Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes

Fang, Xiaodong; Seim, Inge; Huang, Zhiyong; Gerashchenko, Maxim V.; Xiong, Zhiqiang; Turanov, Anton A.; Zhu, Yabing; Lobanov, Alexei V.; Fan, Dingding; Yim, Sun Hee; Yao, Xiaoming; Ma, Siming; Yang, Lan; Lee, Sang-Goo; Kim, Eun Bae; Bronson, Roderick T.; Sumbera, Radim; Buffenstein, Rochelle; Zhou, Xin; Krogh, Anders; Park, Thomas J.; Zhang, Guojie; Wang, Jun; Gladyshev, Vadim N.

Published in:
Cell Reports

DOI:
10.1016/j.celrep.2014.07.030

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
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,2 Zhiqiang Xiong,1 Anton A. Turanov,3 Yabing Zhu,1 Alexei V. Lobanov,3,4 Dingding Fan,1 Sun Hee Yim,3 Xiaoming Yao,1 Siming Ma,3 Lan Yang,1 Sang-Goo Lee,3,4 Eun Bae Kim,4 Roderick T. Bronson,5 Radim Sumbera,6 Rochelle Bjuvenstein,7 Xin Zhou,1 Anders Krogh,2 Thomas J. Park,6 Guojie Zhang,1,2 Jun Wang,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
6University 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.

INTRODUCTION

Subterranean rodents comprise approximately 250 species that spend their lives in dark, unventilated environments and are found on all continents except Australia and Antarctica (Begall et al., 2007). African mole rats (hystricognath rodent family Bathyergidae) are long-lived, strictly subterranean rodents that feed on underground roots and tubers. They are able to flourish in habitats that are poor in oxygen and rich in carbon dioxide and ammonia (Bennett and Faulkes, 2000), conditions that are harmful to mice and rats. It is hypothesized that the African Rift Valley acted as a geographical barrier in shaping the adaptive radiation of African mole rats into southern Africa and from there to other regions (Faulkes et al., 2004). Until now, the only African mole rat genome has been that of the 35 g naked mole rat (NMR, *Heterocephalus glaber*), which resides in northeast Africa and is the most basal African mole rat lineage (Kim et al., 2011). The lack of genomic information for closely related species thus far has prevented detailed analyses of African mole rat traits. To better understand the molecular mechanisms underlying the traits of African mole rats, we developed an improved NMR genome assembly, determined the genome sequence of the related ~160 g Damaraland mole rat (DMR, *Fukomys damarensis*) found in the arid regions of southwest Africa (Figures 1A and 1B), and sequenced the transcriptomes of additional subterranean rodents. Further analyses allowed us to decipher molecular adaptations consistent with subterranean life and shed light on unique traits of a most unusual mammal, the NMR.

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 (32%) 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 (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 (Kasai 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 (Burd et al., 2007; LaVinka et al., 2009). We found that arginase 1 (ARG1), which catalyzes the final step of the hepatic urea cycle and removes ammonia from the body, has a radical residue change in both the NMR and DMR: His254 replaces Leu/Tyr, which is present in 38 other vertebrate species (Figure 2A). This amino acid change was also detected in the distantly related subterranean blind mole rat (Spalacopus cyanus) and the semi-subterranean degu (Octodon degus) of South America. The common ancestor of Octodontoidea (coruro and degu) and Cavioida (guinea pig) diverged ~35 Mya, while African and South American rodents diverged ~41 Mya (Antoine et al., 2012; Meredith et al., 2011; Figure 2B). His254 is located immediately downstream of a conserved motif required for binding manganese and ARG1 function (Dowling et al., 2008; Figure 2C). Moreover, ARG1 is a homotrimer, with the salt bridges formed by Arg255 and Glu256 being critical for its assembly (Lavulo et al., 2001; Sabio et al., 2001). The charged residue flanking the ARG1 core may improve ammonia removal efficiency by interacting with the acidic Glu256 or by strengthening the Arg255-Glu256 salt bridge. In addition, several genes in the urea cycle were expressed at higher levels in NMR and DMR livers compared with mouse and rat (Table S4; Figure 2D). This included arginase 2 (ARG2), the second arginase gene that normally is not expressed in rodent liver. Moreover, expression of the mitochondrial ornithine transporter ORNT1 (SLC25A15), which is essential for the urea cycle (Fierrmonte et al., 2003), was elevated in the NMR and DMR. Taken together, these data indicate that subterranean hystricognath rodents present enhanced ammonia detoxification.

The buildup of CO₂ in underground habitats evokes pain, as CO₂ is converted into acid that stimulates pain receptors in the upper respiratory tract, nose, and eyes (Brand et al., 2010). A recent study found that a negatively charged motif in the sodium channel (Na(V)1.7 protein (SCN9A), which is highly expressed in nociceceptor neurons, prevents acid-induced pain signaling to the NMR brain (Smith et al., 2011). We compared 44 vertebrate sequences and found that the motif is also present in the DMR, two African mole rats in the same genus as the DMR (the Ansell’s mole rat [Fukomys anselli, FA] and the Mashona mole rat [Fukomys darlingi, FD]), the South American subterranean coruro and semi-subterranean degu, the cave-roosting little brown bat (Myotis lucifugus), and the European hedgehog (Erinaceus europaeus). The distantly related subterranean blind mole rat (Spalax galilii) also harbors the same amino acid changes in Na(V)1.7 and adaptation to high CO₂ levels.

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 α (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). Hemo-globin 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) in hemoglobin a 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 subterranean rodents. Overall, our data suggest that globins are constitutively and highly expressed in the brain of hypoxia-resistant rodents and play a major role in their ability to adapt to an oxygen-poor subterranean environment. Experiments are in progress 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 laboratories 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 homogeneous 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 oxidoreduction. 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 observations 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 overexpressed 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 premature aging (Campisi et al., 2011). Furthermore, knockdown of FASTK is also associated with improved neuron elongation and regeneration (Loh et al., 2009). 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 aging 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 Bunkenstein, 2004). We found that the NMR, DMR, and other hystricognath rodents harbor a divergent insulin b-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 a chain and, in particular, the b-chain are highly divergent, with concomitant alterations 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 b-chain result in reduced insulin processing, misfolding, and less effective insulin (based on receptor binding) (Liu et al., 2010). Interestingly, residue 22 of the b-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 porcupine, 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 alternative 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 animals 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 (Duvillié 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 hystrognath 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 (INISIG2), 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 hystrognaths (e.g., porcupine and guinea pig). Our findings support the hypothesis that hystrognath 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., 2009; 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 inducible mutations activating AMPK, a mitochondrial protein, 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 (Qi 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 (Hansen 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., 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.

The calcitonin gene CALCA is responsible for the synthesis of two distinct preprohormones by means of alternative splicing (Rosenfeld et al., 1981). Splicing into exon 4 results in calcitonin (CT), while splicing into exon 5 and the 3’ untranslated exon 6 encodes the sensory neuropeptide CGRP (Figure S3E). The splicing of CALCA is under tight endocrine control, and the regulation of CT and CGRP is complex and involves distal and proximal elements in the CALCA promoter, as well as a range of regulatory elements within and flanking exon 4 (van Oers et al., 1994). Our analysis revealed that there are unique deletions in NMR CALCA, including a conserved 6 bp region, in the 3’ untranslated part of exon 4 (Figure S3F). Given that splicing factors are known to be composite and context dependent (Wang and Burge, 2008), and that elements within exon 4 of CALCA regulate CT and CGRP isoform switching, the deleted region in exon 4 may disrupt the mutually exclusive tissue-specific splicing of CT exon 4 and CGRP exons 5–6, resulting in the observed lack of CGRP expression in sensory neurons of the NMR.

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).
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 talarum*) (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

**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., 2006). 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 76x) high-quality data, including 151 Gbp (or 50x) short insert size reads, were generated (Table 1). The genome was de novo assembled by SOAPdenovo (Li et al., 2010a). Then, 151 Gpb (or 50x) 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 Gpb (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=9400 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 kb 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., 2008). After sorting alignments by the leftmost 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 varFilter, which can be found in the SAMtools package, as the filter tool with parameters -Q 20 -q 20 -d 4 -D -100 -S -20 -i 20 -N 5 -S -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.

**ACCESSION NUMBERS**

The DMR whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under accession code AYUG00000000. The version described in this paper is the first version, AYUG01000000. All short-read data have been sequenced 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**


**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**


Eusociality (Cambridge, UK: Cambridge University Press).


Eusociality (Cambridge, UK: Cambridge University Press).


