Adaptations to a Subterranean Environment and Longevity Revealed by the Analysis of Mole Rat Genomes

Xiaodong Fang,1,2,10 Inge Seim,3,4,10 Zhiyong Huang,1 Maxim V. Gerashchenko,3 Zhiqiang Xiong,1 Anton A. Turanov,3 Yabing Zhu,1 Alexei V. Lobanov,3 Dingding Fan,1 Sun Hee Yim,3 Xiaoming Yao,1 Siming Ma,3 Lan Yang,1 Sang-Goo Lee,3,4 Roderick T. Bronson,5 Radim Sumbera,6 Rochelle Buffenstein,7 Xin Zhou,1 Anders Krogh,2 Thomas J. Park,6 Guojie Zhang,1,2,9 and Vadim N. Gladyshev3,4,*

1BGI-Shenzhen, Shenzhen 518083, China
2Department of Biology, University of Copenhagen, Copenhagen, 2200 Copenhagen N, Denmark
3Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
4Department of Bioinspired Science, Ewha Womans University, Seoul 120-750, South Korea
5Rodent Histopathology Laboratory, Harvard Medical School, Boston, MA 02115, USA
6Department of Biology, University of South Bohemia, Faculty of Science, Ceske Budejovice 37005, Czech Republic
7Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, TX 78245, USA
8Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA
9King Abdulaziz University, Jeddah 21441, Saudi Arabia
10Co-first author
*Correspondence: wangj@genomics.org.cn (J.W.), vgladyshev@rics.bwh.harvard.edu (V.N.G.)
http://dx.doi.org/10.1016/j.celrep.2014.07.030
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

SUMMARY

Subterranean mammals spend their lives in dark, unventilated environments that are rich in carbon dioxide and ammonia and low in oxygen. Many of these animals are also long-lived and exhibit reduced aging-associated diseases, such as neurodegenerative disorders and cancer. We sequenced the genome of the Damaraland mole rat (DMR, *Fukomys damarensis*) and improved the genome assembly of the naked mole rat (NMR, *Heterocephalus glaber*). Comparative genome analyses, along with the transcriptomes of related subterranean rodents, revealed candidate molecular adaptations for subterranean life and longevity, including a divergent insulin peptide, expression of oxygen-carrying globins in the brain, prevention of high CO2-induced pain perception, and enhanced ammonia detoxification. Juxtaposition of the genomes of DMR and other more conventional animals with the genome of NMR revealed several truly exceptional NMR features: unusual thermogenesis, an aberrant melatonin system, pain insensitivity, and unique processing of 28S rRNA. Together, these genomes and transcriptomes extend our understanding of subterranean adaptations, stress resistance, and longevity.

RESULTS AND DISCUSSION

Genome Assembly and Gene Content

The DMR genome yielded a 2.5 Gb sequence (~76-fold coverage) with a scaffold N50 size of 5 Mb (Table 1; Figure S1A). The sequencing depth of 91% of the DMR assembly had more than 10-fold coverage (Figure S1B). We identified 1.3 million...
heterozygous SNPs and estimated a nucleotide diversity (heterozygosity) of 0.06%, which is comparable to that in the NMR, but lower than that in rodents such as mouse and rat (Kim et al., 2011). The low level of nucleotide diversity in the DMR and NMR may reflect their unique social system, which involves a single breeding “queen” per colony, and the low effective size of their populations (Bennett and Faulkes, 2000). The number of repeat elements in the DMR genome was also lower (24.28%) than in other mammals but comparable to that of the NMR (Table 1; Kim et al., 2011). We employed homology and de novo methods as well as RNA sequencing (RNA-seq) data to predict 22,179 protein-coding genes in the DMR genome (Table 1; Figure S1C), which is comparable to what is predicted for other mammals. Our analysis revealed that the common ancestor of the DMR and NMR lived approximately 26 million years ago (Mya) (Figures 1C and S1D), which is similar to the distance between mice and rats, or between humans and macaques.

We further prepared a version of the NMR genome based on the original genome sequence (Kim et al., 2011), additional sequencing, and data generated by the Broad Institute (Table 1). The new NMR assembly had a genome size of 2.7 Gbp (92-fold coverage), with a scaffold N50 of 21 Mb compared with 1.6 Mb in the previously published assembly (Kim et al., 2011). The resulting DMR and NMR genomes, gene models, and transcriptome data for these and related rodents were used to reveal both common and unique features of these animals. We primarily focused on genes that are likely to be involved in the ecophysiology and exceptional longevity of underground-dwelling African mole rats.

Sensory Cues
Analyses of gene family contractions and expansions provide insights into the evolutionary forces that have shaped genomes. Among the 19,839 gene families that are inferred to be present in the most recent common ancestor of mammals, we found that 212 gene families were gained and 59 were lost from the DMR genome (Figure S1E; Table S1). Over the same period, the NMR gained 378 gene families and lost 29. The gene families gained included olfaction (sense of smell) genes that likely play an important role in social interaction and locating food in complete darkness (Heth and Todrank, 2007). The NMR and DMR live exclusively in the dark and display small eyes and poor visual acuity (Bennett and Faulkes, 2000). However, their eyes can still serve to alert the colony to invasion by predators by detecting light entering their tunnels (Nemec et al., 2008; Kott et al., 2010). The visual perception category was enriched in both the DMR (Table S2) (Gene Ontology [GO]: 0007601, p < 0.001, Fisher’s exact test) and NMR pseudogene lists. We found that one visual perception gene (AOC2) was lost and 13 were pseudogenized in the DMR and the new NMR genome assembly (Table S2). Three visual genes (CRB1, GRK7, and GJA10) were inactivated or missing in the DMR and the new NMR genome assembly (Table S2). Positive selection of the rhodopsin gene RHO, which enables dim-light vision, was found in the lineage leading to the common ancestor of the DMR and NMR (Table S3). This is consistent with evidence showing that African mole rat RHO underwent accelerated evolution while preserving sites critical for spectral tuning (Zhao et al., 2009). Interestingly, we observed cataracts in all examined NMRs ranging from 4 to 20 years of age (Figure S2A). This phenotype may be a consequence of captive life under atmospheric oxygen levels, but could also highlight an inadequate antioxidant defense. Low glutathione peroxidase 1 (GPx1) levels may contribute to a decreased protection of the lens against oxidative stress (Kasaikina et al., 2011). A premature stop codon occurs in the gene encoding GPx1 (GPX1) in both the NMR and DMR (Figure S2B), and knockout of this gene in mice results in cataract formation (Reddy et al., 2001; Wang et al., 2009; Wolf et al., 2005).
Adaptations to Hypoxia and a High Carbon Dioxide and Ammonia Environment

The DMR, NMR, and other subterranean rodents rest with conspecifics in underground environments low in oxygen and high in carbon dioxide and ammonia—conditions that would evoke cellular damage and behavioral stress responses in other mammals (Bennett and Faulkes, 2000). Ammonia is a potent irritant that arises from nitrogen and methane accumulation in latrines and nests (Burda et al., 2007; LaVinka et al., 2009). We found that arises from nitrogen and methane accumulation in latrines and nests (Burda et al., 2007; LaVinka et al., 2009). We found that the motif is also present in the DMR, NMR and DMR, we generated the transcriptomes of three subterranean rodents with those of rodents living primarily “aboveground” (surface dwelling). In addition to the NMR and DMR, we generated the transcriptomes of three subterranean hystricognath rodents: the FA, the FD, and the coruro of South America. We further compared them with rat and two guinea pig subspecies (Table S5).

Changes in both gene expression and gene sequences contribute to adaptive mechanisms in subterranean rodents (Avivi et al., 2010). We compared the normoxic brain transcriptomes of subterranean rodents with those of rodents living primarily “aboveground” (surface dwelling). In addition to the NMR and DMR, we generated the transcriptomes of three subterranean hystricognath rodents: the FA, the FD, and the coruro of South America. We further compared them with rat and two guinea pig subspecies (Table S5).

Several genes associated with DNA damage repair and responses to stress showed higher expression in subterranean rodents even during normoxia (Table S5; Figure S2C). Hypoxia induces DNA damage, and in agreement with recent reports on the blind mole rat (Fang et al., 2014; Shams et al., 2013), our data suggest that improved DNA repair is an intrinsic mechanism of adaptation to an underground environment. The most obvious adaptation to a hypoxic subterranean environment is improved oxygen uptake to highly oxygen-demanding tissues, such as the brain. The globin family comprises proteins that are responsible for the delivery and storage of oxygen in cells and tissues. We found that hemoglobin $\alpha$ (HBA1 and HBA2,
identical coding sequences) and neuroglobin (NGB) displayed elevated expression in the brains of subterranean rodents during normoxia (Figure 2F; Table S5), and western blot analysis verified higher hemoglobin a protein expression in the normoxic brain of the NMR compared with that of several surface-dwelling rodents (Figure 2G). We next compared the gene expression of NMR and DMR with the hypoxia-sensitive rat after 8 hr at oxygen levels comparable to those found in NMR burrows (8% O2) (Bennett and Faulkes, 2000). Similar to what was observed under normoxia, hemoglobin a and neuroglobin expression in the hypoxic brain was higher in the NMR and DMR than in the rat (Figure 2H). We observed a 3.4-fold decrease of hemoglobin a mRNA in the DMR under hypoxia, whereas expression in the NMR did not change significantly. Higher NGB expression was observed in the hypoxic rat and NMR brain, but not in the DMR brain, and there was a trend toward higher cytoglobin (CYGB) expression in the DMR. Species-specific expression of globins in response to hypoxia was previously reported in subterranean blind mole rat species (family Spalacidae) that are distantly related to African mole rats (family Bathyergidae) (Avivi et al., 2010). Hemoglobin has a higher affinity for oxygen and is able to unload oxygen more efficiently in the NMR than in the mouse (Johansen et al., 1976), and hemoglobin a plays a neuroprotective role in the brain of rodents during hypoxia (Schelshorn et al., 2009). A unique amino acid change (Pro44His) has recently been hypothesized to convey the adaptation to the subterranean and high-altitude habitats of the NMR and guinea pig, respectively (Fang et al., 2014). Neuroglobin and cytoglobin mRNA expression is elevated in two blind mole rat species compared with the rat under normoxia (Avivi et al., 2010). Importantly, neuroglobin is expressed in blind mole rat glial cells and neurons, whereas its expression is limited to neurons in surface-dwelling rodents, such as the rat, indicating an organ-wide protective function. Glial globin expression has also been
associated with improved oxygen delivery in hypoxia-tolerant 
seals (Schneuer et al., 2012) and shellfish (Kraus and Colacino, 
1986), which is suggestive of a common adaptation with subter-
ranean rodents. Overall, our data suggest that globins are constit-
tutively and highly expressed in the brain of hypoxia-resistant 
rodents and play a major role in their ability to adapt to an oxy-
gen-poor subterranean environment. Experiments are in pro-
gress to further dissect the expression and function of globins in 
African mole rats.

Potential Longevity-Associated Adaptations in the NMR 
and DMR

Subterranean rodents have the highest maximum lifespans for 
their body weight, with species in both the Bathyergidae (e.g., 
the NMR and DMR) and Spalacidae (e.g., the blind mole rat) 
families living for over 20 years (Dammann and Burda, 2007). 
The NMR is the longest-lived rodent known, with a lifespan 
exceeding 30 years, while the longest-lived DMRs in our labora-
tories survived for 20 years. These rodents have a longevity 
quotient similar to that of humans and may show a comparable 
age-related disease pattern (Edrey et al., 2011).

We compared the transcriptomes of the liver (a relatively ho-
mogeneous organ) of the NMR and DMR (Hystricognathi) with 
those of the short-lived rat and mouse (Muridae) (Table S4). 
Compared with the mouse and rat, which spend considerable 
time aboveground, the NMR and DMR showed differential 
expression and enrichment of several genes associated with ox-
idoreduction. Two out of six peroxiredoxins (PRDX2 and PRDX5) 
were expressed at lower levels (Table S4) in NMR and DMR 
livers, which, together with reduced GPx1 activity, may result 
in increased levels of reactive oxygen species (ROS). These ob-
ervations are consistent with reports of oxidative stress in the 
NMR (Andziak et al., 2006), and suggest that the long-lived 
NMR and DMR can thrive despite elevated oxidative stress. 

Loss of FASTK, a Sensor of Mitochondrial Stress, 
in African Mole Rats

We found that the Fas-activated serine/threonine kinase gene 
(FASTK) is inactivated in both the NMR and DMR (Figure S2D). 
FASTK encodes a kinase that serves as a regulator of Fas-
mediated apoptosis and is located at the inner mitochondrial 
membrane. FASTK is associated with cell survival and is over-
expressed in tumors and immune-mediated inflammatory diseases 
such as asthma and AIDS, where it can delay the onset of 
apoptosis and contribute to pathogenesis. Knockdown of this 
gene results in reduced lung inflammation in mice (Simarro 
et al., 2010) and reduced oncogenic potential of cultured human 
cancer cells (Zhi et al., 2013). Chronic inflammation, cancer, and 
cellular senescence are intertwined in the pathogenesis of pre-
mature aging (Campisi et al., 2011). Furthermore, knockdown of 
FASTK is also associated with improved neuron elongation 
and regeneration (Loh et al., 2008). Both the neurons’ ability to 
regenerate and their rate of elongation decrease with age. Loss of FASTK may help maintain neuronal integrity in long-lived 
 mole rats, keeping their brains “younger.” Thus, the loss of 
FASTK in the NMR and DMR suggests a role for FASTK in the ag-
ing phenotype of somatic cells as well as in cancer resistance.

Divergent Insulin in African Mole Rats

It has been reported that NMR insulin cannot be detected using 
rodent antibody-based assays, similar to what was found in the 
guinea pig several decades ago (Chan et al., 1984; Kramer and 
Buffenstein, 2004). We found that the NMR, DMR, and other 
hystricognath rodents harbor a divergent insulin β-chain sequence (Figure 3A). This finding is consistent with the observation that 
insulin in the South American hystricognath is rapidly evolving 
(Opazo et al., 2005). In the guinea pig and other South American 
hystricognaths, the regions encoding the α chain and, in partic-
ular, the β-chain are highly divergent, with concomitant alter-
ations in insulin structure and reduced activity compared with 
most other mammals, and possibly an alternative receptor 
(King et al., 1983; Opazo et al., 2004). Mutations in the human 
β-chain result in reduced insulin processing, misfolding, and 
less effective insulin (based on receptor binding) (Liu et al., 
2010). Interestingly, residue 22 of the β-chain, whose mutation 
(Arg22Gln) is associated with misfolding of insulin and diabetes 
(Liu et al., 2010), is uniquely changed in both African and South 
American hystricognaths (Figure 3A). In the African crested por-
cupine, this residue was previously linked to an altered insulin 
structure with reduced affinity for insulin receptors (Horuk 
et al., 1980). We hypothesize that NMR and DMR insulin exists 
as a monomer with low insulin receptor activity that targets alter-
native receptor(s) outside classic insulin-responsive tissues such 
as liver, muscle, and adipose tissue.

Surprisingly, the NMR (Edrey et al., 2011) and South American 
hystricognaths (Opazo et al., 2004) are able to handle glucose in 
the absence of conventional insulin, suggesting that these ani-
mals have evolved compensatory mechanisms. In mammals,
insulin is not secreted from the pancreas until after birth, and mice lacking insulin die a few days after birth due to acute diabetes mellitus (Duvillé et al., 1997). Until recently, it was unknown how glucose handling in the liver was achieved before birth. It has now been established that insulin growth factor 2 (IGF2), which has high homology to insulin, is abundantly expressed in the fetal liver and signals exclusively via the insulin receptor (IR) to maintain glycemia (Liang et al., 2010a). In most mammals, including mice and rats, IGF2 expression is downregulated after birth in the liver; however, primates and guinea pigs harbor residual IGF2 expression (Lui and Baron, 2013). We found that the NMR and DMR also express IGF2 and its binding protein, IGF2BP2, in the liver (Figure 3B; Table S4). We hypothesize that autocrine/paracrine production of IGF2 in the liver substitutes for insulin and may partly mediate a fetal-like mode of glucose handling in hystricognath rodents (Figure 3C).

Reduced levels of insulin are observed during calorie restriction and inhibition of the growth hormone/IGF1 axis, two manipulations that extend lifespan in various species (Blagosklonny, 2012). Interestingly, molecular innovations of this axis may contribute to the lifespan of the long-lived Brandt’s bat (Seim et al., 2013). In addition to induction of IGF2 expression in NMR and DMR livers, we observed differential expression of genes associated with insulin signaling: decreased IGF1 and insulin induced gene 2 (INSIG2), and increased IGF1R and resistin (RETN) (Table S4). Taken together, these results suggest that a less bioactive insulin and altered downstream signaling may partly explain the enhanced longevity of African mole rats and possibly other hystricognaths (e.g., porcupine and guinea pig). Our findings support the hypothesis that hystricognath rodents have evolved a distinct insulin peptide.

Cancer Resistance

Studies of the NMR (Liang et al., 2010b; Manov et al., 2013; Seluanov et al., 2008, 2009) and the distantly (~70 million years) related blind mole rat (Gorbunova et al., 2012; Nasser et al., 2008; Manov et al., 2013) suggest that many species of long-lived mole rats are resistant to cancer, and even if they do develop pathology, will present a milder phenotype in comparison with short-lived rodents (e.g., mouse) (Azpurua and Seluanov, 2012; de Magalhães, 2013).

A recent study suggested that one potential explanation for mole rats’ cancer resistance lies in the enzyme hyaluronan synthase 2 (HAS2) (Tian et al., 2013). Two amino acid residues in the HAS2 active site were reported to be unique to the NMR and hypothesized to result in the synthesis of high-molecular-mass hyaluronan (HMM-HA), an extracellular matrix polysaccharide. HMM-HA serves as an extracellular signal that results in induction of the tumor suppressor p16INK4A, early contact inhibition, and cancer resistance (Tian et al., 2013). We found that one of the unique amino acid changes in the NMR HAS2 sequence (Asn301Ser) is shared by the DMR, whereas Asn178Ser is unique to the NMR (Figure S2E). In contrast to residue 178, Asn301Ser is present in a highly conserved region. Interestingly, the blind mole rat also secretes HMM-HA (Tian et al., 2013). Taken together, these data suggest that the DMR and the blind mole rat produce HMM-HA that confers cancer resistance. Surprisingly, a recent study found that HMM-HA does not influence the anticancer properties of blind mole rat fibroblasts (Manov et al., 2013), which supports the current evidence showing independent paths to cancer resistance in the blind mole rat (Azpurua and Seluanov, 2012). Future functional studies in mole rats are required to corroborate these observations.

Unique Features of the NMR

Although the NMR and DMR share a relatively recent common ancestor (~26 Mya), the NMR has several exceptional features and is considered a most unusual mammal. Accelerated gene evolution among lineages could indicate an association between genetic changes and the evolution of traits (Qiu et al., 2012). Analysis of nonsynonymous-to-synonymous substitution (Ka/Ks) ratios of 9,367 1:1 orthologs of ten mammalian species revealed that the NMR was significantly enriched for several GO categories, including the respiratory electron transport chain, cell redox homeostasis, and response to oxidative stress (Figure 4A; Table S6). To test whether genes in the rapidly evolving GO categories were under positive selection, we used a branch likelihood ratio test to identify positively selected genes in the NMR and DMR lineages (Table S3).

Body Temperature Regulation

Like other mammals, the DMR tightly controls its body temperature (stable at 35°C). The NMR, in contrast, lacks an insulating layer of fur and cannot maintain thermal homeostasis if it is housed on its own away from the warm confines of its humid burrows (Buffenstein and Yahav, 1991). Over the normal range of ambient temperatures encountered in their natural milieu, they are able to maintain body temperature and employ endothermic mechanisms to fine-tune body temperature. To accomplish this, it employs nonshivering thermogenesis, using large pads of brown adipose tissue interspersed between muscle (Hislip and Buffenstein, 1994). Thermogenin (uncoupling protein 1 [UCP1]) is the major protein used in this kind of heat generation. The NMR UCP1 harbors amino acid changes at the site regulated by fatty acids and nucleotides (Kim et al., 2011), whereas we find that the DMR sequence is typical of other mammals (Figure S3A). Thus, the altered UCP1 is an adaptation of the NMR rather than of mole rats in general, and it is strongly linked to ineffective thermogenesis.

Melatonin is a regulator of circadian rhythm and body temperature (Cagnacci et al., 1992). In rodents, there are two high-affinity receptors for melatonin: melatonin receptor 1a (MTNR1a) and MTNR1b. We found that the NMR is the only known “natural” MTNR1a and MTNR1b knockout animal (Figures S3B and S3C). Both the DMR and NMR lost MTNR1b, although the inactivating mutations are located in different positions (Figure S3B). MTNR1b is also a pseudogene in the distantly related Siberian hamster (Phodopus sungorus) (Prendergast, 2010) and the Syrian hamster (Mesocricetus auratus; GenBank accession number AY145849). However, MTNR1a alone is sufficient to maintain photoperiod and melatonin responses in the Siberian hamster (Prendergast, 2010). Interestingly, MTNR1a is intact in the DMR but inactivated in the NMR (Figure S3C). The lack of cognate melatonin receptors could contribute to the inability of NMR to adequately respond to fluctuating temperature.

Pain Insensitivity

C-fibers are small, unmyelinated axons associated with slow pain signaling in response to a range of external stimuli, which
can be thermal, mechanical, or chemical. The NMR has fewer C-fibers than other rodents, including the DMR (St John Smith et al., 2012), and the C-fibers of the NMR’s skin, eyes, and nose do not produce the pain-relaying neuropeptides substance P (SP) and calcitonin gene-related peptide (CGRP) (Park et al., 2003). It should be noted that there is not a complete loss of expression, as low levels of these neuropeptides can be found in internal organs (Park et al., 2003). It is currently not known how the expression of SP and CGRP is repressed in the sensory neurons of the NMR, but it was shown that NMRs receiving gene therapy with the SP-encoding preprotachykinin gene (TAC1) responded to pain induced by peripheral inflammation (Park et al., 2012), and the C-fibers of the NMR’s skin, eyes, and nose do not produce the pain-relaying neuropeptides substance P (SP) and calcitonin gene-related peptide (CGRP) (Park et al., 2003). It should be noted that there is not a complete loss of expression, as low levels of these neuropeptides can be found in internal organs (Park et al., 2003). It is currently not known how the expression of SP and CGRP is repressed in the sensory neurons of the NMR, but it was shown that NMRs receiving gene therapy with the SP-encoding preprotachykinin gene (TAC1) responded to pain induced by peripheral inflammation (Park et al., 2008). These observations suggest that the SP-encoding gene itself is altered in the NMR. The NMR TAC1 gene harbors an 8 bp deletion in its proximal promoter (Kim et al., 2011). The presence of the 8 bp region in the DMR (Figure S3D) suggests that this region is associated with pain insensitivity. In addition to unique changes in the NMR genes encoding SP and CGRP, genes associated with neurotransmission of pain were under positive selection in the NMR. This included the NMDA receptor NR2B (GRIN2B), the TRP channel TRPC5, and proenkephalin (PENK) (Table S3). β-actin May Mediate Enhanced Oxidative Stress Resistance in the NMR Actins are highly conserved proteins involved in cell structure, motility, and integrity. The vertebrate actin family contains six genes, of which only the cytoplasmic actins, β-actin (ACTB) and γ-actin (ACTG1), are ubiquitously expressed (Herman, 1993). The β-actin protein is highly conserved throughout evolution (Vandekerckhove and Weber, 1978). During oxidative stress, cysteine residues of actins can be oxidized, which is associated with depolymerization and altered regulatory protein interactions (Terman and Kashina, 2013). Increased ROS levels and actin overoxidation are symptomatic of senescence and diseases such as Alzheimer’s (Aksenov et al., 2001). We observed that β-actin (ACTB) is under positive selection in the NMR (Table S3). RNA-seq and synteny analysis (Figure 4B) confirmed the identity of NMR ACTB. We found that both Cys272 and Ala230 of β-actin are converted to serine in the NMR (Figures 4C and 4D). Cys272 is highly redox sensitive and may contribute as a “redox sensor” (Lassing et al., 2007). These observations suggest that NMR β-actin is more resistant to oxidation and may contribute to the longevity of the NMR, which can live at least 10 years longer than the DMR despite its exposure to high ROS levels and lower body mass (Lewis et al., 2013). The potential involvement of ACTB Cys272 in senescence and disease can now be...
evaluated more extensively in the NMR, an animal model that lacks this residue.

**28S rRNA Processing in Evolutionary and Geographically Distant Hystricognath Rodents**

We discovered that NMR rRNA did not display the typical banding pattern, i.e., 28S at ~4.4 kb and 18S at ~1.8 kb, during denaturing gel electrophoresis (Figures 5A and 5B). In contrast, the DMR had the standard pattern (Figure 5B). The unusual NMR pattern occurred in every tissue tested (ovary, kidney, liver, and brain), from separate animals and at any age tested (from 1 to 23 years old). A similar phenomenon, in which 28S rRNA is split into two subunits held together as a single 28S rRNA molecule by hydrogen bonding under native conditions, has been described in insects and plants (Winnebeck et al., 2010). The only vertebrates reported to produce shorter 28S rRNA are South America’s tuco-tucos (Ctenomys) and the degu (Octodontomys gliroides) (Melen et al., 1999). We found that the “break” region is in the D6 domain of 28S. This NMR region corresponds to a cryptic GC- and simple repeat-rich intron in the Talas tuco-tuco (Ctenomys talarnum) (Melen et al., 1999), and there is also a high degree of sequence conservation between these species (Figure 5C). In tuco-tuco, an unknown site within this cryptic intron results in “breakage” of 28S rRNA molecules (Melen et al., 1999). The cryptic intron in the NMR may explain the banding pattern observed (Figure 5D). Two other hystricognaths, the South American guinea pig and the African DMR, do not harbor the cryptic intron (Figure S4). The data suggest that the cryptic intron and the resulting “broken” 28S rRNA were present in a common ancestor prior to the cross-Atlantic migration of small African hystricognath rodents to South America ~41 Mya (Antoine et al., 2012). This proposition is consistent with the predicted ancestor of the NMR, DMR, and guinea pig (Figure 1C). Following submission of our manuscript, Azpurua et al. (2013) also found the unusual 28S processing in NMR tissues and proposed that the unique NMR 28S may result in improved protein synthesis fidelity and a more stable proteome. The NMR thus holds promise as a model organism for investigating the mechanism and function of “hidden breaks” in rRNAs in a laboratory setting.

**Conclusions**

We performed genome sequencing and de novo assembly of two related subterranean rodents, the DMR and NMR. The transcriptomes of five subterranean rodents were also sequenced. African mole rats share a unique ecology and physiology. These animals are also remarkably long-lived for their size and are characterized by similar traits of cancer resistance, maintenance of neuronal integrity, altered insulin structure, and elevated brain globin. However, in many other traits, the use of the DMR genome pinpointed features responsible for the truly unusual characteristics of the NMR, including unusual thermogenesis, an aberrant melatonin system, pain insensitivity, and processing of rRNA. The genomes and transcriptomes can be further mined to provide insights into the fascinating biology of these animals.

**EXPERIMENTAL PROCEDURES**

See Supplemental Experimental Procedures for additional protocols.

**Animals**

A breeding colony of DMRs (Fukomys damarensis) was housed at the University of Illinois at Chicago. The DMR was known as Cryptomys damarensis prior to a recent subclassification into a new genus, Fukomys (Kock et al., 2008). The animals were sacrificed and DNA and RNA were isolated for subsequent sequencing. The genome sequenced was that of a 5-year-old male DMR. Liver and brain transcriptomes were obtained by sequencing individuals from the same colony. Animal experiments were approved by the University of Illinois at Chicago Institutional Animal Care and Use Committee. Species range
maps were obtained and adapted from the World Wildlife Fund’s WildFinder database (http://www.worldwildlife.org/pages/wildfinder).

DMR Genome Sequencing and Assembly

We employed a whole-genome shotgun strategy and next-generation sequencing technologies, using the Illumina HiSeq 2000 as the platform, to sequence the genome of a captive male DMR. We constructed 16 paired-end (PE) libraries with insert sizes of 250 bp, 500 bp, 800 bp, 2 kbp, 5 kbp, 10 kbp, and 20 kbp. In total, 229 Gbp (or 76×) high-quality data, including 151 Gbp (or 50×) short insert size libraries, were generated (Table 1). The genome was de novo assembled by SOAPdenovo (Li et al., 2010a). Then, 151 Gbp (or 50×) data from short-insert-size libraries (250–800 bp) were split into 63-mers and contigs, with unambiguous connections in de Bruijn graphs retained. All reads were aligned onto contigs for scaffold building using PE information. We used k-mer analysis (Li et al., 2010b) to estimate the genome size of the DMR. In this study, K was 17, K_num was 61,533,145,821, and K_depth was 22.5. Therefore, the DMR genome size was estimated to be 2.73 Gbp (Figure S1A).

Assembly of the NMR Genome

To develop an improved assembly of the NMR genome, we used lastz (Harris, 2007), with the parameter “=M=60=y=9000 =T=2=format=axt=”, to align previously generated genome sequences (Kim et al., 2011) and additional sequence data to a recently released genome assembly from the Broad Institute (GenBank accession number AHKG00000000). ChainNet (Kent et al., 2003) was used to combine traditional alignments into larger structures. After the primary alignment was obtained, PE reads with insert sizes from 2 to 20 kbp were mapped to the “newly formed” genome. A new NMR assembly with a scaffold N50 of 21.3 Mb was generated (Table 1).

Whole-Genome Heterozygosity Analysis

We aligned all high-quality, short-insert-size reads to the genome assembly using BWA (Li and Durbin, 2009). Since the alignment results were stored in BAM/SAM format, we selected SAMtools, which is based on the Bayesian model, for variation analysis (Li et al., 2009). After sorting alignments by the left-most coordinates and removing potential PCR duplicates, we used SAMtools mpileup to call SNPs and short InDels. We rejected SNPs and InDels within reads with a depth that was either much lower or much higher than expected, since a large copy-number variation might lead to miscalling of SNPs. The sequencing depth ranged from 4 to 100 and the upper limit was approximately triple the sequencing depth. SNP miscalling due to alignment around short InDels and low-quality sequences was removed. We applied samtools.pl var-filter, which can be found in the SAMtools package, as the filter tool with parameters -G 20 -g 20 -d 4 -D 100 -S 20 -i 20 -N 5 -s 5 -W 5 -N 1.

Repeat Annotation

RepeatProteinMask and RepeatMasker (Tarailo-Graovac and Chen, 2009) were used to identify and classify transposable elements by aligning the DMR genome sequences against a library of known repeats, Repbase, with default parameters. The repeats obtained were combined together to form a list of nonredundant repeats of DMR. The same approach was used to identify repeats in related mammals, including the NMR.

ACCESS NUMBER

The DMR whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under accession code ALU00000000. The version described in this paper is the first version, ALU001000000. All short-read data have been deposited in the Short Read Archive under accession code SRA099445. Raw sequencing data of the transcriptome have been deposited in the Gene Expression Omnibus under accession code GSE50726.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.07.030.

AUTHOR CONTRIBUTIONS

V.N.G. coordinated the study. T.J.P. collected and prepared DMR, NMR, and rat samples. R.S. collected and prepared FA, FD, and coruro samples. A.A.T., M.V.G., S.H.Y., and R.T.B. prepared and analyzed biological samples. X.F., Z.H., Z.X., Y.Z., X.Y., D.F., and L.Y. performed genome sequencing, assembly, and annotation. X.F., X.Z., G.Z., and J.W. supervised genome sequencing, assembly, and annotation. X.F., I.S., Z.H., M.V.G., Z.X., Y.Z., A.V.L., D.F., Y.Y., S.M., L.Y., S.-G.L., E.B.K., R.B., X.Z., A.K., T.J.P., and V.N.G. performed genome and transcriptome analyses. All authors contributed to data interpretation. I.S. and V.N.G. wrote the paper with significant contributions from X.F., Z.X., and Z.H., and input from all authors.

ACKNOWLEDGMENTS

This work was supported by the NIH (AG047745, AG047200, AG038004, AG021518, and GM061603), the WCU Program (R31-2008-000-10010-0), and the National Natural Science Foundation of China (31171190). NMR and DMR photos were provided by Joel Sartore (http://www.joelsartore.com).

Received: October 3, 2013
Revised: May 11, 2014
Accepted: July 17, 2014
Published: August 28, 2014

REFERENCES

Aksenov, M.Y., Aksenova, M.V., Butterfield, D.A., Geddes, J.W., and Markevsky, W.R. (2001). Protein oxidation in the brain in Alzheimer’s disease. Neurosience 103, 373–383.
Andziak, B., O’Connor, T.P., Qi, W., DeWaal, E.M., Pierce, A., Chaudhuri, A.R., Van Remmen, H., and Buffenstein, R. (2006). High oxidative damage levels in the longest-living rodent, the naked mole-rat. Aging Cell 5, 483–471.
Antoine, P.O., Marivaux, L., Croft, D.A., Billet, G., Ganerød, M., Jaramillo, C., Martin, T., Orliac, M.J., Tejada, J., Altamirano, A.J., et al. (2012). Middle Eocene rodents from Peruvian Amazonia reveal the pattern and timing of cavemorph origins and biogeography. Proc. Biol. Sci. 279, 1319–1326.
Arivi, A., Gerlaich, F., Joel, A., Reuss, S., Burnmester, T., Nevo, E., and Hankeln, T. (2010). Neuroglobin, cytoglobin, and myoglobin contribute to hypoxia adaptation of the subterranean mole rat Spalax. Proc. Natl. Acad. Sci. USA 107, 21570–21575.
Azpurua, J., and Seluanov, A. (2012). Long-lived cancer-resistant rodents as new model species for cancer research. Front. Genet. 3, 319.
Azpurua, J., Ke, Z., Chen, I.X., Zhang, Q., Ermoilenko, D.N., Zhang, Z.D., Gorbunova, V., and Seluanov, A. (2013). Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. Proc. Natl. Acad. Sci. USA 110, 17350–17355.
Begall, S., Burda, H., and Schleich, C.E. (2007). Subterranean rodents: news from underground. In Subterranean Rodents—revisited. In Subterranean Rodents: News from Under-
Cagnacci, A., Elliott, J.A., and Yen, S.S. (1992). Melatonin: a major regulator of the circadian rhythm of core temperature in humans. J. Clin. Endocrinol. Metab. 75, 447–452.

Campisi, J., Andersen, J.K., Kapahi, P., and Melov, S. (2011). Cellular senescence: a link between cancer and age-related degenerative disease? Semin. Cancer Biol. 21, 354–359.

Chan, S.J., Episkopou, V., Zeitlin, S., Karathanasis, S.K., MacKrell, A., Steiner, D.F., and Efstratiadis, A. (1984). Guinea pig preproinsulin gene: an evolutionary compromise? Proc. Natl. Acad. Sci. USA 81, 5046–5050.

Dammann, P., and Burda, H. (2007). Senescence patterns in African mole-rats (Bathyergidae, Rodentia). In Subterranean Rodents, S. Begall, H. Burda, and C.E. Schleich, eds. (Berlin, Germany: Springer), pp. 251–262.

de Magalhães, J.P. (2013). How ageing processes influence cancer. Nat. Rev. Cancer 13, 357–365.

Dolbow, D.P., Di Costanzo, L., Gennadios, H.A., and Christianson, D.W. (2008). Evolution of the arginase fold and functional diversity. Cell. Mol. Life Sci. 65, 2039–2055.

Duvillié, B., Cordonnier, N., Delport, L., Dandoy-Dron, F., Itier, J.M., Monthoux, E., Jami, J., Joshi, R.L., and Bucchi, D. (1997). Phenotypic alterations in insulin-deficient mutant mice. Proc. Natl. Acad. Sci. USA 94, 5137–5140.

Edrey, Y.H., Park, T.J., Kang, H., Biney, A., and Buffenstein, R. (2011). Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. Exp. Gerontol. 46, 116–123.

Fang, X., Nevo, E., Han, L., Levinson, E.Y., Zhao, J., Avivi, A., Larkin, D., Jiang, X., Feranchuk, S., Zhu, Y., et al. (2014). Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. Nat. Commun 5, 3966.

Faulkner, C.G., Verheyen, E., Verheyen, V., Jarvis, J.U., and Bennett, N.C. (2004). Phylogenographical patterns of genetic divergence and speciation in African mole-rats (Family: Bathyergidae). Mol. Ecol. 13, 619–628.

Fiermonte, G., Dolce, V., David, L., Santorelli, F.M., Dionisi-Vici, C., Palmtier, F., and Walker, J.E. (2003). The mitochondrial ornithine transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. J. Biol. Chem. 278, 32778–32783.

Gorbunova, V., Hine, C., Tian, X., Ablaeva, J., Gudkov, A.V., Nevo, E., and Seluanov, A. (2012). Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. Proc. Natl. Acad. Sci. USA 109, 19392–19396.

Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. PhD thesis (University Park, Pennsylvania: Pennsylvania State University).

Herman, I.M. (1993). Actin isoforms. Curr. Opin. Cell Biol. 5, 48–55.

Heth, G., and Todrank, J. (2007). Using odors underground. In Subterranean Rodents: News from Underground. Subterranean Rodents, S. Begall, H. Burda, and C.E. Schleich, eds. (Berlin, Germany: Springer), pp. 85–96.

Hislop, M.S., and Buffenstein, R. (1994). Noradrenaline induces nonshivering thermogenesis in both the naked mole-rat (Heterocephalus glaber) and the thermoregulation. J. Therm. Biol.

Horuk, R., Blundell, T.L., Lazarus, N.R., Neville, R.W., Stone, D., and Wollmer, M. (1981). The mitochondrial ornithine transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. J. Biol. Chem. 278, 32778–32783.

Kasaikina, M.V., Lobanov, A.V., Malinouski, M.Y., Lee, B.C., Seravalli, J., Finney, L., Vogt, S., Park, T.J., et al. (2011). Genome sequencing by concerted necrotic cell death mechanism. Proc. Natl. Acad. Sci. USA 108, 17005–17014.

King, G.L., Kahn, C.R., and Heidt, C.H. (1983). Sharing of biological effect and receptors between guinea pig insulin and platelet-derived growth factor. Proc. Natl. Acad. Sci. USA 80, 1308–1312.

Kock, D., Ingram, C., Frabotta, L., Honeycutt, R., and Burda, H. (2006). On the nomenclature of Bathyergidae and Fukomys n. gen. Mammalia: Rodentia). Zootaxa 1142, 51–55.

Kott, O., Sumbera, R., and Nemec, P. (2010). Light perception in two strictly subterranean rodents: life in the dark or blue? PLoS ONE 5, e11810.

Kramer, B., and Buffenstein, R. (2004). The pancreas of the naked mole-rat (Heterocephalus glaber); an ultrastructural and immunocytochemical study of the endocrine component of thermoneutral and cold acclimated animals. Gen. Comp. Endocrinol. 139, 206–214.

Kraus, D.W., and Colacino, J.M. (1986). Extended oxygen delivery from the nerve hemoglobin of Telanta alternata (Bivalvia). Science 232, 90–92.

Lassing, I., Schmitzberger, F., Björnstedt, M., Holmgren, A., Nordlund, P., Schutt, C.E., and Lindberg, U. (2007). Molecular and structural basis for redox regulation of beta-actin. J. Mol. Biol. 370, 331–348.

LaVinka, P.C., Brand, A., Landau, V.J., Wirtshaffer, D., and Park, T.J. (2009). Extreme tolerance to ammonia fumes in African naked mole-rats: animals that naturally lack neuroneptides from trigeminal chemosensory nerve fibers. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 195, 419–427.

Lavulo, L.T., Sossong, T.M., Jr., Brigham-Burke, M.R., Doyle, M.L., Cox, J.D., Christianson, D.W., and Ash, D.E. (2001). Subunit-subunit interactions in trimeric arginase. Generation of active monomers by mutation of a single amino acid. J. Biol. Chem. 276, 14242–14248.

Lewis, K.N., Andziak, B., Yang, T., and Buffenstein, R. (2013). The naked mole-rat response to oxidative stress: just deal with it. Antioxid. Redox Signal. 19, 1388–1399.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMTools. Bioinformatics 25, 2078–2079.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., et al. (2010a). De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265–272.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., et al. (2010b). The sequence and de novo assembly of the giant panda genome. Nature 463, 311–317.

Liang, L., Guo, W.H., Esquilliano, D.R., Asai, M., Rodriguez, S., Giraud, J., Kushiner, J.A., White, M.F., and Lopez, M.F. (2010a). Insulin-like growth factor 2 and the insulin receptor, but not insulin, regulate fetal hepatic glycogen synthesis. Endocrinology 151, 741–747.

Liang, S., Mele, J., Wu, Y., Buffenstein, R., and Hornsby, P.J. (2010b). Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). Aging Cell 9, 626–635.

Liu, M., Hodish, I., Haataja, L., Lara-Lemos, R., Rajpai, G., Wright, J., and Arvan, P. (2010). Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth. Trends Endocrinol. Metab. 21, 652–659.

Loh, S.H., Francescut, L., Lingor, P., Bähr, M., and Nicotera, P. (2008). Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. Cell Death Differ. 15, 283–298.

Lui, J.C., and Baron, J. (2013). Evidence that Igf2 down-regulation in postnatal tissues and up-regulation in malignancies is driven by transcription factor E2f3. Cancer Res. 73, 66–72.

Manov, I., Hirsh, M., Iancu, T.C., Malik, A., Sotnichenko, N., Band, M., Avivi, A., and Shams, I. (2013). Pronounced cancer resistance in a subterranean rodent, the blind mole-rat, Spalax: in vivo and in vitro evidence. BMC Biol. 11, 91.
Melen, G.J., Pesce, C.G., Rossi, M.S., and Kombilhit, A.R. (1999). Novel processing in a mammalian nuclear 28S pre-rRNA: tissue-specific elimination of an ‘intrin’ bearing a hidden break site. EMBO J. 18, 3107–3118.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodblad, A., Eizirik, E., Simão, T.L., Stadler, T., et al. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334, 521–524.

Nasser, N.J., Avivi, A., Shafat, I., Edovitsky, E., Zcharia, E., Ilan, N., Vladovsky, I., and Nevo, E. (2009). Alternatively spliced Spalax heparanase inhibits extraacellular matrix degradation, tumor growth, and metastasis. Proc. Natl. Acad. Sci. USA 106, 2253–2258.

Nemec, P., Cvekova, P., Benada, O., Wielkopolska, E., Oklikowicz, S., Turlejski, K., Burda, H., Bennett, N.C., and Peichl, L. (2009). Expression of hemoglobin in rodent neurons. J. Cereb. Blood Flow Metab. 29, 585–595.

Ono, K., Li, Y., Rabinovitch, P.S., and Martin, G.M. (2005). Age-related cataract progression in five mouse strains. Invest. Ophthalmol. Vis. Sci. 46, 3247–3255.

Opazo, J.C., Soto-Gamboa, M., and Bozinovic, F. (2004). Blood glucose concentration in caviomorph rodents. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 137, 57–64.

Park, T.J., Lu, Y., Jüttner, R., Smith, E.S., Hu, J., Brand, A., Wetzel, C., Milenikovic, N., Erdmann, B., Heppenstall, P.A., et al. (2008). Selective inflammatory pain insensitivity in the African naked mole-rat (Heterocephalus glaber). PLoS Biol. 6, e13.

Prendergast, B.J. (2010). MT1 melatonin receptors mediate somatic, behavioral, and reproductive neuroendocrine responses to photoperiod and melatonin in Siberian hamsters (Phodopus sungorus). Endocrinology 151, 714–721.

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yak genome and adaptation to life at high altitude. Nat. Genet. 44, 946–949.

Reddy, V.N., Giblin, F.J., Lin, L.R., Dang, L., Unakar, N.J., Musch, D.C., Boyle, D.L., Takemoto, L.J., Ho, Y.S., Knoernschild, T., et al. (2001). Glutathione peroxidase-1 deficiency leads to increased nuclear light scattering, membrane damage, and cataract formation in gene-knockout mice. Invest. Ophthalmol. Vis. Sci. 42, 3247–3255.

Rosenfeld, M.G., Amara, S.G., Roos, B.A., Ong, E.S., and Evans, R.M. (1981). Altered expression of the calcitonin gene associated with RNA polymorphism. Nature 290, 63–65.

Sabio, G., Mora, A., Rangel, M.A., Quesada, A., Marcos, C.F., Alonso, J.C., Sabio, G., Mora, A., Rangel, M.A., Quesada, A., Marcos, C.F., Alonso, J.C., and Bozinovic, N., Erdmann, B., Heppenstall, P.A., et al. (2008). Selective inflammatory pain insensitivity in the African naked mole-rat (Heterocephalus glaber). PLoS Biol. 6, e13.

Seluanov, A., Hine, C., Bozzella, M., Hall, A., Sasahara, T.H., Ribeiro, A.A., Catania, K.C., Presgraves, D.C., and Gorbunova, V. (2008). Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. Aging Cell 7, 813–823.

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

Shams, I., Malik, A., Manov, I., Joel, A., Band, M., and Avivi, A. (2013). Transcription pattern of p53-targeted DNA repair genes in the hypoxia-tolerant subterranean mole rat Spalax. J. Mol. Biol. 425, 1111–1118.

Simarro, M., Giannattasio, G., De la Fuente, M.A., Benarafa, C., Subramanian, K.K., Ishizawar, R., Balestrieri, B., Andersson, E.M., Luo, H.R., Ordunia, A., et al. (2010). Fas-activated serine/threonine phosphoprotein promotes immune-mediated pulmonary inflammation. J. Immunol. 184, 5325–5332.

Smith, E.S., Omerbasir, D., Lechner, S.G., Anirudhan, G., Lapatinska, L., and Lewin, G.R. (2011). The molecular basis of acid insensitivity in the African naked mole-rat. Science 334, 1557–1560.

St John Smith, E., Purfürst, B., Grigoryan, T., Park, T.J., Bennett, N.C., and Lewin, G.R. (2012). Specific paucity of unmethylated C-fibers in cutaneous peripheral nerves of the African naked-mole rat: comparative analysis using six species of Bathyrhagidae. J. Comp. Neurol. 519, 2785–2803.

T каролина graoac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics. Chapter 4, Unit 4.10.

Terman, J.R., and Kashina, A. (2013). Post-translational modification and regulation of actin. Curr. Opin. Cell Biol. 25, 30–38.

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

van Oers, C.C., Adema, G.J., Zandberg, H., Moen, T.C., and Baas, P.D. (1994). Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol. Cell. Biol. 14, 951–960.

Vandekerckhove, J., and Weber, K. (1978). At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. 126, 783–802.

Wang, Z., and Burge, C.B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14, 802–813.

Wang, H., Gao, J., Sun, X., Martinez-Wittingham, F.J., Li, L., Varadaraj, K., Farrell, M., Reddy, V.N., White, T.W., and Mathias, R.T. (2009). The effects of GPX-1 knockout on membrane transport and intracellular homeostasis in the lens. J. Membr. Biol. 227, 25–37.

Winnebeck, E.C., Millar, C.D., and Warman, G.R. (2010). Why does insect RNA look degraded? J. Insect Sci. 10, 159.

Wolf, N., Penn, P., Pendergrass, W., Van Remmen, H., Bartke, A., Rabinovich, P., and Martin, G.M. (2005). Age-related cataract progression in five mouse models for anti-oxidant protection or hormonal influence. Exp. Eye Res. 81, 276–285.

Zhao, H., Ru, B., Teeling, E.C., Faulkes, C.G., Zhang, S., and Rossiter, S.J. (2009). Rhodopsin molecular evolution in mammals inhabiting low light environments. PLoS ONE 4, e8326.
Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes

Xiaodong Fang\textsuperscript{1,2,10}, Inge Seim\textsuperscript{3,4,10}, Zhiyong Huang\textsuperscript{1}, Maxim V. Gerashchenko\textsuperscript{3}, Zhiqiang Xiong\textsuperscript{1}, Anton A. Turanov\textsuperscript{3}, Yabing Zhu\textsuperscript{1}, Alexei V. Lobanov\textsuperscript{3}, Dingding Fan\textsuperscript{1}, Sun Hee Yim\textsuperscript{3}, Xiaoming Yao\textsuperscript{1}, Siming Ma\textsuperscript{3}, Lan Yang\textsuperscript{1}, Sang-Goo Lee\textsuperscript{4}, Eun Bae Kim\textsuperscript{4}, Roderick T. Bronson\textsuperscript{5}, Radim Šumbera\textsuperscript{6}, Rochelle Buffenstein\textsuperscript{7}, Xin Zhou\textsuperscript{1}, Anders Krogh\textsuperscript{2}, Thomas J. Park\textsuperscript{8}, Guojie Zhang\textsuperscript{1,2}, Jun Wang\textsuperscript{1,2,9,*}, Vadim N. Gladyshev\textsuperscript{3,4,*}

\textsuperscript{1}BGI-Shenzhen, Shenzhen, 518083, China
\textsuperscript{2}Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark
\textsuperscript{3}Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, 02115, USA
\textsuperscript{4}Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, South Korea
\textsuperscript{5}Rodent Histopathology Laboratory, Harvard Medical School, Boston, MA 02115, USA
\textsuperscript{6}University of South Bohemia, Faculty of Science, Ceske Budejovice, 37005, Czech Republic
\textsuperscript{7}Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, TX 78245, USA
\textsuperscript{8}Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA
\textsuperscript{9}King Abdulaziz University, Jeddah, 21441, Saudi Arabia
\textsuperscript{10}Co-first authors

*Correspondence: wangj@genomics.org.cn (J.W.), vgladyshev@rics.bwh.harvard.edu (V.N.G)
Supplemental Information

Figure S1, Related to Figure 1. Features of the DMR genome

(A) DMR genome size estimation using k-mer analysis. The red curve refers to the k-mer frequency distribution, and the green curve refers to the cumulative distribution of k-mer frequency. More than 25% high depth kmer(>255) may due to repeats in the genome. An unexpected peak at depth of 50 may result from segmental duplication or other repetitive elements. (B) Sequencing depth distribution of the DMR genome. High quality short insert size reads were mapped to the associated genome with an average depth of 34, and approximately 91.64% of the genome was covered by more than 10 reads. The red curve denotes the proportion of the genome in a given sequencing depth, and the blue curve shows the cumulative coverage of the genome. (C) Summary of evidence for DMR gene models from three types of gene source (D: de novo prediction; H: homology-based; R: RNA-seq). (D) Phylogenetic tree of 10 mammalian species. The tree was constructed with PhyML under an ml+JTT+gamma model. The bootstrap values were calculated based on 1,000 replicates. (E) Distribution of lost and newly evolved gene families in mammalian lineages. Numbers designate the number of gene families that have expanded (green) and contracted (red) since the split from the common ancestor. The most recent common ancestor (MRCA) has 19,839 gene families.
Figure S2, Related to Figures 2 and 3. Morphological and molecular changes of subterranean rodents

(A) NMR eye (x10, insert at x60). NMRs have normal eye morphology with the exception of cataracts, which were identified in every specimen examined (n=4). (B) The gene encoding glutathione peroxidase 1 (GPX1) of NMR and DMR harbors a premature stop codon and results in a truncated protein. Sequence alignment of vertebrate GPx1 proteins (C-terminal regions) is shown, and the C-terminal amino acid residue of NMR and DMR is highlighted in red. (C) Expression pattern of DNA damage response and repair genes in subterranean and ‘aboveground’ (surface-dwelling) rodents. The selected genes include mutS homolog 3 (MSH3), high mobility group box 2 (HMGB2), NME/NM23 nucleoside diphosphate kinase 1 and 6 (NME1 and NME6), and ubiquitin specific peptidase 1 and 5 (USP1 and USP5). (D) Overview of inactivation events in FASTK of African mole rats. The exon structure of the human gene, which is identical to the mouse and rat ortholog, is shown. Boxes indicate exons and lines introns, while red arrows indicate inactivation events (insertions or deletions that change the frame, or stop point mutations resulting in premature termination of translation). (E) Sequence alignment of mammalian hyaluronan synthase 2 (HAS2). Amino acids 178 and 301 are highlighted in red. Identical residues are shaded in black.
Figure S3, Related to Figure 4. Unique molecular changes in the naked mole rat

(A) Amino acid changes in functionally important motifs of the thermogenesis protein UCP1 are restricted to the NMR. Sequence alignment of rodent and human UCP1 is shown. Identical residues are shaded in black and conserved motifs are shown in blue. Amino acid residues unique to the poikilothermic NMR are highlighted in red. (B) Melatonin receptor 1b is inactivated in the DMR and NMR. Sequence alignment of rodent and human MTNR1b nucleotide sequences is shown. Identical residues are shaded in black, and premature stop codons are highlighted in red. (C) Melatonin receptor 1a is intact in the DMR. Sequence alignment of DMR and other mammalian MTNR1a sequences is shown. Identical residues are shaded in black. (D) A deletion within the promoter of TAC1, which encodes substance P (SP) is unique to the NMR and is not present in the related African mole rat, the DMR. Identical residues are shaded in black and the region deleted in NMR is highlighted in red. (E) Schematic genomic organization of human CALCA and the derived calcitonin/CT and calcitonin gene-related peptide/CGRP exons. Exons are shown as boxes. Exons employed by CT and GCRP are shown in blue and orange, respectively, while exons 1-3 shared by both isoforms are shown in both colors. The splicing pattern and the peptide sequences of prepro-calcitonin and prepro-calcitonin gene-related peptide are shown at the top and bottom of the exons, respectively. (F) NMR-specific deletions within exon 4 of CALCA. Multiple sequence alignment of NMR, DMR, mouse, rat and human exon 4 is shown. Exons are highlighted in blue and polyadenylation sites are shown in red. Regions deleted in the NMR are highlighted in red. Stars indicate nucleotide positions that are shared among all sequences. A highly conserved region deleted in the NMR is shown in yellow.
| DMR | tucu-tuco | NMR | mouse | human | chimp | guinea pig |
|-----|-----------|-----|-------|-------|-------|------------|
| exon 5' D6 | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC | ACCTGGCCAGCACTGGGACAGAGGTGCGGTAAGTCTCACTCCTGCAGGAGC |
| exon 3' D6 | TCTGGTGC | TCTGGTGC | TCTGGTGC | TCTGGTGC | TCTGGTGC | TCTGGTGC |
**Figure S4, Related to Figure 5. Multiple sequence alignment of the mammalian D6 region of 28S ribosomal RNA**

A highly conserved region of the mammalian 28S rRNA D6 domain is shown in black. Divergent regions of the D6 domain, exon 5' D6 and exon 3' D6, are shown in blue, and the cryptic intron is shown in orange. Dashes (−) indicate gaps, stars (*) indicate complete conservation and dots (.) indicate similarity. Repeat elements are indicated in bold and underlined.
Table S1, Related to Figure 1. Significantly expanded and contracted gene families in African mole rat lineages
(provided in a separate file).

Table S2, Related to Figure 1. Pseudogenes and lost genes in DMR, and pseudogenes common to the DMR and NMR
(provided in a separate file).

Table S3, Related to Figure 4. Genes under positive selection in the ancestral lineage of DMR/NMR, the DMR and the NMR, and associated GO categories
(provided in a separate file).

Table S4, Related to Figures 2 and 3. RPKM values and GO enrichment of genes differentially expressed in the liver of subterranean rodents
(provided in a separate file).

Table S5, Related to Figure 2. RPKM values and GO enrichment of genes differentially expressed in the brain of subterranean rodents
(provided in a separate file).

Table S6, Related to Figures 1 and 4. Lineage-specific accelerated GO categories of DMR and NMR (P ≤ 0.05)
(provided in a separate file).
SUPPLEMENTAL EXPERIMENTAL PROCEDURES

DMR and NMR genome features and evolution
Prediction of protein-coding genes
To predict genes in the DMR genome, we used both homology-based and de novo methods, and in addition, RNA-seq data were incorporated. For the homology-based prediction, human, mouse and rat proteins were downloaded from Ensembl (release 64) and mapped onto the genome using tBLASTn (Kent, 2002). Then, homologous genome sequences were aligned against the matching proteins using GeneWise (Birney et al., 2004) to define gene models. For de novo prediction, Augustus (Stanke, 2003) and GENSCAN (Salamov and Solovyev, 2000) were employed to predict coding genes, using appropriate parameters. RNA-seq data were mapped to genome using TopHat (Trapnell et al., 2009), and transcriptome-based gene structures were obtained by cufflinks (http://cufflinks.cbcb.umd.edu). Finally, homology-based, de novo derived and transcript gene sets were merged to form a comprehensive and non-redundant reference gene set using GLEAN (http://sourceforge.net/projects/glean-gene), removing all genes with sequences less than 50 amino acid as well as those that only had the de novo support. We obtained a reference gene set that contained 22,179 DMR genes, and the parameter of DMR genes is comparable to other published mammals, including the related naked mole rat (Kim et al., 2011) (Table 1).

Functional annotation of genes
Gene functions were assigned according to the best match of the alignments using BLASTp against SWISS-PROT databases (Bairoch and Apweiler, 1997). Gene Ontology IDs for each gene were obtained from the corresponding InterPro entry (Ashburner et al., 2000; Mulder and Apweiler, 2007). All genes were aligned against KEGG proteins, and the pathway in which the gene might be involved was derived from the matching genes in KEGG (Kanehisa and Goto, 2000).

Phylogenetic analysis
A gene family is a group of similar genes descended from a single gene in the last common ancestor of the targeted species. We used the TreeFam methodology (Li et al., 2006) to define gene families in 10 mammalian genomes (human, rhesus macaque, rabbit, mouse, rat, Chinese hamster, dog, guinea pig, DMR and NMR). We carried out the same pipeline and parameters that we used in our previously published work (Kim et al., 2011; Li et al., 2010a). In total, we obtained 19,839 gene families of which 6,133 are single-copy orthologous families. CDS sequences from each single-copy family were aligned guided by MUSCLE (Edgar, 2004) alignments of protein sequences and concatenated to one super gene for each species. Next, RAxML (Stamatakis et al., 2005) was applied to build phylogenetic trees under JTT+gamma model. 1,000 bootstrap replicates were employed to assess the branch reliability in RAxML. Divergence time was estimated by PAML MCMCTree (Yang and Rannala, 2006) using 4d site sequences of 5,838 single copy genes (Figure S1D).

Gene family expansion and contraction
To examine expansion and contraction of gene families in the DMR, we inferred the rate and direction of change of gene family size for DMR and a group of other mammals (human, rhesus, dog, NMR, rabbit, Chinese hamster, guinea pig, mouse and rat) using CAFÉ (De Bie et al., 2006), which is based on a stochastic birth-death model.
DMR gained and lost genes
To determine the orthologous relationship between DMR and human proteins, sequences of the human protein dataset were downloaded from Ensembl (release 64). The longest transcript was chosen to represent each gene with alternative splicing variants. We next subjected human and DMR proteins to BLASTp analysis with a similarity cutoff threshold E-value=1e-5. With the human protein set as a reference, we found the best hit for each DMR protein, with the criteria that more than 30% of the aligned sequence showed an identity above 30%. Reciprocal best-match pairs were defined as orthologs. Then, gene order information was used to filter out false positive orthologs caused by incorrect draft genome assembly and annotation. Orthologs not in gene synteny blocks were removed from further analysis. For example, in the case of 3 continuous genes in the human genome A, B and C, even if all three orthologs could be identified in both humans and DMR based on the cutoff threshold described above, but the B gene in the DMR genome was not located between A and C genes, it was filtered out, as it could be located in another scaffold or another place within the same scaffold. Using this method, we identified gene synteny relationships for human and DMR lineages.

Orthology information was obtained as described above. Since it showed synteny information at the protein level, it could be used to analyze gene-gain and -loss between human and DMR lineages. In the protein synteny blocks, if a human protein had no DMR ortholog, and excluding false positive predictions that could be caused by annotation or genome assembly (gap ≥ 5%), a protein could be defined as either being lost in the DMR lineage or gained in the human lineage. Using DMR as a reference to generate the orthology relationship, we applied this procedure to identify genes gained in the DMR lineage compared to the human lineage.

DMR pseudogenes
To detect homozygous pseudogenes in the DMR genome in silico, we first aligned all human genes (protein sequences downloaded from Ensembl (release 64) onto the DMR genome using BLASTp (with parameters -F F -e 1e-5). SOLAR (Sorting Out Local Alignment Results) (Yu et al., 2006) was used to conjoin the fragmental alignments for each gene. Best hit regions of each gene with 5 kb flanking sequence were cut down and re-aligned with their corresponding human orthologous protein sequences using GeneWise (with the parameters “-genesf -tforn –quiet”), which helped define the detailed exon-intron structure of each gene (Birney et al., 2004). Genes with frameshifts and premature stop codons, as reported by GeneWise, were considered candidate pseudogenes. We further carried out a series of filtering processes: (i) To avoid the frameshifts and premature stop codons incorrectly reported due to flaws in the GeneWise algorithm, we also aligned all human proteins to their corresponding loci in the human genome using GeneWise. Genes with frameshifts and premature stop codons in the human-to-human alignment were filtered out; (ii) Using the results of the human-to-human alignment from GeneWise, candidate pseudogenes with obvious splicing errors near their frameshifts and premature stop codons were filtered out; (iii) Candidate pseudogenes with a low number of reads covering their frameshift and premature stop codon sites were considered assembly errors. In addition, cases with a considerable number of reads resulting from different genotypes at these sites were treated as heterozygous. Cases of assembly error and heterozygosis were filtered out.
Assessment for visual perception pseudogenes

To identify visual perception genes lost or pseudogenised in the DMR genome, we first employed GeneWise (Birney et al., 2004). To examine the visual perception pseudogenes, we estimated the rate ratio (ω) of non-synonymous to synonymous substitutions (Yang, 2007). The coding sequences of DMR visual perception genes were aligned with their human, mouse, rat and dog homologs using MUSCLE (Edgar, 2004). We used two branch models: Model H0, which considers all branches with the same ω, was used to compute the average ω; Model H1, which examined the DMR and other branches with different ω, was used to compute the ω of the DMR branch (ω2) and other branches (ω1). Then, the likelihood ratio test (LRT) was used to compute the p-value that rejected the model H0. If the p-value was less than 0.05, we could reject the model H0 and accept model H1.

Among the 13 identified visual perception genes, two were absent in the rat (RP1LI and GRK7), but other genes were present in all examined mammals (human, mouse, rat and dog). Visual genes inactivated or lost in the DMR are indicated in Table S2.

Identification of proteins with unique amino acid changes in African mole rats

To associate predicted DMR and NMR peptide sequences with human RefSeq IDs, as presented in UCSC multiway alignments, NCBI tBLASTn v2.2.26+ of the BLAST+ suite (Sayers et al., 2012) with an E-value cut-off set at 1e-5 was employed. The best match (≥50% overall amino acid sequence identity along the entire sequence and spanning ≥75% of the length of the query sequence) was used to annotate the sequences. African mole rat proteins were aligned to orthologs from 38 vertebrate species using ClustalW v2.1 (Larkin et al., 2007). In addition to the Damaraland mole rat (Fukomys damarensis) and the and the naked mole rat (Heterocephalus glaber), the following organisms obtained via the UCSC multiway track (Miller et al., 2007) were examined (the full common name includes the text in brackets): the little brown bat (Myotis lucifugus), the megabat/fruit bat/flying fox Pteropus vampyrus, human (Homo sapiens), (common) chimpanzee (Pan troglodytes), (Sumatran) orangutan (Pongo pygmaeus abelii), rhesus monkey (Macaca mulatta), (hamadryas) baboon (Papio hamadryas), (common) marmoset (Callithrix jacchus), (Philippine) tarsier (Tarsier syrichta), (gray) mouse lemur (Microcebus murinus), bushbaby/small-eared greater galago (Otolemur garnettii), (Northern) tree shrew (Tupaia belangeri), (house) mouse (Mus musculus), (brown) rat (Rattus norvegicus), (Ord's) kangaroo rat (Dipodomys ordii), (domestic) guinea pig (Cavia porcellus), (thirteen-lined ground) squirrel (Spermophilus tridecemlineatus), (European) rabbit (Oryctolagus cuniculus), (American) pika (Ochotona princeps), alpaca (Vicugna pacos), cow (Bos taurus), horse (Equus caballus), cat (Felis catus), dog (Canis lupus familiaris), (European) hedgehog (Erinaceus europaeus), (common) shrew (Sorex araneus), (African) elephant (Loxodonta africana), (lesser hedgehog) tenrec (Echinops telfairi), (tammar) wallaby (Macropus eugenii), (gray short-tailed) opossum (Monodelphis domestica), platypus (Ornithorhynchus anatinus), (Western clawed) frog (Xenopus tropicalis), medaka/Japanese rice fish (Oryzias latipes), (three-spined) stickleback (Gasterosteus aculeatus), (tora)fugu (Takifugu rubripes), tetraodon/green spotted puffer (Tetraodon nigroviridis), zebrafish (Danio rerio), and sea lamprey (Petromyzon marinus).
In-house Perl scripts were used to parse the ClustalW output and identify unique amino acids. The Perl script scanned orthologous proteins for sites where types/groups of residues are shared by African mole rats only. The script groups residues into four groups: acidic (ED), basic (KHR), cysteine (C) and “other” (STYNQGAVLIFPMW). For example, it identifies cases where the NMR and DMR harbor E or D, whilst the other organisms contain basic, cysteine, or “other” residues at that particular site. Candidate proteins were manually examined for alignment quality and conservation of the region with amino acid changes.

The false positive rate of the detection of unique amino acids was estimated as follows: the error rate for SOAPdenovo assembly is assumed to be 1 nucleotide per 81,025 base pairs (Li et al., 2010b). Since only coding regions were used for amino acid conservation analysis, and the total predicted size of coding sequences (CDS) of the DMR and NMR is approximately 33 Mb, then 418 nucleotides could be considered as candidates for false positives. Potential tBLASTn misalignments are of minor importance since the parameters used were quite strict (E-value 1e-5). The unique amino acid method calls for a particular amino acid to be conserved among the tested vertebrate genomes, but differ in the genome-in-question (e.g. the DMR). In a test dataset, 3,287 amino acids were conserved among 36 genomes, out of 1,018,988 tested cases; therefore, only 0.32% of all sequences fit our search criteria. Taken together, with 418 candidates, one would expect a false positive rate of 1.33 per analysis; or in other words, one false positive per 315 unique amino acid residue candidates.

To expand on results pertaining to specific genes, we acquired genomic or transcriptome data of two additional African and two South-American hystricognath rodents: we generated brain RNA-seq data from the African mole rats Ansell's mole rat (Fukomys anselli, FA) and Mashona mole rat (Fukomys darlingi, FD). For comparison, we obtained the recently released genome sequence of the semi-subterranean degu (Octodon degus) (GenBank accession code AJSA00000000.1), as well as brain transcriptome data from the subterranean coruro (Spalacopus cyanus) of South-America. We acknowledge the efforts of the Broad Institute in assembling the genome sequence of the degu.

**Orthologous gene set for evolutionary analysis using PAML**

The orthologous sets of the 10 species from the phylogenetic tree were collected. For genes with alternative splice variants, the longest transcript was selected to represent the gene. Orthologs were identified using the method described in our previous study (Kim et al., 2011). MUSCLE (Edgar, 2004) was used for multi-protein sequence alignments of orthologs, using the human proteins as the reference. A series of filtering steps were performed on the alignments to assess alignment quality and conservation of exon-intron structure. Briefly, an alignment with more than 20% gaps or less than 50% identity between each protein and its human ortholog was identified as false positive and discarded. Next, the alignment was scanned exon-by-exon, and genes with more than 10% gaps or less than 60% identity, were manually checked against the genome sequences to validate whether these polymorphisms were real or caused by annotation errors. After these filtering processes, high quality alignments of protein sequences were retained and their alignments of associated codon sequences were used for further analysis. In total 9,367 1:1 ortholog alignments were obtained.
Identification of gene categories under accelerated evolution

The orthologous sets of the 10 mammals from the phylogenetic tree were used to identify proteins in each lineage that show accelerated evolution using a branch model implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) program. We aligned GO categories containing no fewer than 10 genes in the 10-species ortholog data set. PAML’s free-ratio model (Yang, 2007) was used to calculate different ω ratios (also known as Ka/Ks) across every branch in the phylogenetic tree. To investigate whether fast-evolving genes in the NMR or DMR lineage were enriched for specific biological processes, a binomial test (P ≤ 0.05) was used to identify GO categories with putatively accelerated non-synonymous divergence (increased dN/dS ratios) in either the DMR lineage or the NMR lineage.

Identification of genes under positive selection

To detect genes under positive selection in either the NMR or the DMR lineage, a series of evolutionary models were tested using PAML’s branch model (Zhao et al., 2010). Briefly, the average ω across the tree (ω0), ω of the branch tested (ω2) and ω of all other branches (ω1) were determined with the following parameter settings: Codonfreq=2, kappa=2.5, initial omega=0.2. A chi-square test were used to determine if ω2 was significantly higher than ω1 and ω0, which implies that these genes may be fast-changed or fast-evolved in the appointed branch. After obtaining potential positively selected genes (ω2 > ω1 and p value ≤ 0.05 adjusted by the FDR method for multiple testing) (Benjamini and Yekutieli, 2001), we performed a final manual check of the alignment to confirm our results.

Expression analysis

Survey of gene expression in the brain

To identify genes that could play a role in the adaptation to an underground environment, we compared the brain transcriptomes of five subterranean and three surface-dwelling (‘aboveground’) hystricognath rodents. Total RNA was isolated from brain tissue (frontal lobe region) of Fukomys damarensis (DMR, n=2), Fukomys anselli (FA, n=1), Fukomys darlingi (FD, n=1), and Spalacopus cyanus (coruro, n=1). All animals were from captive colonies and the experiments were approved by the respective Animal Care and Use Committees (the University of Illinois at Chicago in the case of DMR; the University of South Bohemia in the case of FA, FD and coruro).

RNA sequencing libraries were constructed using the Illumina mRNA-Seq Prep Kit as per the manufacturer’s instructions. Briefly, oligo(dT) magnetic beads were used to purify mRNA molecules. Next, mRNA was fragmented and randomly primed during the first strand synthesis by reverse transcription. Double-stranded cDNA fragments were then obtained by second-strand synthesis with DNA polymerase I, and the double stranded cDNA was subjected to end-repair by Klenow and T4 DNA polymerases, and finally A-tailed by Klenow lacking exonuclease activity. Ligation to Illumina Paired-End Sequencing adapters, size selection by gel electrophoresis and then PCR amplification completed the library preparation procedure. 200 bp paired-end libraries were sequenced using Illumina HiSeq 2000 (90 bp at each end).

Because no reference genomes are available for FA, FD and the coruro, the brain transcriptomes were assembled de novo. We also assembled paired-end reads from published RNA-seq data from the naked mole rat Heterocephalus glaber (NMR, n=2) (Kim et al., 2011) (NCBI GEO accession no. GSE30337) and the South-American
hystricognath rodents the domestic guinea pig *Cavia porcellus* (n=3) and the Brazilian guinea pig *Cavia aperea* (n=3), as well as a tame line of the brown rat (*Rattus norvegicus*; n=3) in the rodent family Muridae (Albert et al., 2012) (ArrayExpress accession no. E-MTAB-1249). With the exception of the NMR, all samples examined were males. Since many genes would be expected to change their expression between the surface-dwelling rodents the guinea pig (Rodentia, Hystricognathi) and the rat (Rodentia, Muridae) (Konopka et al., 2012), the present study included the rat as an outgroup in order to further narrow down the candidate driver genes for adaptation to a subterranean environment.

Reads were preprocessed to remove adapters and overrepresented sequences identified by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The software package Trimmomatic (Lohse et al., 2012) was used to trim Illumina reads falling below the established quality threshold (default settings) as well as trim adapter sequences and remove unpaired reads. Paired reads were concatenated, keeping only reads with ≥36 bp length (after merging and adapter trimming). Next, transcriptome assembly was completed using the short-read *de novo* assembler Trinity (Grabherr et al., 2011). Abundance estimates were obtained by running RSEM (RNA-seq by Expectation Maximization) separately for the left and right reads from each paired-end run (sample) (Haas et al., 2013; Li and Dewey, 2011). A database of mouse UniRef90 protein sequences (Wu et al., 2006) was interrogated using transcripts predicted by Trinity as a query (BLASTx E-value cut-off 1e-20) (Grabherr et al., 2011). Next, The RSEM output files (RSEM.genes.results) were parsed using a custom Perl script, retaining Trinity transcripts that aligned across at least 80% of the corresponding mouse protein sequence (Grabherr et al., 2011). We did not exclude any genes from the annotations during quantification (e.g. paralogs were kept), and did not modify the gene annotations or attempt to identify new genes using our data.

Differentially expressed genes were detected using the method described by Chen and colleagues (Chen et al., 2010), which is based on the Poisson distribution (Audic and Claverie, 1997) and normalization for differences in the RNA output size and sequencing depth between samples, as well as accounting for different gene lengths. Genes with at least a two-fold difference in expression between the subterranean rodents (DMR, FD, FD, NMR and coruro) compared to the mouse and guinea pigs, with a false discovery rate (FDR) ≤ 0.05, were defined as differentially expressed genes. The resulting gene set was manually curated to remove likely false positive calls (e.g., if FA has a very low RPKM value compared the other subterranean rodents and is similar to guinea pigs and the rat). While we appreciate that, unlike techniques such as *in situ* hybridization and laser microdissection-coupled RNA-seq, transcriptome analysis of a tissue pieces from a complex organ such as the brain may not correspond to a uniform cell population, we believe that this work provides a foundation for the study of genes that that could play a role in the adaptations to a subterranean environment.

Survey of globin gene expression response to short-term hypoxia
DMR, NMR and rat (n=3) were exposed to short-term hypoxia (8 hours) at 8% O₂, the estimated oxygen condition in NMR tunnels (Bennet and Faulkes, 2000). Samples were sequenced, processed and analyzed exactly as detailed above. *HIF3A*, which displays rapid elevated systemic mRNA expression upon short-term hypoxia in rats (Heidbrederet et al., 2003), was used to confirm a hypoxia response.
Survey of gene expression in the liver

Total RNA was isolated from the liver from 3- to 7-year-old female and male DMRs. Transcriptome sequencing was performed as outlined above. Briefly, in addition to the DMR, liver RNA-seq data generated from three additional species were obtained: NRM (Kim et al., 2011) (NCBI GEO accession no. GSE30337), and the mouse and rat (Merkin et al., 2012) (NCBI GEO accession no. GSE41637). The RNA-seq data analyzed comprised at least three individuals from each species. Transcriptome reads were mapped to reference genomes using TopHat (Trapnell et al., 2009), and mapped reads were analyzed using in-house Perl scripts. We identified orthologs and obtained expression data using gene annotations spanning coding regions. Differentially expressed genes were detected as outlined above. Genes with at least a two-fold difference in expression between the NMR and DMR compared to the mouse and rat, with a false discovery rate (FDR) ≤ 0.05, were defined as differentially expressed genes.

Gene enrichment analysis

Gene Ontology (GO) term analyses were performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) (Huang da et al., 2009). Briefly, to test for enrichment we compared genes that were significantly differentially expressed in the subterranean rodents against DAVID’s GO FAT database. The DAVID functional annotation tool categorizes GO terms and calculates an “enrichment score” or EASE score (a modified Fisher's exact test-derived p-value). Categories with smaller p-values (P < 0.01) (Kosiol et al., 2008; Qiu et al., 2012) and larger fold-enrichments (≥2.0) were considered interesting and most likely to convey biological meaning (Huang da et al., 2009).

Western blot analysis for hemoglobin α in rodent brain tissue

NMR, mouse, rat and guinea pig brains (frontal lobes) were used for preparation of protein lysate. Tissues were rinsed in cold PBS to minimize blood contamination, homogenized in RIPA buffer (Abcam) with protease inhibitor (Sigma), centrifuged 10 min at 150,000 rpm and total protein concentration in soluble fraction was determined by Coomassie Protein Assay Reagent (Thermo Scientific). 50 µg of total protein from each tissue samples was resolved by SDS-PAGE on 10% Bis-Tris gel in non-reducing conditions and transferred onto a PVDF membrane (Invitrogen). Goat polyclonal antibody against mouse hemoglobin α-chain (sc-31333, Santa-Cruz) was used as a primary antibody in dilution 1:500 and donkey anti-goat IgG-HRP (horseradish peroxidase) (sc-2020, Santa-Cruz) was used as a secondary in dilution 1:1000. Monoclonal anti-β-actin antibody (A2228, Sigma) was used for loading control. Western blots were developed with Clarity Western ECL blot substrate (Bio-Rad) and visualized with ChemiDoc XRS+ imaging system (Bio-Rad).

Identification of the novel 28S RNA in the NMR

Denaturing agarose gel electrophoresis

Total RNA from the NMR, DMR and mouse was isolated with a RNAqueous Midi total RNA isolation kit (Invitrogen) using 50-100 mg of frozen tissue. The resulting RNA samples (1 µg of total RNA) were electrophoresed on 1% denaturing agarose gels and stained with ethidium bromide. RNA Millennium Marker (Invitrogen) was used as a molecular weight marker.
Characterization of the 5` end of the 28S NMR fragment
Total RNA from the NMR liver was extracted with TRIzol reagent (Invitrogen) according to the manufacturer`s protocol. A 5` RACE RNA adapter (5`-GCUGAUUGCGAUGAAGAACUCUGGUUGCUGCUUUGAUGAAA-3`) was ligated with T4 RNA ligase 1 (NEB) in the following reaction mixture, which was incubated for 2 hours at 37 °C: 300 ng total RNA, 50 pmol 5` RACE adapter 1 µl PEG 8000, 1 µl buffer, 1 µl T4 RNA ligase 1 enzyme, and water to 10 µl. The reaction product was purified by phenol-chloroform method and precipitated by ethanol, and the resulting RNA was used for PCR using AccuPrime polymerase (Invitrogen) with primers 5RACEouter (5`-GCTGATGGCGATGAATGAACACTG-3`) and D6R (5`-GACTGACCCATGTTCAAAAGTTGTCGT-3`). Cycling conditions included 25 cycles at 60° for 30 sec, with elongation at 72° over 20 sec. PCR products were purified using a QIAGEN purification kit, subcloned into pGEM-T Easy (Promega) and transformed into DH5α E.coli strain (Invitrogen). Colonies were grown in LB media, and the plasmids purified with a QIAGEN miniprep kit and sequenced by capillary electrophoresis.

Characterization of the 3` end of the 28S NMR RNA fragment
300 ng of total RNA from the NMR liver was treated with polyA polymerase (NEB) with 1 mM ATP to attach a polyA tail to ribosomal RNA fragments. After ethanol precipitation, reverse transcription and second strand synthesis was set up using Invitrogen`s SuperScript Double-Stranded cDNA Synthesis kit. DNA was purified by phenol-chloroform and precipitated with ethanol. DNA ends were blunted with Klenow Polymerase (Fermentas) according to the manufacturer`s suggestions. After blunting, DNA fragments were phosphorylated by T4 kinase to enable subcloning into blunted pGEM-T Easy vector (Promega). Plasmids were prepared for sequencing as described above.

Multiple sequence alignments
BLASTn (Altschul et al., 1990) was used to interrogate nucleotide databases for sequences similar to the NMR and Talas tuco-tuco. 28S rRNA sequences of the DMR, NMR, mouse (GenBank accession no. AL355742), human (GenBank accession no. M11167), chimpanzee (GenBank accession no. K03429), guinea pig (GenBank accession no. NT_176417), Talas tocu-tuco (GenBank accession no. AF119340 cDNA and AF119339 genomic) were aligned using Clustal Omega (Sievers et al., 2011). Repeat elements were identified using RepeatMasker (Tarailo-Graovac and Chen, 2009).

Protein modeling
For protein modeling of ARG1 (Figure 2B), we used Molsoft ICM Browser Pro software (http://www.molsoft.com/icm_browser_pro.html) and a crystal structure model of human arginase I (D`Antonio et al., 2011), available from the RCSB Protein Data Bank (http://www.pdb.org) (PDB entry 3tf3) (Berman et al., 2003).

Validation of genes by PCR
The following genes were selected for validation by PCR and Sanger sequencing of NMR genomic DNA or cDNA: UCP1 (5`-CTCTTTCTGTGGTGGCTCCTTGAGGGACG-3 and 5`-TGTATGCCAAATCCTTTTCTGGAAGCTGA-3; 981 bp; cDNA template), CDKN2A (5`-ATCCGAGACTTTTTAAGGTGTCTG-3 and 5`-
ATCCGAGACTTTTAAGGGTTGCTG-3; 834 bp; cDNA template), MTNRIA (5’-TGACAAGAATTCTGCTCTGCT-3 and 5’-
GACCTTGAGCAATAACAGGTAACTATAATAATACT-3; 675 bp; gDNA
template), MTNR1B (5’-GTAATTATCCTTGTTGGTCTGCTGCT-3 and 5’-
ATTCATAGGTGGTTGGTCCGCTGTAGAGCAGA-3; 903 bp; gDNA template),
GJA10 (5’-AGACTAGAAGAGGGTAAGAAGTAAACTGAG-3 and 5’-
GAGAGAATATTGGCCTATCTACACACCCCTAC-3; 792 bp; gDNA template)
SUPPLEMENTAL REFERENCES

Albert, F.W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., Blanco-Aguilar, J.A., Plyusnina, I.Z., Trut, L., Villafuerte, R., et al. (2012). A comparison of brain gene expression levels in domesticated and wild animals. PLoS Genet. 8, e1002962.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet. 25, 25-29.

Audic, S., and Claverie, J.M. (1997). The significance of digital gene expression profiles. Genome Res. 7, 986-995.

Bennett, N.C., and Faulkes, C.G. (2000). African mole-rats: ecology and eusociality (Cambridge, UK: Cambridge University Press).

Bairoch, A., and Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. Nucleic Acids Res. 25, 31-36.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annal. Stat. 29, 1165-1188.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nature Struct. Biol. 10, 980.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. 14, 988-995.

Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z., and Kang, L. (2010). De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. PloS ONE 5, e15633.

D'Antonio, E.L., and Christianson, D.W. (2011). Crystal structures of complexes with cobalt-reconstituted human arginase I. Biochemistry 50, 8018-8027.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22, 1269-1271.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biot. 29, 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grubherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Prot. 8, 1494-1512.

Heidbreder M., Fröhlich F., Jöhren O., Dendorfer A., Qadri F., Dominiak P. (2003). Hypoxia rapidly activates HIF-3alpha mRNA expression. FASEB J. 11, 1541-3.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Prot. 4, 44-57.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27-30.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res. 12, 656-664.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M.,
Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479, 223-227.

Konopka, G., Friedrich, T., Davis-Turak, J., Winden, K., Oldham, M.C., Gao, F., Chen, L., Wang, G.Z., Luo, R., Preuss, T.M., et al. (2012). Human-specific transcriptional networks in the brain. Neuron 75, 601-617.

Kosiol, C., Vinar, T., da Fonseca, R.R., Hubisz, M.J., Bustamante, C.D., Nielsen, R., and Siepel, A. (2008). Patterns of positive selection in six Mammalian genomes. PLoS Genet. 4, e1000144.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., et al. (2010a). The sequence and de novo assembly of the giant panda genome. Nature 463, 311-317.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., et al. (2010b). De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265-272.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M., and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 40, W622-627.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338, 1593-1599.

Miller, W., Rosenbloom, K., Hardison, R.C., Hou, M., Taylor, J., Raney, B., Burchans, R., King, D.C., Baertsch, R., Blankenberg, D., et al. (2007). 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. Genome research 17, 1797-1808.

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol. Biol. 396, 59-70.

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yak genome and adaptation to life at high altitude. Nature Genet. 44, 946-949.

Salamov, A.A., and Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10, 516-522.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetverinin, V., Church, D.M., Dicuccio, M., Federhen, S., et al. (2012). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 40, D13-25.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21, 456-463.
Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 Suppl 2, ii215-225.

Tarailo-Graovac M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. In Current Protocols in Bioinformatics, A. D Baxevanis et al. ed. (Hoboken, NJ: John Wiley & Sons), pp. 4.10.1-4.10.14.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105-1111.

Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., et al. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 34, D187-191.

Yang, Z., and Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23, 212-226.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586-1591.

Yu, X.J., Zheng, H.K., Wang, J., Wang, W., and Su, B. (2006). Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics 88, 745-751.

Zhao, H., Yang, J.R., Xu, H., and Zhang, J. (2010). Pseudogenization of the umami taste receptor gene Tas1r1 in the giant panda coincided with its dietary switch to bamboo. Mol. Biol. Evol. 27, 2669-2673.