). Within the Bathyergidae, so far no other taxa are known to live as long as NMRs. Species such as the dune mole-rats (genus *Bathyergus*) and *Georychus* generally have short life spans in the order of 4-6 years (Bennett *et al.*, 2006; N.C. Bennett unpublished data), although two *Georychus* are known to have lived for 10 and 11 years respectively in captivity (Weigl, 2005). The

Silvery mole-rat (*Heliophobius*) and some *Fukomys* (e.g. Damaraland and Zambian mole-rats) may live more than 7 and 10 years respectively (Tacutu *et al.*, 2013; Weigl, 2005). The absence of cancer has only been noted in NMRs, but there is a paucity of studies on other species in this context. Thus, the role of HAS2 and HA in other species remains unclear, but it is likely that multiple factors contribute to longevity. Nevertheless, our results and analysis provide the basis for further studies to establish the functional significance of HAS2 variants, using *in vitro* methods to characterise the different versions of HA produced, and the role they may play in cancer resistance.

### **Acknowledgments**

We thank the Ruth Rose for cloning, and the following for sample collection: Walter Verheyen (*Tachyoryctes*), Patricia Mirol (*Ctenomys*) and Clair and Mark Rylands (*Talpa*). NCB Acknowledges funding from the DST-NRF SARChI Chair for Behavioural Ecology and Physiology; SJR and KTJD were supported by the European Research Council.

### **References**

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry R, Huchon D. 2009 Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evol. Biol.* **9**, 71. (doi:10.1186/1471-2148-9-71)

Bennett NC, Maree S, Faulkes CG. 2006 *Georychus capensis*. Mammalian Species, **799**, 1–4.

Buffenstein R. 2008 Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *J. Comp. Physiol. B* **178**, 439–445.

Delaney MA, Nagy L, Kinsel MJ, Treuting PM. 2013 Spontaneous histologic lesions of the adult naked mole rat (*Heterocephalus glaber*): A retrospective survey of lesions in a zoo population. *Vet. Pathol.* **50**, 607–621.

Faulkes CG, Verheyen E, Verheyen W, Jarvis JUM, Bennett NC. 2004 Phylogeographic patterns of speciation and genetic divergence in African mole-rats (Family Bathyergidae). *Mol. Ecol.* **13**, 613–629.

Fraser JRE, Laurent TC, Laurent UBG. 1997 Hyaluronan: its nature, distribution, functions and turnover. *J. Intern. Med.* **242**, 27–33. doi:10.1046/j.1365-2796.1997.00170.x. PMID 9260563.

Gorbunova V, Hinea C, Tian X, Ablueva J, Gudkov AV, Nevo E, Seluanov A. 2012 Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. *P. Natl. Acad. Sci. USA* **109**, 19392–19396.

Gorbunova V, Seluanov A, Zhang Z, Gladyshev VN, Vijg J. 2014 Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat. Rev. Genet.* **15**, 531–540.

Maddison WP, Maddison DR. 2014 Mesquite: a modular system for evolutionary analysis. Version 3.01 <http://mesquiteproject.org>

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, et al. 2011 Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524. (doi: 10.1126/science.1211028)

Seluanov A, Hine C, Azpurua J, Feigenson M, Bozzella M, Mao Z, Catania KC, Gorbunova V. 2009 Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *P. Natl. Acad. Sci. USA* **106**, 19352–19357.

Stern R. 2009 *Association Between Cancer and “Acid Mucopolysaccharides”*: An Old Concept Comes of Age, Finally. In *Hyaluronan in Cancer Biology* (ed. R. Stern), pp. 3–16. San Diego, CA: Academic Press/Elsevier.

Stone EA, Siddow A. 2005 Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986.

Tacutu R, Craig T, Budovsky T, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhães JP. 2013 Human Ageing Genomic Resources: Integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.* **41**, D102–D1033.

Tamura K, Stecher G, Peterson D, FilipSKI A, Kumar S. 2013 MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.

Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, AbLaeva J, Mao Z, Nevo E, Gorbunova V, Seluanov A. 2013 High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499**, 346–349. (doi: 10.1038/nature12234)

Watanabe K, Yamaguchi Y. 1996 Molecular identification of a putative human hyaluronan synthase. *J. Biol. Chem.* **271**, 22945–22948.

Weigl R. 2005 Longevity of Mammals in Captivity; from the Living Collections of the World. Kleine Senckenberg-Reihe 48: Stuttgart.

Yang ZH. 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

## ***S1. SUPPLEMENTARY METHODS***

### ***(a) Sampling, PCR and sequencing***

Genomic DNA was extracted from muscle using a standard protocol (Faulkes *et al.*, 1997), and PCR amplification of exons 2 and 4 of the HAS2 gene carried out using primers designed from conserved regions in multiple species alignments of published HAS2 sequences. Exon 2 (627bp; 209 amino acids): HAS2E1F (5'-TGCATTGTGAGAGGTTTMTATG-3') and HAS2E1R (5'-CTGWACATARTCCACRCTBCG-3'); exon 4 (930bp; 310 amino acids): HAS2E3F (5'-CAARTAYGAYTCSTGGATYTCYTT-3') and HAS2E3R (5'-CATACTCMAGCACCATGTC-3'). PCR reactions followed a standard protocol using 10 ng of DNA in a 50 µl reaction containing Taq buffer, 1 Unit of Taq polymerase, 0.4 µM of each primer and 200 µM of dNTPs. After initially denaturing at 94°C for 2 min, 35 cycles of 94°C (30 s), 50°C (30 s), and 72°C (30 s) were carried out, followed by a final 72°C for 10 min. Negative controls were also included in each set of reactions. PCR products were purified using QIAquick PCR purification kits® (Qiagen, UK). Exons 1 and 3 were not sequenced as the former is non-transcribed, and the latter short (102bp; 34 amino acids) and invariant in amino acid sequence across mammals. Sequencing was carried out in both directions using the PCR primers to obtain complementary partially or fully overlapping strands, using the Eurofins Genomics Value Read automated sequencing service (Eurofins Genomics, Ebersberg, Germany). In species where sequencing failed using the above primers, PCR products were cloned using a pGEM®-T Easy Vector System (Promega, UK), and positive clones sequenced using standard plasmid primers T7 and SP6.

### ***(b) Phylogenetic analysis***

In addition to *de novo* sequencing of DNA from 13 species, a further 57 representative mammalian HAS2 sequences from all species available at the time of the study were retrieved from GenBank (Table 1). Sequences were aligned manually for analysis using Mesquite (Madison & Madison, 2014), and the species tree topology (e.g. Figures S1-3) used for selection analyses based on published studies (Faulkes *et al.*, 2004; Blanga-Kanfi *et al.*, 2009; Meredith *et al.*, 2011). Genetic distances were calculated and phylogenetic trees based on both nucleotide and amino acid sequences constructed using MEGA 6 (Tamura *et al.*, 2013).

### ***(c) Testing for selection***

#### *Site models*

We characterised the signature of selection acting along *HAS2* across all 70 mammal species included in our study. Site-wise  $\omega$  values were estimated across all branches in the phylogeny for the alignment consisting of 519 codons run with the codeml package in PAMLv4.4 (Yang, 2007). These  $\omega$  values were assigned to predefined site classes according to each site model (e.g. M1a had two and M2a had three) (Nielsen and Yang 1998; Wong *et al.* 2004). We estimated the mean  $\omega$  of each site class, and the proportions of sites falling into each class. To test where and how  $\omega$  varied among sites; three model comparisons were carried out. Firstly, we assessed whether  $\omega$  varied among sites by comparing a model with a single free  $\omega$  (M0) to one in which  $\omega$  fell into three discrete classes (M3). Secondly, to test for positive selection, we compared model M1a (Nearly Neutral) in which site classes were neutral ( $\omega=1$ ) and

purifying ( $0 < \omega < 1$ ) to model M2a (Positive Selection) that had a third site class corresponding to positive selection ( $\omega > 1$ ). For a second test of adaptation, we compared model M7 (beta) to M8 (beta& $\omega$ ), in which the latter had an additional site class in which  $\omega$  could exceed one. Likelihood ratio tests (LRT) were used for all model comparisons.

#### *Branch-site models*

In order to determine if any particular sites along *HAS2* were under positive selection in two particular branches of interest we implemented branch-site models (Zhang et al. 2005). In our first test we set the naked mole-rat branch as the foreground branch of interest; in the second test the ancestral mole-rat + cane rat branch was set as the foreground branch of interest. In each branch-site test for positive selection the site-wise estimates of  $\omega$  for the branch that has been designated as the foreground branch of interest are compared with estimates across the remaining background branches in the species phylogeny under model A. The model parameters are then compared with those of the null model A with a likelihood ratio test (LRT), and the significance of the model fit assessed by ChiSq statistic with one degree of freedom, with *P*-values  $< 0.05$  indicating the alternative model has a significantly better fit compared to the null.

#### *Clade models*

We also tested for evidence of divergent selection pressures acting on any of the branches within a focal clade of interest corresponding to the Bathyergidae + cane rat. In this test all branches of the Bathyergidae + cane rat clade, including the ancestral branch is set as the foreground clade and the remaining branches are the background clade (see Figure S1 for tree topology). We implemented the clade model C (Bielawski and Yang 2004), in which the estimated averaged  $\omega$  of the foreground



clade is compared to that of the averaged  $\chi^2$  of the background clade. Values estimated by each clade model were then compared with model M1a (nearly neutral) via a LRT with three (DF), again with  $P$ -values  $<0.05$  indicating the alternative model has a significantly better fit compared to the null.

#### ***(d) Multivariate Analysis of Protein Polymorphism***

Multivariate Analysis of Protein Polymorphism (MAPP) uses an amino acid sequence alignment to calculate the predicted impact of each potential SNP at each position. These predictions are based on a set of scales of physicochemical properties (hydropathy, polarity, charge, volume and free energy in alpha-helix and beta-strand conformation), for which each amino acid has a numeric value.  $P$ -value interpretations of the MAPP scores are then calculated, predicting the impact of each amino acid variant (Stone & Sidow, 2005; [http://mendel.stanford.edu/supplementarydata/stone\\_MAPP\\_2005](http://mendel.stanford.edu/supplementarydata/stone_MAPP_2005)).

## ***S2. SUPPLEMENTARY RESULTS***

#### ***(a) Phylogenetic analysis***

A summary of pairwise genetic distances for the complete mammalian dataset of 70 species are presented in Table S1. Nucleotide uncorrected  $p$ -distances ranged from 0.001 (humans versus chimps) to 0.195 (Philippine tarsier versus opossum), reflecting phylogenetic proximity on the species tree (Figures S1-3). Amino acid differences ranged from zero among all anthropoid primates except the bonobo, where there is a unique phenylalanine (F) to leucine (L) substitution at site 398 (an extracellular region), to 0.153 (Cape elephant shrew versus opossum). Interestingly, the opossum consistently across all sites shows greater numbers of amino acid differences with eutherian mammals than the more divergent platypus, despite greater nucleotide distances in the latter (Table S1). Maximum likelihood trees were inferred by using

the K2+G+I model (for nucleotide sequences) and the JTT+G model (for amino acid sequences). The nucleotide based tree recovered many of the main clades of the species tree (as in Figures S1-3) with bootstrap support of >90%, with anomalous groupings occurring as polytomies or clades with very low bootstrap values. The amino acid based tree was not informative as only three clades were recovered with bootstrap support of >90%, one with the African mole-rats *excluding* the naked mole-rat (93%), another with four Afrotherians (Cape elephant shrew, aardvark, tenrec and golden mole; 98%) and the third with the microchiropteran bats (93%).

### ***(b) Selection analysis***

#### *Site models*

Our site-wise analysis of *HAS2* sequences across 70 mammal species found no evidence of positive selection as both pair-wise tests for positive selection (M1a vs. M2a and M7 vs. M8) were not significant (LRT:  $P=1.000$  and  $P=0.427$ ). Instead, sites along *HAS2* were found to be under purifying selection with all three estimated  $\omega$  categories of Model 3 (discrete) being  $<1.0$  (LRT: M0 vs. M3  $P<0.0001$ ) (Table 1).

#### *Branch-site models*

There was no evidence of sites under positive selection along the naked mole-rat branch (LRT:  $P=1.000$ ) or along the ancestral branch of the Bathyergidae + cane rat (LRT:  $P=1.000$ ) in either of our branch-site tests.

#### *Clade models*

Clade models performed to test for divergent selection between the focal clade of Bathyergidae + cane rat compared to the remaining mammal species detected significant divergent selection (LRT:  $P<0.001$ ). However, the estimated omega of both the foreground and background clades was  $<1.00$  and, therefore, both clades were found to be under purifying selection ( $FG\omega = 0.156$ ;  $BG\omega = 0.126$ ).

Clade models performed to test for divergent selection between the focal clade of Bathyergidae + cane rat compared to the remaining Euarchontoglires species (Rodentia, Lagomorpha, Primates and Scandentia) detected significant divergent selection (LRT:  $P < 0.001$ ). However, the estimated omega of both the foreground and background clades was  $< 1.00$  and, therefore, both clades were found to be under purifying selection ( $FG\omega = 0.215$ ;  $BG\omega = 0.112$ ).

Clade models performed to test for divergent selection between the focal clade of Bathyergidae + cane rat compared to a background clade of four outgroup Hystricomorph species (porcupine, tuco-tuco, chinchilla and guinea pig) did not detect significant divergent selection (LRT:  $P = 1.000$ ). Therefore, the null model (M1a nearly neutral) was found to fit the data better, with the majority of sites ( $\sim 0.99$ ) found to be under purifying selection ( $\omega = 0.020$ ).

### ***(c) Multivariate Analysis of Protein Polymorphism (MAPP) analysis***

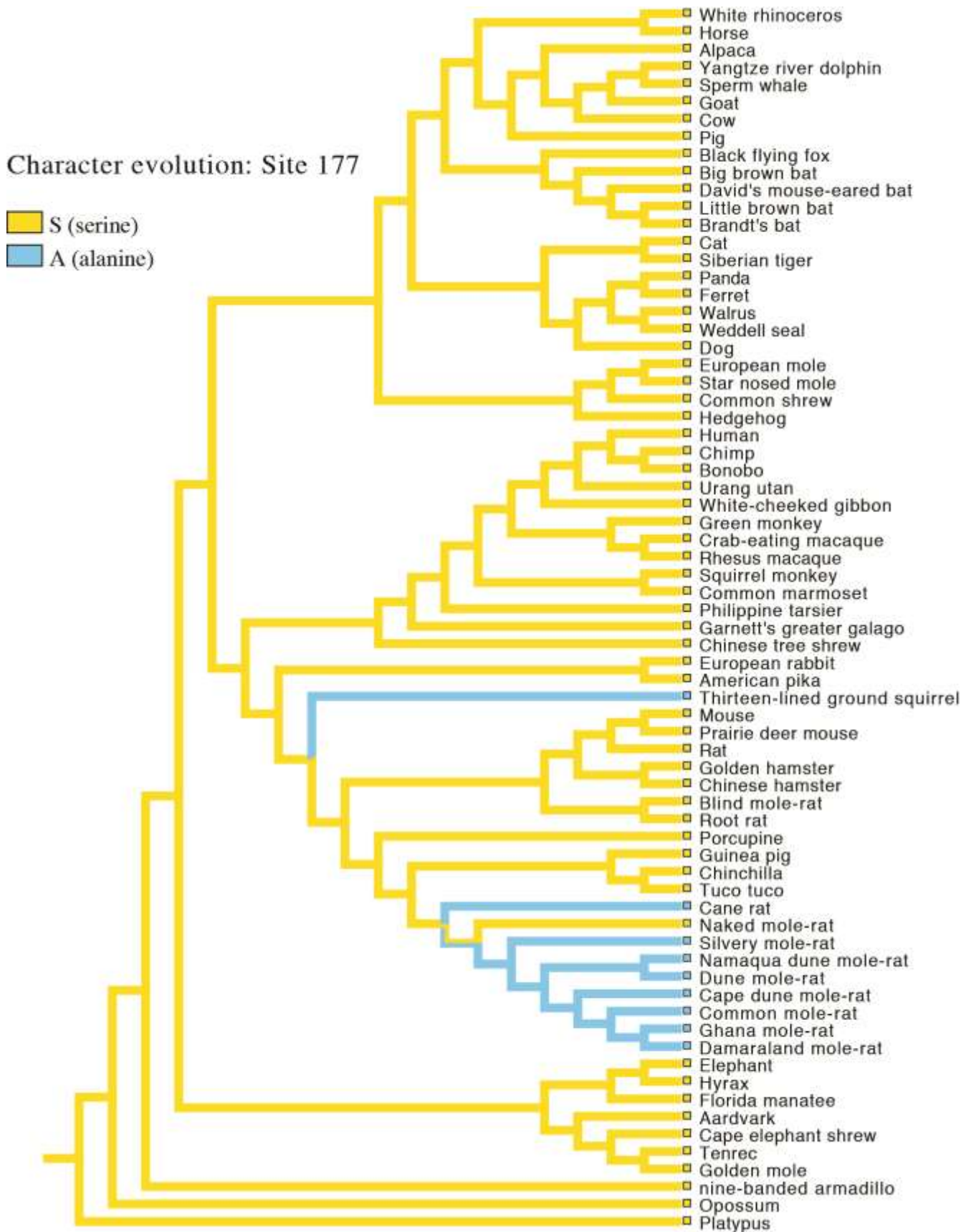
A full summary of the numbers of substitutions per taxon that are predicted by MAPP analysis to have a significant impact on protein function is presented in Table S3. These values ranged from zero to a maximum of 21 significant substitutions in the Cape elephant shrew, closely followed by the opossum with 20 substitutions. Third ranking was a much lower value of 11, observed in the armadillo. In this context, the two significant substitutions of the naked mole-rat fall at the bottom end of the distribution. Amino acid residues that are crucial for *N*-acetylglucosaminyltransferase activity at sites 212 (aspartic acid), 314 (aspartic acid), 350 (glutamine), 353 (arginine) and 354 (tryptophan) (Watanabe & Yamaguchi, 1996) are fully conserved across all mammals in our dataset.



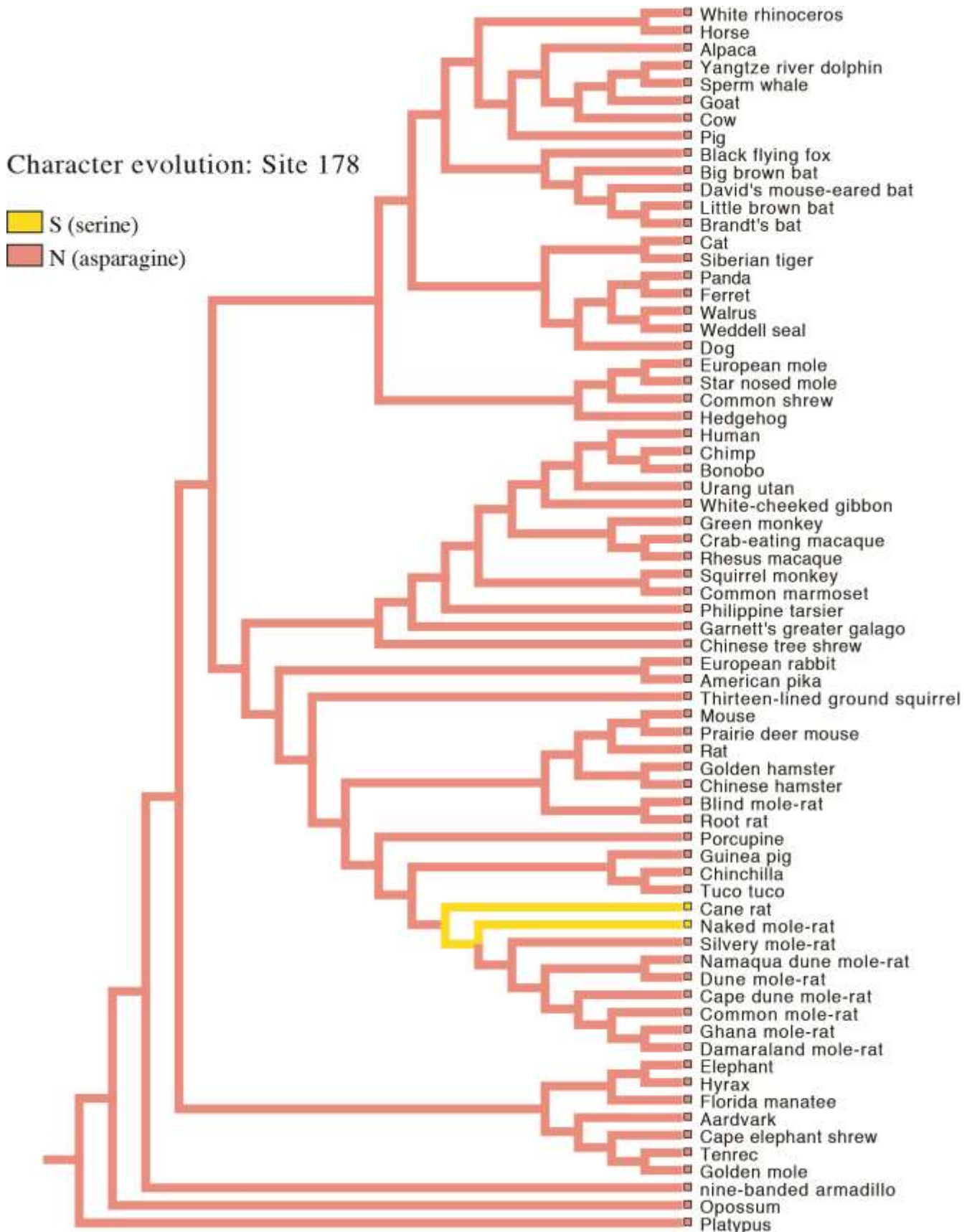


#### S4. SUPPLEMENTARY FIGURES

**Figure S1:** Species tree illustrating the character evolution at Site 177 of exon 2 of the HAS2 gene, indicating gains and losses of alanine.

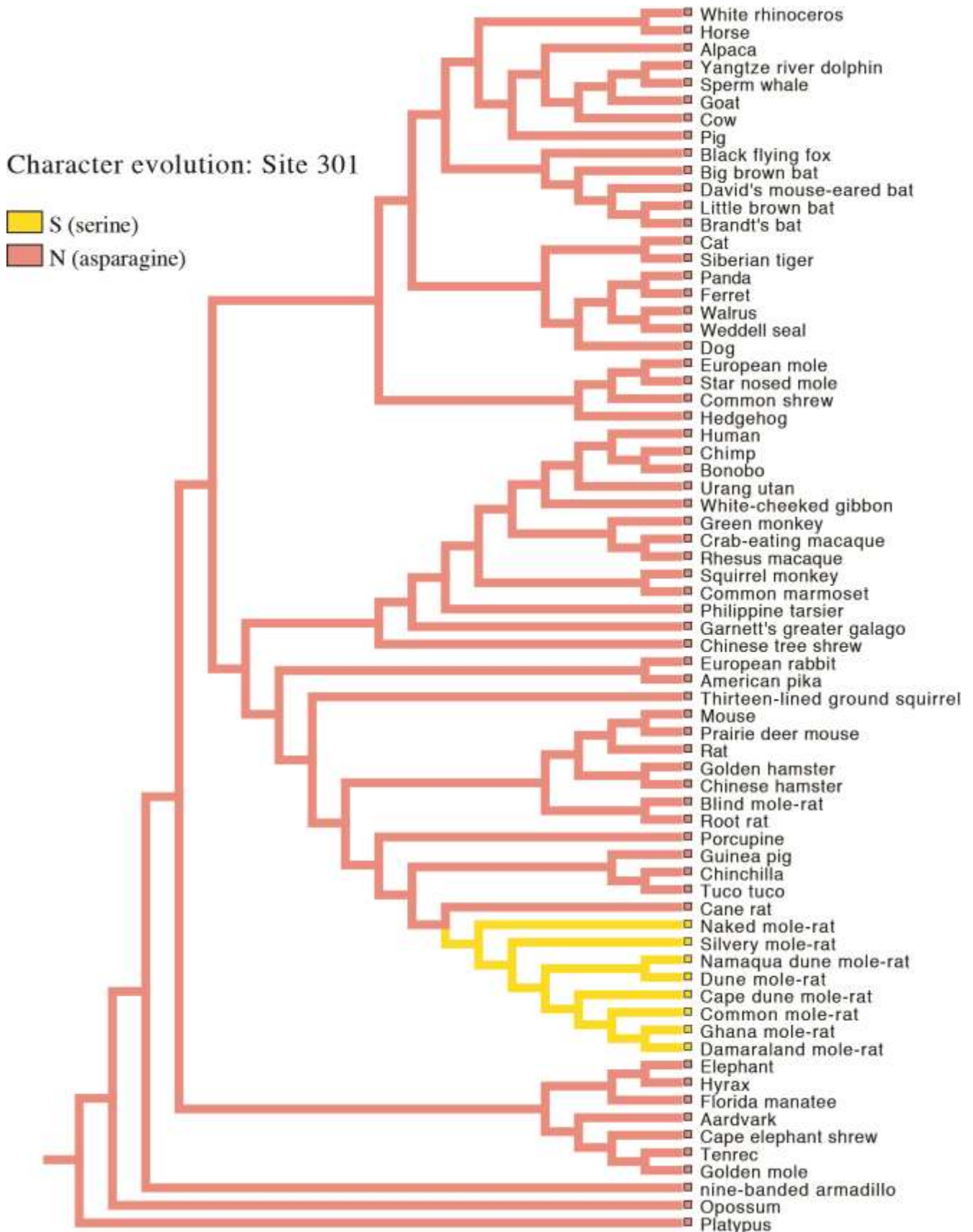


**Figure S2:** Species tree illustrating the character evolution at Site 178 of exon 2 of the HAS2 gene, indicating gains and losses of serine.





**Figure S3:** Species tree illustrating the character evolution at Site 301 of exon 4 of the HAS2 gene, indicating gain of a serine substitution in the bathyergid clade.





## Supplementary references

Bielawski JP, ZH Yang. 2004 A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121–132.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry R, Huchon D. 2009 Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evol. Biol.* **9**, 71. (doi:10.1186/1471-2148-9-71)

Faulkes CG, Verheyen E, Verheyen W, Jarvis JUM, Bennett NC. 2004 Phylogeographic patterns of speciation and genetic divergence in African mole-rats (Family Bathyergidae). *Mol. Ecol.* **13**, 613–629.

Faulkes CG, Abbott DH, O'Brien HP, Lau L, Roy M, Wayne RK, Bruford MW. 1997 Micro- and macro-geographic genetic structure of colonies of naked mole-rats, *Heterocephalus glaber*. *Mol. Ecol.* **6**, 615–628.

Maddison WP, Maddison DR. 2014 Mesquite: a modular system for evolutionary analysis. Version 3.01 <http://mesquiteproject.org>

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, et al. 2011 Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification.

*Science* **334**, 521–524. (doi: 10.1126/science.1211028)

Nielsen R, Yang ZH. 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936.

Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013 MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.

Stone EA, Siddow A. 2005 Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986.

Wong WSW, Yang Z, Goldman N, Nielsen R. 2004 Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051.

Zhang JZ, Nielsen R, Yang ZH. 2005 Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479.