PhD thesis
Xiaodong Fang

Convergent and divergent adaptations of subterranean rodents
Genome wide studies on Heterocephalus, Fukomys and Spalax

Academic advisors:
Anders Krogh
University of Copenhagen, Denmark

The PhD School of Science, Faculty of Science, University of Copenhagen
Submitted: 30/03/2015
Dissertation for the degree of philosophiae doctor (PhD)
Department of Biology, University of Copenhagen
Copenhagen, Denmark
And
BGI-Shenzhen
Shenzhen, China
March 2015

Author: Xiaodong Fang

Title: Convergent and divergent adaptations of subterranean rodents
Subtitle: Genome wide studies on Heterocephalus, Fukomys and Spalax

Academic advisors:
Anders Krogh
University of Copenhagen, Denmark

The PhD School of Science, Faculty of Science, University of Copenhagen
Submitted: March 2015
This PhD project started in 2011 as collaboration between the Department of Biology, University of Copenhagen and BGI-Shenzhen. The work presented here has been performed at both institutions under the supervision of Prof. Anders Krogh.
ACKNOWLEDGEMENTS

I take this opportunity to thank all people who supported me during my PhD project. First of all, I would like to express my thanks to my super advisor Professor Anders Krogh. He is very kind and nice to me for providing mentor and help during my PhD project. Thanks have to be expressed to the staffs in the university, their helps in administration and support should not be missed. Secondly, I have to thank BGI for providing such opportunity for me to join the PhD program, without the support from BGI, it will never happen. Thirdly, my PhD program, mainly consists of three genome projects, it is quite huge and never can be finished by myself only, thus I am using this great opportunity to thank all my collaborators around the world, I can not finish the projects without their ultra important contributions, here I would like to especially express my appreciations to Professor Vadim N. Gladyshev and Eviatar Nevo and Dr. Lijuan Han, their ideas, knowledge and leadership are the basis of the success of these genome projects. Finally, a special thanks to my families, I always received encouragement and power from my families when I had tough time during my PhD program.
ABSTRACT

Subterranean rodents comprise approximately 250 species that spend their entire lives in underground, unventilated tunnels, distributed along all continents except Australia and Antarctica. Subterranean rodents escape from predators and extreme climatic fluctuations in their underground habitats, but subject to various stressors such as darkness, oxygen deficiency, hypercapnia, food shortage and high infectivity. As representative of strictly subterranean rodents, the naked mole rat (NMR, Heterocephalus glaber), the Damaraland mole rat (DMR, Fukomys damarensis) and the blind mole rat (BMR, Spalax galili) have evolved convergent and divergent traits in many of their morphological, physiological, and/or behavioral characteristics, which facilitate their adaptations to a similar underground burrowing life style. For example, all these three rodents show degenerate visual acuity and advanced sensory systems in the dark; they display remarkable tolerance to a living environment with an excess of carbon dioxide and ammonia, but lack of oxygen; they exhibit extraordinarily long lives, and keep a fantastic resistance to cancer and other aging-associated diseases. In this study, we reported the genomic and transcriptomic information of DMR and BMR. By comparing the genomes and transcriptomes of subterranean rodents with that of their aboveground counterparts, we uncovered candidate molecular mechanisms of mammalian adaptation to subterranean environment. The extreme biological features of the NMR, BMR and DMR, coupled with the reported genetic information, will promote the utilization of subterranean animal models for biological and biomedical research in the fight against aging, cancer, stroke and other related diseases.
Summaries

Summaries in English:
Stress is an evolutionary driving force of great importance. Subterranean rodents are a widely distributed group of species that living in underground burrows for most of their lives. The subterranean ecotope protects them to escape from aboveground climatic fluctuations and predators, while challenge them with various stresses including light deficiency, low oxygen, high carbon dioxide, food scarcity, and high pathogenicity. In all continents excepting Australia and Antarctica, about 250 extant rodent species primarily live underground and therefore share several common traits, including highly developed sensory except vision, remarkable tolerance to hypoxia and hypercapnia, similar energy metabolism pattern (such as low body temperature and low metabolic rate), etc. In the meantime, every species show divergent adaptations to their separated feeding niches and phylogenies. The evolutionary biology of subterranean rodents provides excellent animal models to elucidate how the organisms evolutionarily adapt to harsh environments. What’s more, given the remarkable features of certain subterranean rodents, such as longevity, anti-cancer, hypoxia tolerance, comparing them with short-lived, cancer-prone, hypoxia susceptible aboveground rodents at the molecular level, would greatly widen our understanding of disease and aging, and thus benefit future biomedical applications.

In this thesis, we sequenced the genomes and transcriptomes of three subterranean rodents: the naked mole rat (NMR, *Heterocephalus glaber*), the Damaraland mole rat (DMR, *Fukomys damarensis*) and the blind mole rat (BMR, *Spalax galili*), not only because they are subterranean representatives that achieve extraordinary success in flourishing underground, but also due to their medically useful features in anti-aging and cancer resistance. By comparing subterranean rodents NMR, DMR and BMR with the well-known aboveground rodents mouse (*Mus musculus*) and rat (*Rattus norvegicus*), we discussed the shared adaptions and divergent evolutions among them. We characterized the genomic features of subterranean rodents, as well as their specific transcriptomic regulation pattern in response to hypoxia, aging, or inducing tumorigenesis. The extreme traits of the NMR, DMR and BMR, coupled with the reported genomic and transcriptomic information, provide excellent opportunities to understand convergent and divergent adaptations, and would offer new insight into cancer prevention and lifespan extension in human.

Summaries in Danish:
Stress er en evolutionær drivkraft af stor betydning. Underjordiske gnavere er en vidt udbredt gruppe af arter, der tilbringer det meste af deres liv i underjordiske gange. Det underjordiske levested beskytter dem mod rovdyr og climatiske udsving over jorden, men udfordrer dem med adskillige stressfaktorer såsom mørke, hypoxi, hyperkapni, fødeknapthed og patogene. Fordelt på alle kontinenter undtagen Australien og Antarktis findes omkring 250 nulevende gnaverarter, der fortrinsvis lever underjordisk og derfor deler adskillige fællestræk, der omfatter højtudviklede senser fraregnet syn, en bemærkelsesværdig tolerance over for hypoxi, hyperkapni, et ens mønster for energimobilismen (lav kropstemperatur og metabolisk rate) etc. Imidlertid udviser hver art også afvivende tilpasninger i forhold til deres forskellige føde nicher og fyllogenese. De underjordiske gnaveres evolutionsbiologi bidrager derfor med fremragende dyremodeller til studiet af organismens evolutionære tilpasning til barske miljøer. Yderligere giver de bemærkelsesværdige træk hos visse underjordiske gnavere, såsom et langt liv, lille kæfrisiko og hypoxitolerance, i modsætning til de kortlivede, kæftudsatte og hypoxifølsomme overjordiske gnavere, en mulighed
for sammenligning på det molekylære niveau, som kan øge vores forståelse for sygdomme og aldring betydeligt og dermed gavne biomedicinsk anvendelser i fremtiden.

## Contents

**SUMMARIES** .................................................................................................................................................. 6

**1 INTRODUCTION** ......................................................................................................................................... 11

1.1 Subterranean rodents ................................................................................................................................ 11

1.2 Next generation sequencing (NGS) technology ............................................................................................ 15

1.3 Bioinformatic approaches .......................................................................................................................... 15

**2 AIMS............................................................................................................................................................ 19

**3 METHODS.................................................................................................................................................... 20

3.1 Genome assembly ......................................................................................................................................... 20

3.2 Genome annotation ....................................................................................................................................... 21

3.3 Whole genome synteny analysis .................................................................................................................... 22

3.4 Gene family analysis ....................................................................................................................................... 22

3.5 Phylogenetic tree and divergence time .......................................................................................................... 22

3.6 Pseudogenes detection .................................................................................................................................. 22

3.7 Gene gain and loss ........................................................................................................................................... 22

3.8 Transcriptome analysis ................................................................................................................................... 23

**4 RESULTS..................................................................................................................................................... 24

4.1 NMR, DMR and BMR genomes statistics ...................................................................................................... 24

4.2 Vision degeneration ....................................................................................................................................... 25

4.3 Circadian rhythm ........................................................................................................................................... 27

4.4 Hypercapnia, hypoxia, and ammonia tolerance ............................................................................................ 31

4.5 Longevity ....................................................................................................................................................... 34

4.6 Cancer resistance .......................................................................................................................................... 35

4.7 Other traits .................................................................................................................................................... 37

**5 CONCLUSIONS ........................................................................................................................................... 40

**6 PERSPECTIVE............................................................................................................................ 41
7 REFERENCES .................................................................................................................. 42
8 APPENDIX .................................................................................................................. 46
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG1</td>
<td>arginase 1</td>
</tr>
<tr>
<td>ARG2</td>
<td>arginase 2</td>
</tr>
<tr>
<td>ASS1</td>
<td>argininosuccinate synthase</td>
</tr>
<tr>
<td>ASL</td>
<td>argininosuccinate lyase</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BMR</td>
<td>blind mole rat, <em>Spalax galili</em></td>
</tr>
<tr>
<td>BRMC</td>
<td>Bayesian relaxed molecular clock</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CPS1</td>
<td>carbamoyl phosphate synthase 1</td>
</tr>
<tr>
<td>DAVID</td>
<td>database for annotation, visualization and integrated discovery</td>
</tr>
<tr>
<td>DEG</td>
<td>differentially expressed gene</td>
</tr>
<tr>
<td>DMR</td>
<td>Damaraland mole rat, <em>Fukomys damarensis</em></td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>HGP</td>
<td>Human Genome Project</td>
</tr>
<tr>
<td>HMM-HA</td>
<td>high-molecular-mass hyaluronan</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>NGS</td>
<td>Next generation sequencing</td>
</tr>
<tr>
<td>NJ</td>
<td>Neighbor-Joining</td>
</tr>
<tr>
<td>NMR</td>
<td>naked mole rat, <em>Heterocephalus glaber</em></td>
</tr>
<tr>
<td>ORNT1</td>
<td>ornithine transporter 1</td>
</tr>
<tr>
<td>OTC</td>
<td>ornithine transcarbamylase</td>
</tr>
<tr>
<td>RPKM</td>
<td>Reads Per Kilo bases per Million reads</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>SINEs</td>
<td>short interspersed nuclear elements</td>
</tr>
<tr>
<td>SP</td>
<td>substance P</td>
</tr>
<tr>
<td>TE</td>
<td>transposable element</td>
</tr>
<tr>
<td>TR</td>
<td>tandem repeat</td>
</tr>
<tr>
<td>TRAIN</td>
<td>transcription of repeats activates interferon</td>
</tr>
<tr>
<td>TRF</td>
<td>Tandem Repeats Finder</td>
</tr>
<tr>
<td>WGS</td>
<td>whole-genome shotgun</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Subterranean rodents
There are approximately 250 species (38 genera, 6 families) subterranean rodents distributed on all continents except Australia and Antarctica. They spend their entire lives in dark, unventilated, self-constructed underground tunnels, mainly feeding on underground roots and tubers. The underground living condition protects residents from the threat of predators and fluctuations of temperature and humidity. However, they had to evolve genetic adaptive complexes [1] to deal with underground stressors of darkness, food shortage, oxygen deficiency, high carbon dioxide, and high infectivity. In this thesis we focused on three representative subterranean rodents: NMR, DMR and BMR, and carried out intensive comparative genomics studies among underground and aboveground rodents.

Figure 1. Representative subterranean families all across the world.
NMR

The NMR belongs to *hystricomorpha* rodent family *Bathyergidae*, resembling a saber-toothed sausage, with pink, wrinkly skin and sparse hair (Figure 2). Although looks ugly, as the first African mole rat being sequenced, the NMR has some very fantastic traits. For example, the NMR has a lifespan exceeding 30 years, 10 times longer than its cousins mouse and rat, making it the longest living rodent [2]. Moreover, the NMR shows delaying senescence, characterized by no age-related increase in mortality, and normal reproductive and physiological functions until death [3]. The NMR is a remarkable cancer-proof rodent, shows extraordinary resistance to not only spontaneous cancer [4] but also experimentally induced tumorigenesis [5]. NMRs are unable to sustain thermogenesis [6], and are insensitive to certain types of inflammatory pain [7]. Recent research reported that high-molecular-mass hyaluronan (HMM-HA) mediates the early contact inhibition in the NMR, contributing to its cancer resistance; HMM-HA may also relates with longevity by increasing stress resistance as a result of the antioxidant and cytoprotective properties of hyaluronan. What’s more, the NMR live in eusocial groups like ants and bees, making it one of the only two known eusocial mammals (another one is the DMR, which we will introduce later). Different NMR individuals within one colony carry out specialized jobs; normally there are one breeding female (the queen), one to three breeding males, and hundreds of fertility suppressed workers[8] living together.
Closely related with NMR, the DMR (Figure 3) also belongs to *hystricomorpha* rodent family *Bathyergidae*, distributed mainly in the arid regions of southwest Africa. DMR and NMR share some common traits such as poor vision, hypoxia and hypercapnia tolerance, longevity, etc. And they are the only two known eusocial mammals, naturally live in large colonies with one breeding female, the ‘queen’, and her offspring. Aside from the queen and her one to three breeding male, the sexual maturity of all subordinates were suppressed [8].
The BMR (Figure 4) belongs to the Muroidea family Spalacidae, with a distribution mainly along the Eastern Mediterranean and North Africa. Although with similar sounding names, BMR and NMR species diverged more than 70 million years ago [9], and BMRs are more closely related with mice and rats than to NMRs. Like other subterranean rodents, the darkness in underground burrows leads to a mosaic evolution of visual function, including ocular regression, coupled with photoperiodic perception progression [10,11]. In the meanwhile, the BMR evolved a larger brain volume with developed tactile sense, olfactory, vibration perception, and magnetic orientation systems to compensate vision degeneration [12]. Most remarkably, since oxygen as low as 7.2% and carbon dioxide as high as 6.1% were recorded in their underground tunnels [13], the BMR has to evolve physiological adaptations to survive and carry out intense activities under severe hypoxic and hypercapnic stress. The BMR is one of the long-lived rodents (with a lifespan exceeding 20 years) that is also resistant to cancer. There is a unique substitution in a DNA-binding domain of BMR p53, which is suggested to inhibit hypoxia-induced apoptosis while in favour of cell cycle arrest [14] and necrosis. Our genomic investigation uncovered that anti-cancer in the BMR is possibly regulated by an interferon-mediated necrotic cell death mechanism [9].

Since environmental stress is considered as an important evolutionary driving force [15,16], most of subterranean rodents share convergent evolutions in adapt to a shared ecotope, e.g. all of them have more or less regressive vision ability and progressive hypoxia and hypercapnia tolerance. In the meanwhile, they show divergent adaptations to their separated feeding niches and different phylogenies. Taking the behavioral traits as examples, BMR is solitary style whereas NMR and DMR are eusocial. With the fast-developing of NGS technology, recent progress in rodent comparative genomics offers an inimitable opportunity for the investigation on convergent and divergent evolution, and the discovery of molecular mechanisms that underlie longevity and cancer susceptibility.
1.2 Next generation sequencing (NGS) technology

DNA sequencing refers to the process of recognizing the precise nucleotide composition within a DNA molecule. Sanger sequencing, also known as first-generation sequencing, was the first mainstream sequencing technology developed by Edward Sanger in 1975. Since then, Sanger sequencing was considered to be the gold standard for DNA sequencing for two and a half decades [17]. The Human Genome Project (HGP) was accomplished by using Sanger sequencing. However, the pressing need for reducing sequencing cost has driven the development of next generation sequencing (NGS), or high-throughput sequencing technologies, which parallelize the sequencing process and generate millions of sequencing reads concurrently. Commonly used NGS technologies, including Illumina Solexa, ABI SOLiD, Roche 454, and Helicos, are intended to lower the cost of DNA sequencing to a large extent, compared with standard Sanger sequencing. To date, NGS technologies have been widely used in a variety of contexts, including exome sequencing, targeted sequencing, whole-genome sequencing, transcriptome profiling, etc. The recent accumulation of sequenced genomes from all branches of life presents us with an extraordinary opportunity to decipher life.

1.3 Bioinformatic approaches

As an interdisciplinary area of science, bioinformatics integrates mathematics, statistics, computer science and bioscience, develops a mass of computational algorithms, computer databases and software tools to elucidate biological data. Common used bioinformatic approaches include sequence alignment, de novo assembly, variant identification, etc. The domain of bioinformatics includes the study of genes, transcripts, proteins, and metabolites by trans-omic approaches.

1.3.1 Genome assembly

Genome assembly refers to the process of taking a great many fragmented sequencing reads and stitching them back together to construct a reference genome from scratch. Since adjacent sequencing reads usually overlap by tens of base pairs, the assembly process is essentially turned into a jigsaw puzzle with millions of pieces. Although greatly lower the cost of sequencing, the very short reads length, together with the significantly large amount of reads, increase the computational complexity of genome assembly of NGS. Level of genomic heterozygosity, proportion of repetitive sequence, as well as rate of sequencing errors, all those factors can impede the construction of high quality assembly.

To de novo assemble a genome based on NGS sequencing reads, computer programs typically make use of both overlapping information and pair end relationship. First, short sequencing reads can be joined up through overlapping regions into a continuous sequence, known as a ‘contig’. Factors such as genomic repeats, polymorphisms, missing data and sequencing errors would eventually restrict the length of the contigs that assemblers can build. Next, since paired reads were generated from the same piece of DNA, using the PE information can help link contigs into ‘scaffolds’, ordered assemblies of contigs with gaps in between. We can also use paired-read data to estimate the size of genomic repetitive regions.
1.3.2 Genome annotation

Annotation bridges the gap from the genome assembly to the biological features of the species. The high-quality annotation would help to identify the key traits of the genome, including repetitive sequences which might serve as function regulationary elements, protein-coding genes and their products, as well as non-coding RNA genes.

Genomic repetitive sequences include tandem repeats (TRs) and transposable elements (TEs). TRs refers to a pattern that two or more nucleotides occurs repeated and the repeated unit are adjacent to each other. TEs are DNA sequences that can move or duplicate within genomes, by themselves or with the assistance of other elements. TEs are recognized to affect genome size, genome structure, and chromosomal rearrangements [19-21]. TEs may act as “controlling” elements and regulate the transcriptional expression of adjacent genes [22]. Moreover, TEs could modify, re-wire and create gene regulatory networks [23]. Therefore, it is of great importance to understand the distribution and classification of repeat elements in a newly sequenced genome.

In addition to genomic repetitive element, there are two types of genes in the genome assembly: protein-coding genes and non-coding RNA (ncRNA) genes. Protein-coding genes can be transcribed to mRNA, then translated to functional proteins. Annotion of protein-coding genes is of great importance, since their protein products directly involved in biological or physiological process. ncRNA genes, including transfer RNA (tRNA), ribosomal RNA (rRNA), microRNAs (miRNA), long non-coding RNA (lncRNA), small nuclear RNA (snRNA), etc, represent a category of genes that do not encode proteins, but produce functional RNA molecules. Although being ignored by early genomic investigation,
ncRNA genes are suggested to have particularly abundant regulatory functions, such as guiding post-transcriptional regulation, or directing RNA modifications.

1.3.3 Orthologous and gene family analysis
Orthologous genes, or orthologs, refer to genes in different species that descended from a single gene of the last common ancestor. By orthologous genes identification and cross-species comparison, we could gain useful information in taxonomic classification and phylogenetic studies among species.

Genes are categorized into families according to sequences similarity. A gene family refers to a set of genes with high similarity in nucleotide or amino acid composition, formed by duplication of a single original gene, and generally with similar biochemical functions. Although the expansion or contraction of gene families along a specific lineage can be due to chance, in most cases it can be the result of natural selection, thus related with certain species-specific traits.

1.3.4 Phylogenetic relationship
A phylogenetic tree, which can be inferred according to nucleotide or amino acid sequences similarity, could represent the evolutionary relationships among species in a very intuitive way.

There are various computational methods to construct phylogenetic trees based on a number of input sequences. Distance-matrix methods, including neighbor-joining and UPGMA, calculate genetic distance from multiple sequence alignments, represent the simplest approach to construct phylogenetic relationship. Many sequence alignment programs, such as ClustalW, also use the distance-based algorithms to create tree. Maximum parsimony is another simple method of estimating phylogenetic trees, but implies an implicit model of evolution (i.e. parsimony). More advanced methods use the optimality criterion of maximum likelihood, often within a Bayesian Framework, and apply an explicit model of evolution to phylogenetic tree construction.

1.3.5 Functional genomic features
By comparing the genome with that of closely related species, we could gain the information about conserved genomic regions that potentially related with maintaining of the essential cellular function, as well as divergent genomic elements that potentially associated with species-specific traits. There are several types of divergent genomic traits we need to pay attention to: pseudogenes, gain or lost genes, positively selected genes, etc.

Pseudogenes are DNA sequences that resemble functional genes in nucleotide composition, but are inactivated by frameshift mutation or early termination. Pseudogenes often result from the accumulation of multiple mutations within a gene whose product is not required for the survival of the organism.

Gain or loss events also represent the role of the evolutionary force on shaping the genome. Genes gain event might related with enhancements of certain function that benefit the adaptiveness of the organism; while gene loss event occurs when there are unnecessary function or functional redundancy.

Positively selected genes contain certain mutations that are favored and preserved during adaptive evolution, which potentially benefit the species’s adapt to environment stresses. Identification of genes or genetic loci that underwent positive selection could enhance our understanding about evolutionary adaptions.
1.3.6 Transcriptome analysis

Transcriptome includes all mRNA transcripts in the cell, reflecting the genes that actively expressed at any given time. Unlike the genome, which is fixed for a given cell line (excluding mutations) in the rough, the transcriptome can change with external environmental conditions. Therefore, through elaborate experiments and rational sampling time, comparative transcriptomic analysis would enable us to identify those differentially expressed genes (DEGs), which are closely related with certain biological processes that we are interested in.
2 Aims

With the widely application of whole-genome sequencing, as well as the accumulation of reference genome information, recent progress in rodent comparative genomics offers an inimitable opportunity for the investigation on convergent and divergent adaptations, and the discovery of molecular mechanisms that underlie longevity and cancer susceptibility.

Here we present the most comprehensive genomic and transcriptomic study of three subterranean rodents: NMR, DMR and BMR. Together with the genome sequence of mouse and rat, as well as that of other sequenced mammals, we carried out extensive comparative biology investigation among subterranean and aboveground rodents.

One of the most important destinations of this thesis is to elucidate how the organisms evolutionarily adapt to living environments. By sequencing and comparing the whole genomes of different subterranean rodent species, we are able to analyze the convergent and divergent evolution of subterranean rodents at the molecular level with a tremendous efficiency, and thus get better understanding on how these species convergently and divergently deal with their special underground niches.

What’s more, given the facts that maximum lifespans for rodents range from 3-4 years in mice and rats, to over 20 years in BMRs and DMRs, and exceeding 30 years in NMRs; Cancer incidence rate also differ substantially between cancer-prone mice/rats and cancer-resistant NMRs/BMRs, compare genomic and transcriptomic information of NMR, DMR and BMR with that of those short-lived, cancer-prone aboveground rodents, would greatly widen our understanding of genetic basis that control longevity and cancer susceptibility, and thus benefit future biomedical application.
3 Methods

3.1 Genome assembly

Based on the features of genomic repetitive sequences, libraries of different insert sizes, including small insert-size libraries (e.g., 250bp, 500bp, 800bp, etc) and large insert-size libraries (e.g., 2Kbp, 5Kbp, 10Kbp, 20Kbp, etc), are sequenced to obtain the whole genome sequencing depth of over 70X coverage, ensuring the precision of each single base and genome integrality. The short sequencing reads were assembled using SOAPdenovo [24]—a assembler that developed specifically for NGS short reads. SOAPdenovo employs the de Bruijn graph algorithm to reduce computational complexity and simplify the assembly process.

First, the k-mer frequency methodology was used to filter out sequencing reads with low quality, and correct potential sequencing errors. After this quality control step, SOAPdenovo first splits the reads from small insert size libraries (170-800 bp) into Kmers and merges Kmers, constructing the de Bruijn graphs; unambiguous connections in de Bruijn graphs are then collected to build contigs. Then the sequencing reads from large insert size libraries were mapped onto contigs for scaffolds building; Paired-end information was subsequently employed to link contigs into scaffolds. Pipeline for assembling NGS sequencing reads using SOAPdenovo was shown in Figure 6.
Figure 6. Pipeline for assembling NGS sequencing reads using SOAPdenovo.

3.2 Genome annotation

In this thesis, TRs were annotated using Tandem Repeats Finder (TRF) [25]. TEs were identified and classified by homology to RepBase sequences, which contains many known TEs, using the softwares RepeatProteinMask and RepeatMasker [26] with default parameters.

To predict genes in the NMR, DMR and BMR genomes, we used a combination of homology-based method, de novo method, as well as transcriptome-based evidence. For the homology-based prediction, mouse and human proteins (Ensembl release 64) were mapped onto the assembly using tBLASTn [27]. Then Genewise[28] were used to define gene models. For de novo prediction, Augustus [29] and Genscan [30] were employed to predict coding genes, using appropriate parameters. RNA-seq data were mapped to genome using Tophat [31], 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 integrate to form a comprehensive and non-redundant reference gene set using GLEAN (http://sourceforge.net/projects/glean-gene/).
Functions of genes were assigned based on the best hit derived from the alignments to proteins annotated in SwissProt and TrEMBL databases using Blastp. Motifs and domains were annotated via InterPro by searching against publicly available databases such as Pfam, PROSITE, PRINTS, ProDom, and SMART. Gene Ontology information was also retrieved from InterPro. By searching across KEGG database and finding the best hit for each gene, we could gain the KEGG pathway annotation.

3.3 Whole genome synteny analysis
To detect synteny blocks between NMR/DMR/BMR and other mammals, pairwise whole-genome alignment was performed using LASTZ with parameters T=2 and Y=9400 (http://www.bx.psu.edu/miller_lab/). ChainNet, which can integrate genomic inversions, translocations, duplications, and deletions, was used to accommodate traditional alignments into larger structures.

3.4 Gene family analysis
In order to identify gene families, DNA and protein sequences for other sequenced mammals (for example, human, mouse, rat, etc) were downloaded from the Ensembl database release 64. Then the Treefam methodology was used to identify gene families among considered species. Detailed parameters were described in supplementary information of reference [37].

3.5 Phylogenetic tree and divergence time
Phylogenetic relationship among NMR/DMR/BMR and other related mammals were constructed using single-copy orthologous gene sequences, as well as 4-fold degenerate sites. We employed Modeltest to choose the best fit substitution model, and Mrbayes to reconstruct the tree. The Bayesian relaxed molecular clock (BRMC) method was used to estimate the divergence time among species via the program multidivtime. Detailed parameters were described in supplementary information of reference [37].

3.6 Pseudogenes detection
To predict pseudogenes in the NMR/DMR/BMR genome, the longest transforms of human proteins in ensemble release 64 were used to call homologs in the NMR/DMR/BMR genomes. Once frameshift mutation or early termination events were observed in homologous regions, we carried out further manual review on the genomic reads mapping quality nearby. Gene with high mapping quality but inactivated by frameshift mutation or early termination were identified as pseudogenes.

3.7 Gene gain and loss
Since orthology information showed synteny information at the protein level, it could be used to analyze gene gain and loss events between human and NMR/DMR/BMR. Within the protein synteny blocks, if a human protein had no NMR/DMR/BMR ortholog, and excluding false positive predictions that could be caused by annotation or genome assembly (gap > 5%), this protein could be defined as either being lost in the NMR/DMR/BMR lineage or gained in the human lineage. Using NMR/DMR/BMR as a reference to generate the orthology relationship, we applied similar procedure to identify genes gained in the NMR/DMR/BMR lineage compared to the human lineage.
3.8 Transcriptome analysis

Gene expression levels were calculated according to RPKM (Reads Per Kilo bases per Million reads) [42]. Transcriptome reads were mapped onto corresponding reference genome by Tophat [31], and the mapped reads were analyzed by using in-house Perl scripts. The total read numbers were normalized by multiplying a normalization factor, so as to minimize the influence of difference in RNA output between the samples[43].

DEGs were detected using the method described by Chen et al [44], based on the Poisson distribution [45] 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 level between the subterranean rodents (NMR, DMR and BMR) compared to the mouse and rat, with a false discovery rate (FDR) ≤ 0.05, were defined as DEGs. The resulting gene set was manually curated to remove likely false positive calls.

GO term analyses were performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) [46]. Briefly, we compared genes that were significantly differentially expressed in the subterranean rodents against DAVID’s GO FAT database to test for enrichment. 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) [47,48] and larger fold-enrichments (≥2.0) were considered interesting and most likely to convey biological meaning [46].
4 Results

In this study, genomes of the three subterranean rodents (NMR, DMR and BMR) were sequenced. By compared with aboveground counterparts (mouse and rat), we characterize the genomic and transcriptomic features of subterranean rodents, discussed the shared convergent and divergent adaptations among them. The extreme traits of the NMR, BMR and DMR, coupled with the reported genetic information, not only enable us better understand adaptive evolution, but also provide opportunities for addressing some of the most challenging questions in biology and biomedicine, such as anti-aging, cancer resistance, hypoxia and hypercapnia tolerance, thermogenesis, vision degeneration, circadian rhythms, etc.

4.1 NMR, DMR and BMR genomes statistics

We applied a whole-genome shotgun (WGS) strategy to sequence the genomes of NMR, DMR and BMR. The global statistics of the assemblies were shown in Table 1.

Table 1. Statistics of the NMR, DMR and BMR genomes assembly [9,37,49].

Sequencing data

<table>
<thead>
<tr>
<th>Species</th>
<th>Insert size (bp)</th>
<th>Total data (Gb)</th>
<th>Sequencing coverage (X)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>170-800</td>
<td>126.5</td>
<td>46.9</td>
</tr>
<tr>
<td>NMR</td>
<td>2,000-20,000</td>
<td>120.7</td>
<td>44.7</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>247.2</td>
<td>91.6</td>
</tr>
<tr>
<td></td>
<td>170-800</td>
<td>151.6</td>
<td>50.5</td>
</tr>
<tr>
<td>DMR</td>
<td>2,000-20,000</td>
<td>77.5</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>229.0</td>
<td>76.4</td>
</tr>
<tr>
<td></td>
<td>170-800</td>
<td>172.6</td>
<td>57.5</td>
</tr>
<tr>
<td>BMR</td>
<td>2,000-20,000</td>
<td>87.1</td>
<td>29.0</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>259.7</td>
<td>86.6</td>
</tr>
</tbody>
</table>

Assembly results

<table>
<thead>
<tr>
<th>Species</th>
<th>N50 (kp)</th>
<th>Longest (kb)</th>
<th>Size (Gb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>contigs</td>
<td>19.3</td>
<td>178.9</td>
</tr>
<tr>
<td></td>
<td>scaffolds</td>
<td>1,585</td>
<td>7,787</td>
</tr>
<tr>
<td>DMR</td>
<td>contigs</td>
<td>22.9</td>
<td>229.6</td>
</tr>
<tr>
<td></td>
<td>scaffolds</td>
<td>4,996</td>
<td>22,231</td>
</tr>
<tr>
<td>BMR</td>
<td>contigs</td>
<td>27.6</td>
<td>285.5</td>
</tr>
<tr>
<td></td>
<td>scaffolds</td>
<td>3,618</td>
<td>15,899</td>
</tr>
</tbody>
</table>
4.2 Vision degeneration

Vision degeneration and other sensory (e.g., vibrational, tactile, vocal, olfactory, et al.) progression are shared traits of underground-dwelling mammals in adapt to darkness, thus be considered as a clear evidence of convergent evolution.

We identified 59 and 29 lost gene families in DMR and NMR genomes, respectively, including tens of genes associate with visual perception function (Table 2 and Figure 7). In the meanwhile, we found 22 BMR visual related genes were missing or inactivated by frameshift mutation or premature termination (Table 3); and a gene family contraction in the BMR beta/gamma crystalline cluster was also observed. These results potentially linked with subterranean rodents’ degeneration of vision.

Table 2. Visual related genes that are inactivated or are missing in the NMR genome[37].

<table>
<thead>
<tr>
<th>Gene</th>
<th>Inactivation event</th>
<th>Time of gene loss</th>
<th>(\omega_0) (average)</th>
<th>(\omega_1) (other)</th>
<th>(\omega_2) (NMR)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBP3</td>
<td>F</td>
<td>NMR</td>
<td>0.121</td>
<td>0.085</td>
<td>0.232</td>
<td>1.187E-11</td>
</tr>
<tr>
<td>ARR3</td>
<td>F/S</td>
<td>NMR</td>
<td>0.420</td>
<td>0.260</td>
<td>0.912</td>
<td>6.263E-06</td>
</tr>
<tr>
<td>PDE6C</td>
<td>F</td>
<td>NMR</td>
<td>0.171</td>
<td>0.139</td>
<td>0.316</td>
<td>0.0001</td>
</tr>
<tr>
<td>GUCA1B</td>
<td>F</td>
<td>NMR</td>
<td>0.083</td>
<td>0.056</td>
<td>0.217</td>
<td>0.001</td>
</tr>
<tr>
<td>GJA10</td>
<td>F/S</td>
<td>NMR</td>
<td>0.308</td>
<td>0.248</td>
<td>0.524</td>
<td>0.002</td>
</tr>
<tr>
<td>GUCY2E</td>
<td>F</td>
<td>NMR</td>
<td>0.124</td>
<td>0.105</td>
<td>0.182</td>
<td>0.002</td>
</tr>
<tr>
<td>CRYBA4</td>
<td>S</td>
<td>NMR</td>
<td>0.055</td>
<td>0.036</td>
<td>0.123</td>
<td>0.001</td>
</tr>
<tr>
<td>GNAT2</td>
<td>F</td>
<td>NMR</td>
<td>0.055</td>
<td>0.039</td>
<td>0.108</td>
<td>0.017</td>
</tr>
<tr>
<td>SLC24A1</td>
<td>F</td>
<td>NMR</td>
<td>0.389</td>
<td>0.355</td>
<td>0.517</td>
<td>0.035</td>
</tr>
<tr>
<td>CRYBB3</td>
<td>S</td>
<td>NMR</td>
<td>0.071</td>
<td>0.048</td>
<td>0.122</td>
<td>0.037</td>
</tr>
<tr>
<td>RP1L1</td>
<td>S</td>
<td>NMR</td>
<td>0.448</td>
<td>0.424</td>
<td>0.513</td>
<td>0.186</td>
</tr>
<tr>
<td>GRK7</td>
<td>F</td>
<td>NMR</td>
<td>0.154</td>
<td>0.135</td>
<td>0.201</td>
<td>0.335</td>
</tr>
<tr>
<td>PDE6H</td>
<td>F</td>
<td>NMR</td>
<td>0.091</td>
<td>0.082</td>
<td>0.127</td>
<td>0.648</td>
</tr>
<tr>
<td>EYS</td>
<td>F/S</td>
<td>Ancestor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GUCA1C</td>
<td>F</td>
<td>Ancestor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OPN1LW</td>
<td>L</td>
<td>NMR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OPN1MW</td>
<td>L</td>
<td>NMR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PRCD</td>
<td>L</td>
<td>NMR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RD3</td>
<td>L</td>
<td>NMR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: F, frameshift; S, premature stop codon; L, gene loss with synteny region. Ancestor/NMR indicates that the gene was lost in a rodent ancestor or the NMR, respectively.
Figure 7. Inactivation events observed within NMR visual perception pseudogenes [37]. Gene models are drawn according to the mouse or human ortholog counterpart. Green squares represent exons and blue lines introns. Red arrows indicate inactivation events, such as insertion or deletion that change the reading frame, or point mutations leading to premature termination.
Table 3. Visual related genes that are inactivated or are missing in the BMR genome [9].

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Gene name</th>
<th>Kegg class</th>
<th>F/S</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPN1SW</td>
<td>opsin 1 (cone pigments), short-wave-sensitive</td>
<td>Signaling Molecules and Interaction;</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>IMPG1</td>
<td>interphotoreceptor matrix proteoglycan 1</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>CRYBA1</td>
<td>crystallin, beta A1</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>BFSP2</td>
<td>beaded filament structural protein 2, phakinin</td>
<td>Cell Motility;</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>BEST1</td>
<td>bestrophin 1</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>CABP4</td>
<td>calcium binding protein 4</td>
<td>NA</td>
<td>S</td>
<td>Yes</td>
</tr>
<tr>
<td>RBP3</td>
<td>retinol binding protein 3, interstitial</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>CHML</td>
<td>choroideremia-like (Rab escort protein 2)</td>
<td>NA</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>CACNA2D4</td>
<td>calcium channel, voltage-dependent, alpha 2/delta subunit 4</td>
<td>Circulatory System; Signaling Molecules and Interaction; Cardiovascular Diseases; Signal Transduction;</td>
<td>F</td>
<td>Yes</td>
</tr>
<tr>
<td>ALMS1</td>
<td>Alstrom syndrome 1</td>
<td>NA</td>
<td>F</td>
<td>Yes</td>
</tr>
<tr>
<td>C2orf71</td>
<td>chromosome 2 open reading frame 71</td>
<td>NA</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>RGR</td>
<td>retinal G protein coupled receptor</td>
<td>Signaling Molecules and Interaction;</td>
<td>F</td>
<td>Yes</td>
</tr>
<tr>
<td>CRYBB1</td>
<td>crystallin, beta B1</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>CNGB3</td>
<td>cyclic nucleotide gated channel beta 3</td>
<td>Signaling Molecules and Interaction;</td>
<td>S</td>
<td>Yes</td>
</tr>
<tr>
<td>CRYBA4</td>
<td>crystallin, beta A4</td>
<td>NA</td>
<td>S</td>
<td>No</td>
</tr>
<tr>
<td>PPEF2</td>
<td>protein phosphatase, EF-hand calcium binding domain 2</td>
<td>NA</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>GJA1</td>
<td>gap junction protein, alpha 8, 50kDa</td>
<td>NA</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>SLC24A1</td>
<td>solute carrier family 24 (sodium/potassium/calcium exchanger), member 1</td>
<td>Sensory System; Nucleotide System; Metabolism;</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>GUCT2F</td>
<td>guanylate cyclase 2F, retinal</td>
<td>Metabolism of Cofactors and Vitamins;</td>
<td>F&amp;S</td>
<td>No</td>
</tr>
<tr>
<td>RDH5</td>
<td>retinol dehydrogenase 5 (11-cis/9-cis)</td>
<td>Metabolism of Cofactors and Vitamins;</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>RDH8</td>
<td>retinol dehydrogenase 8 (all-trans)</td>
<td>Metabolism of Cofactors and Vitamins;</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>ABCA4</td>
<td>ATP-binding cassette, sub-family A (ABC1), member 4</td>
<td>Membrane Transport</td>
<td>F&amp;S</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: F: frameshift; S: premature stop; A/S: ‘Yes’ means this gene contains some alternative splicing forms to avoid the mutation sites; ‘No’ means this gene contains no alternative splicing forms to avoid the mutation sites.

4.3 Circadian rhythm

Melatonin serves as a regulator of body temperature and circadian rhythm [50]. Normally, periodic light/dark switches affect synthesis of melatonin, thus modulates circadian rhythms in mammals. The fact that subterranean rodents live in naturally dark underground burrows make it interesting to study their melatonin signalling.

Previous study showed that the pineal glands of NMR, which synthesizes and secretes melatonin, are atrophied [51]. However, all the genes involved in melatonin synthesis, including TPH1, TPH2,
**DDC, AANAT** and **ASMT**, were detected with intact gene structures. It's worth noting that that genes **AANAT** and **ASMT**, which involved in the last two steps of melatonin synthesis, showed extremely low expression levels in the NMR liver, kidney and brain at any age (Table 4 and Figure 8). What’s more, two major melatonin receptors, **MTNRI A** and **MTNRI B**, which encode MT1 and MT2, respectively, were inactivated by early termination events (Figure 9). It seems that although the melatonin signalling was disrupted, the NMR could maintain its circadian rhythms in accordance with locomotor activity and body temperature when exposed to periodic light/dark changes[52]. This result is in conformity with a previous report that MT1/MT2 knockout mice could maintain normal circadian rhythms [53].

Further comparative study revealed that although both the DMR and NMR lost **MTNR1b**, the inactivating mutations are located in different loci (Figure 10B). It is also of interest that **MTNR1a** is intact in the DMR assembly, but inactivated in the NMR (Figure 10A).

**Table 4.** RPKM of genes involved in melatonin synthesis of liver, kidney and brain for newborn (0), 4-year-old (4) and 20-year-old (20) NMRs [37].

<table>
<thead>
<tr>
<th>Genes</th>
<th>Liver (age groups)</th>
<th>Kidney (age groups)</th>
<th>Brain (age groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td><strong>TPH1</strong></td>
<td>0.05</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>TPH2</strong></td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>DDC</strong></td>
<td>3.97</td>
<td>20.88</td>
<td>13.67</td>
</tr>
<tr>
<td><strong>AANAT</strong></td>
<td>1.76</td>
<td>0.9</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>ASMT</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 8.** Compounds, genes and enzymes required for melatonin biosynthesis [37].
Figure 9. Inactivation of melatonin receptors, \textit{MTNR1A} and \textit{MTNR1B}, in the NMR genome [37]. (A) Premature stop codons (red) in the NMR \textit{MTNR1A} and \textit{MTNR1B}. (B) Topology of human melatonin receptors and the locations of stop codons in the NMR protein.
**Figure 10.** Comparison of melatonin receptors, *MTNR1a* and *MTNR1b*, in NMR and DMR. (A) *MTNR1a* is of intact sequence in the DMR, but inactivated in NMR (Figure 8). Alignment of DMR and other mammalian *MTNR1a* nucleotide sequences is shown. Identical residues are shaded in black. (B) *MTNR1b* is inactivated in both DMR and NMR, as an evidence of convergent evolution. Alignment of rodent and human *MTNR1b* nucleotide sequences is shown. Identical residues are shaded in black, and premature stop codons are highlighted in red.
4.4 Hypercapnia, hypoxia, and ammonia tolerance

Subterranean rodents rest with conspecifics in underground environments low in O2 and high in CO2 and ammonia (a potent irritant [54,55]). The very harsh living conditions could trigger cellular damage, as well as physiological and behavioral stress responses in other mammals [56].

Protein ARG1 (arginase 1) can catalyzes the last step of the hepatic urea cycle, so as to removes ammonia from the body. Unique residue change (ARG1: His254 replaces Leu/Tyr) in both the NMR and DMR were observed, although the wide type is present in 38 other vertebrate species (Figure 11A). This amino acid change was also detected in the distantly related subterranean coruro (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 [57,58] (Figure 11B). His254 is located immediately downstream of a conserved motif required for binding manganese and ARG1 function [59] (Figure 11C). Moreover, ARG1 is a homotrimer, with the salt bridges formed by Arg255 and Glu256 being critical for its assembly [60,61]. 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 (Figure 11D). 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[62], was elevated in the NMR and DMR. Taken together, these data indicate that subterranean hystricognath rodents present enhanced ammonia detoxification.
**Figure 11.** Subterranean Adaptations in Hystricognath Rodents [49]. (A) Subterranean hystricognath rodents share a charged residue at position 254 of arginase 1 (ARG1). The manganese-binding site, residues critical for enzyme trimer assembly (Arg255 and Glu256), and unique His254 changes are highlighted in purple, orange, blue, and red, respectively. Identical residues in vertebrates are shaded in black. (B) Phylogenetic relationship of hystricognath rodent lineages examined in this study. Approximate divergence times (Myr) are indicated. (C) Structural model of human ARG1 monomer. Residues are highlighted as in (A). (D) Schematic representation of the roles of components of the urea cycle with altered sequence (purple box) or expression in NMR and DMR (green boxes). (E) Species of hypercapnic habitats share a negatively charged three-residue motif in the Na(V) 1.7 sodium channel protein. Acidic amino acid residues in the motif, corresponding to amino acids 1718 and 1720 of the human sequence, are shown in red. Identical residues in vertebrates are shaded in black. (F) Heatmap of globin expression in normoxic...
rodent brains. Scaled log2 transformed normalized read counts (denoted as the row Z score) are plotted in beige–blue color, with blue indicating high expression and beige indicating low expression. B. guinea pig, Brazilian guinea pig; HBA1/2, hemoglobin a; NGB, neuroglobin; CYGB, cytoglobin. Red stars indicate differentially expressed genes in subterranean rodents. (G) Western blot of hemoglobin a in normoxic rodent brains with antibodies against the mouse protein. (H) Comparison of globin gene expression under normoxia (21% O2) and hypoxia (8% O2 over 8 hr). Annotated as in (F).

Although there are compensation mechanisms reported [63,64] to clear excess CO2, underground rodents NMR, DMR and BMR tissues would subject to acidosis and acidification [65]. According to our study, BMR contains a mutation in gene *Scn9a*, which encodes the proton-gated nociceptor sodium channel Nav1.7. The mutation results in the replacement of a highly conserved, positively charged amino acid motif (KKV) in domain IV of Nav1.7 by the negatively charged EKD motif, attracting protons thereby blocking the channel and thus protecting the BMR tissue from acid pain [66]. Interestingly, a similar mutation, resulting in a negatively charged EKE motif, was found in the NMR and DMR Nav1.7. And the same Nav1.7 mutation was also identified in the distantly related, cave-dwelling microbat *Myotis lucifugus*, which also encounters hypercapnia living conditions [66]. A phylogenetic interpretation of Nav1.7 sequence evolution in mammals (Figure 12) indicates that the pain-blocking mutation is an adaptive trait, which has arisen independently in these species by convergent evolution.

![Figure 12. Evolution of the adaptive amino-acid sequence motif from the sodium channel nociceptor protein Nav1.7 in mammals [9]. The pain-blocking mutation is an evidence of adaptive evolution in hypercapnia-exposed BMR, DMR, NMR and the cave microbat (*Myotis*).](image)
4.5 Longevity

Life spans change significantly among rodents, range from 3–4 years in mice and rats, to over 20 years in DMRs and BMRs, and more than 30 years in NMRs. Subterranean rodents have the highest maximum lifespans for their body weight [67]. These rodents have a longevity quotient similar to that of humans and may show a comparable age-related disease pattern [68].

For NMR, we identified positive selection signals on gene TEP1, which encoded a telomerase component, and gene TERF1, a telomeric repeat binding factor involved in protecting telomeres [69] and regulating telomere length [70]. TOP2A, along with TEP1 and TERF1 from the set of positively selected genes, are part of a five-protein complex of alternate lengthening of telomere pathway [71]. By analyzing transcriptome data, it is also interesting that TERT, the telomerase reverse transcriptase, displayed stable expression level at any age (Figure 14). Overall, both transcriptomic and genomic data suggest that altered telomerase function in the NMR potentially associate with its longevity and anti-cancer.

![Figure 13. NMR-specific amino acid change in TERF [37]. Asterisks (*) indicate residues involved in telomere binding in human TRF1 protein encoded by TERF1. The Ala75Pro mutation in the human protein is known to inhibit dimerization of TRF1 and telomere binding. The same amino acid changed in the NMR sequence.](image13.png)

![Figure 14. Expression of TERT in liver, kidney and brain of 4-year-old and 20-year-old NMRs. TERT showed stable expression at any age in NMR [37.](image14.png)
4.6 Cancer resistance

Although both BMR and NMR showed remarkable cancer resistance, they may follow divergent molecular strategies. The NMR mostly relies on HMM-HA to mediate early contact inhibition to avoid cancer. In addition, altered telomerase function by TOP2A, TEP1 and TERF1 may also contribute to its longevity and cancer resistance.

By contrast, the BMR utilizes an unique tumour suppression mechanism, in which necrosis takes a more important role than apoptosis. We observed two genes from the interferon signalling pathway (Ifnb1 and Mx1), as well as multiple genes involved in regulation of cell death and inflammation (Nfkb, Tnfrsf1a, Birc3, Fem1b and Aifs1), underwent copy number duplications in the BMR. Furthermore, three genes involved in necrosis and inflammation (Tnfrsf1a, Tnfsf15 and Nfkb1) show evidence of positive selection (Tables 5 and 6, Figure 15). Our findings fit the following model: BMR specific mutations on gene p53 weakened its tumour suppressor activity, and weakened its suppression on TRAIN (transcription of repeats activates interferon)-mediated releasing of suicidal interferon. As a result, the increased necrosis evoked by releasing of interferon serve as a highly effective compensatory mechanism to complement insufficient p53-mediated tumour suppression [72].

Table 5. The potential positively selected cancer-related genes with branch model [9].

<table>
<thead>
<tr>
<th>Species</th>
<th>Symbol</th>
<th>Gene</th>
<th>oθ (average)</th>
<th>o1 (other)</th>
<th>o2 (target)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMR</td>
<td>TNFSF10</td>
<td>tumor necrosis factor (ligand) superfamily</td>
<td>0.30884</td>
<td>0.27198</td>
<td>0.62821</td>
<td>2.40E-02</td>
</tr>
<tr>
<td></td>
<td>NGFRAP1</td>
<td>nerve growth factor receptor (TNFSF16)</td>
<td>0.10377</td>
<td>0.07424</td>
<td>0.40086</td>
<td>1.55E-02</td>
</tr>
<tr>
<td></td>
<td>PYGL</td>
<td>phosphatase, glycogen, liver</td>
<td>0.05611</td>
<td>0.0479</td>
<td>0.09986</td>
<td>1.08E-02</td>
</tr>
<tr>
<td></td>
<td>MAG</td>
<td>myelin associated glycoprotein</td>
<td>0.04731</td>
<td>0.03914</td>
<td>0.08381</td>
<td>8.00E-03</td>
</tr>
<tr>
<td></td>
<td>CAPN5</td>
<td>calpain, small subunit 1</td>
<td>0.07805</td>
<td>0.02881</td>
<td>0.21947</td>
<td>4.04E-07</td>
</tr>
<tr>
<td></td>
<td>RAB25</td>
<td>RAB25, member RAS oncogene family</td>
<td>0.05311</td>
<td>0.03558</td>
<td>0.09442</td>
<td>2.20E-02</td>
</tr>
<tr>
<td></td>
<td>CAPN6</td>
<td>calpain 6</td>
<td>0.0748</td>
<td>0.05435</td>
<td>0.16796</td>
<td>1.95E-04</td>
</tr>
<tr>
<td>BMR</td>
<td>NFKB1</td>
<td>nuclear factor of kappa light polypeptide 1</td>
<td>0.10003</td>
<td>0.08851</td>
<td>0.20265</td>
<td>4.36E-04</td>
</tr>
<tr>
<td></td>
<td>HSPB4P1</td>
<td>HSPB (heat shock 27kDa) associated protein 1</td>
<td>0.30927</td>
<td>0.27818</td>
<td>0.53635</td>
<td>1.57E-02</td>
</tr>
</tbody>
</table>

Table 6. The potential positively selected cancer-related genes with branch site model [9].
<table>
<thead>
<tr>
<th>Species</th>
<th>Symbol</th>
<th>Gene</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>P48</td>
<td>Fas (TNF receptor superfamily, member 6)</td>
<td>2.12E-02</td>
</tr>
<tr>
<td></td>
<td>PYGL</td>
<td>phosphatase, glycogen, liver</td>
<td>3.17E-02</td>
</tr>
<tr>
<td></td>
<td>TNFRSF1A</td>
<td>tumor necrosis factor receptor superfamily, member 1A</td>
<td>4.21E-02</td>
</tr>
<tr>
<td></td>
<td>PaRP2</td>
<td>poly (ADP-ribose) polymerase 2</td>
<td>8.51E-03</td>
</tr>
<tr>
<td></td>
<td>CAPN1</td>
<td>calpain 1, (mu1) large subunit</td>
<td>3.13E-02</td>
</tr>
<tr>
<td></td>
<td>CD40</td>
<td>CD40 molecule, TNF receptor superfamily member 5</td>
<td>7.75E-03</td>
</tr>
<tr>
<td></td>
<td>FUS</td>
<td>fused in sarcoma</td>
<td>3.55E-02</td>
</tr>
<tr>
<td></td>
<td>PVR</td>
<td>poliovirus receptor</td>
<td>4.28E-02</td>
</tr>
<tr>
<td>BMR</td>
<td>TNFRSF1A</td>
<td>tumor necrosis factor receptor superfamily, member 1A</td>
<td>2.39E-02</td>
</tr>
<tr>
<td></td>
<td>NFKB1</td>
<td>nuclear factor of kappa light polypeptide gene enhancer in B-cells 1</td>
<td>2.62E-03</td>
</tr>
<tr>
<td></td>
<td>TNFRSF15</td>
<td>tumor necrosis factor (ligand) superfamily, member 15</td>
<td>1.67E-09</td>
</tr>
<tr>
<td></td>
<td>HSPBAP1</td>
<td>HSPB (heat shock 27kDa) associated protein 1</td>
<td>7.68E-04</td>
</tr>
</tbody>
</table>

Figure 15. BMR adaptive complex related with hypoxia tolerance and anti-cancer [9]. The colours of genes Ccnd2, Ccng1, Ccnb1, Cdk4, Cyc, Casp9, Irf7 and B1 SINEs represent transcriptome expression changes according to RPKM, while colours of Mx1, Birc3, Ifnb1, A1fm1, Nfkb1, Tnfrsf1a and Fem1B indicate gene copy number amplification. The green check marks on Tnfrsf1a and Nfkb1 represent positive selection signal.
4.7 Other traits

Hairless of NMR

Although NMRs have about 100 sensory hairs around their bodies that serve as whiskers to sense their surroundings, they represent the only mole rat species that naturally has no fur (or “naked”). According to previous studies, the nuclear receptor co-repressor, Hairless (Hr), potentially involved in the molecular pathways of hair growth. Mutations in Hr result in hair loss in mice, rat and human. Our analyses of NMR Hr revealed substantial divergence of this protein from known mammalian orthologs; the presence of NMR specific mutations potentially associated with its hairless phenotype (Figure 16).

Figure 16. Hairless homolog (Hr) of NMR [37]. (A) The Neighbor-Joining (NJ) tree demonstrating relationships between the predicted NMR Hr homolog protein and ten Hr proteins from other mammals. (B) Schematic representation of human HR protein functional domains. (C) Protein alignment of Hr proteins from NMR and ten other mammals. The position of C397Y mutation associated with rat hairless phenotype is indicated on the top. In the NMR sequence, this Cys is replaced with Trp.
Thermoregulation of NMR

The NMR does not regulate its body temperature like other mammals. More specifically, they are thermoconformers rather than thermoregulators, with body temperature change with ambient temperatures. Therefore, NMRs have to use behavioral thermoregulation at lower or higher temperatures. For example, NMRs tend to crowd together in more sun-warmed parts of the burrows when cold; conversely, they tend to stay in the deeper, cooler parts of the burrows when hot.

According to our genomic study, we observed unique mutations in protein UCP1, which encodes mitochondrial uncoupling protein 1, mediating non-shivering thermogenesis (Figures 17 and 18). Since the major heat production process is mainly regulated by UCP1, the NMR unique mutations suggest a tight association of UCP1 function with the specific thermoregulation pattern of the NMR.

Figure 17. NMR-Specific mutations in UCP1 protein sequences and their roles in thermoregulation [37]. a, Alignment of UCP1 sequences of NMR, Guinea pig, mouse and human. NMR-Specific amino acids are highlighted in red, and conserved motifs in blue. b, Topology of UCP1. Regions affected by mutations in the NMR are highlighted. c, Structural model of UCP1. Location of the channel and the nucleotide binding loop with altered sequences in the NMR are shown.
Pain insensitivity of NMR

Pain insensitivity is another unique features of NMRs. Based on pervious studies, the skin of NMR naturally lacks an important neurotransmitter, substance P (SP), which is involved in sending pain signals to the central nervous system (CNS) in mammals. When injected with SP, however, the NMR pain signaling works properly, but only with capsaicin or acids stimuli. This absence of SP was proposed to be an adaptation to the hypercapnia living condition, which would cause acid accumulation in the NMR body tissues[7]. NMRs’ SP deficiency has also been related with their lack of the histamine-induced itching and scratching behavior [73].

Gene TAC1 encodes SP in mammals. In our study, we observed an intact TAC1 sequences in NMR. However, there is a deletion in the TAC1 core promoter region, which was highly conserved among mammals (Figure 19). Our finding indicates that the synthesis of neurotransmitter SP appears to be functional, but may be under unique regulation pattern, leading to NMR’s pain insensitivity.

Figure 19. A deletion within the promoter of gene TAC1 is unique to the NMR and is not present in the related African mole rat, the DMR [49]. Identical residues are shaded in black, and the deletion region in NMR is highlighted in red.
5 Conclusions

Stress is an evolutionary driving force of great importance. Subterranean rodents are a widely distributed group of species that live in underground tunnels most of their lives. The subterranean ecotope protects them to escape from aboveground climatic fluctuations and predators, while challenge them with multiple stresses including light deficiency, hypoxia, hypercapnia, food shortage, and pathogenicity. In all continents excepting Australia and Antarctica, about 250 extant rodent species primarily live underground and therefore share several common traits, including highly developed sensory except vision, remarkable tolerance to hypoxia and hypercapnia, similar energy metabolism pattern (such as low body temperature and low metabolic rate), etc. In the meantime, every species show divergent evolutions in adapt to their phylogenies and separated feeding niches. The evolutionary biology of subterranean rodents provides excellent animal models to elucidate how the organisms evolutionarily adapt to harsh environments. What’s more, given the remarkable features of certain subterranean rodents, such as longevity, anti-cancer, hypoxia tolerance, comparing them with short-lived, cancer-prone, hypoxia susceptible aboveground rodents at the molecular level, would greatly widen our understanding of disease and aging, and thus benefit future biomedical applications.

In this thesis, we sequenced the genomes and transcriptomes of three subterranean rodents: the naked mole rat (NMR, *Heterocephalus glaber*), the Damaraland mole rat (DMR, *Fukomys damarensis*) and the blind mole rat (BMR, *Spalax galili*), not only because they are subterranean representatives that achieve extraordinary success in flourishing underground, but also due to their medically useful features in anti-aging and cancer resistance. By comparing subterranean rodents NMR, DMR and BMR with the well-known aboveground rodents mouse (*Mus musculus*) and rat (*Rattus norvegicus*), we discussed the shared adaptions and divergent evolutions among them. We characterized the genomic features of subterranean rodents, as well as their specific transcriptomic regulation pattern in response to hypoxia, aging, or inducing tumorigenesis. The extreme traits of the NMR, DMR and BMR, together with the reported genomic and transcriptomic information, provided opportunities for understanding convergent and divergent adaptations, and offered new insights for lifespan extension and cancer prevention in human.
6 Perspective

Subterranean rodents offer us great animal models to investigate adaptive evolution to environment stresses. The results of the thesis lead to new questions for possible further research; some of the questions are still only ideas, and some have already been worked on.

Transposable elements regulation

TEs in different mammalian lineages display various distribution patterns. For example, Alu SINE expanded in human and other primate genomes, ID SINE increased in NMR and rats genomes, while B4 SINE amplified in the genome of Lesser Egyptian jerboa *Jaculus jaculus*.

In this thesis, we observed an B1/B2 SINE copy number expansion in BMR, as well as the transcription activity up-regulation of B1 SINE upon hypoxia. These finding providing us insight into adaptive evolutionary force affecting the transposon insertion pattern. However, whether TE expansions were adaptive and what were the evolutionary stresses related with them is still poorly understood.

Genetics of social organization

Life underground has been proposed either to favor solitariness or, to the contrary, to promote sociality. In concordance with the first idea, most specialized diggers are solitary, such as BMR. However, group-living in NMR and DMR seem to support the second hypothesis. Therefore, subterranean rodents provide an ideal model system to examine adaptive evolution of social organization.

In this study, we reported functional enrichment of olfaction gain genes likely play an important role in social interaction. More sophisticated regulation mechanism is still poorly understood. We will carry out more intensive investigations to address the genetic, transcriptomic, and epigenetic mechanisms that control social organization.

Genetics of longevity and cancer

Life span of rodents varies widely across species. Although phylogenetically related, lifespans range from 3-4 years in rats and mice, to over 20 years in DMRs and BMRs, and exceeding 32 years in NMRs. Cancer rates are also extremely diverse, from cancer-prone mice to cancer-resistant NMRs and BMRs. With the accumulation of whole genome sequencing data among rodents, recent progress in comparative genomics offers an inimitable opportunity for the discovery of molecular mechanisms that underlie anti-aging and cancer resistance.

Different cancer-resistant species employ distinct tumor suppressor strategies. As we have mentioned, the NMR mostly relies on HMM-HA to mediate early contact inhibition to avoid cancer, while the BMR uses the interferon-mediated cell death mechanism[9]. Interestingly, other rodents are likely to harbor additional novel tumor suppressor mechanisms. For example, the long-lived, aboveground diurnal rodent, grey squirrel, show remarkable high telomerase activity and slow proliferation rate in its tissue culture, indicating a novel cell cycle control mechanism that is still poorly undertand. In addition, it is reported that rodents with high body mass (e.g., >10 kg) have unique adaptations associated with telomerase activity and anti-aging. As evolutionary path for longevity and anticancer adaptations is largely depend on unique species ecology, the investigation of such species-specific adaptations would promote the subsequent application of these strategies in humans.
7 References


8 Appendix


Paper I

Genome sequencing reveals insights into physiology and longevity of the naked mole rat

Eun Bae Kim1*, Xiaodong Fang2*, Alexey A. Fushan1*, Zhiyong Huang2*, Alexei V. Lobanov3, Lijuan Han2, Stefano M. Marino3, Xiaqing Sun2, Anton A. Turanov3, Pengcheng Yang2, Sun Hee Yim3, Xiang Zhao2, Marina V. Kasaikina3, Nina Stoletzki3, Chunfang Peng1, Paz Polak1, Zhiqiang Xiong2, Adam Kiezun3, Yabing Zhu2, Yuanxin Chen2, Gregory V. Kryukov3,4, Qiang Zhang2, Leonid Peshkin5, Lan Yang2, Roderick T. Bronson6, Rochelle Buffenstein7, Bo Wang2, Changlei Han2, Qiye Li2, Li Chen2, Wei Zhao2, Shamil R. Sunyaev3,4, Thomas J. Park8, Guojie Zhang2, Jun Wang2,9,10 & Vadim N. Gladyshev1,3,4

1 Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, Korea.
2 BGI-Shenzhen, Shenzhen, 518083, China.
3 Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.
4 Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.
5 Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.
6 Rodent Histopathology Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA.
7 Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, Texas 78245, USA.
8 Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA.
9 Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.
10 Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.
*These authors contributed equally to this work.

The naked mole rat (Heterocephalus glaber) is a strictly subterranean, extraordinarily long-lived eusocial mammal. Although it is the size of a mouse, its maximum lifespan exceeds 30 years making this animal the longest-living rodent. Naked mole rats show negligible senescence, no age-related increase in mortality, and high fecundity until death. In this manner of living, they are resistant to both spontaneous cancer and experimentally induced tumorigenesis.

Naked mole rats pose a challenge to the theories that link aging, cancer and redox homeostasis. Although characterized by significant oxidative stress, the naked mole rat proteome does not show age-related oxidative susceptibility to oxidative damage or increased ubiquitination. Naked mole rats naturally reside in large colonies with a single breeding female, the 'queen', who suppresses the sexual maturity of her subordinates. They also live in full darkness, at low oxygen and high carbon dioxide concentrations, and are unable to sustain thermogenesis nor feel certain types of pain.

Here we report the sequencing and analysis of the naked mole rat genome, which reveals unique genome features and molecular adaptations consistent with cancer resistance, poliklothery, hairlessness and insensitivity to low oxygen, and altered visual function, circadian rhythms and taste sensing. This information provides insights into the naked mole rat's exceptional longevity and ability to live in hostile conditions, in the dark and at low oxygen. The extreme traits of the naked mole rat, together with the reported genome and transcriptome information, offer opportunities for understanding aging and advancing other areas of biological and biomedical research.

We applied a whole genome shotgun strategy to sequence the genome of an individual male naked mole rat (NMR) (Table 1 and Supplementary Tables 1–3). The sequencing depth of 98.6% of the genome assembly was more than 20-fold (Supplementary Figs 1–4).

The mitochondrial genome was also assembled. Approximately 25% of the NMR genome was represented by transposon-derived repeats, which is lower than in other mammals (40% in human, 37% in mouse, and 35% in rat genomes) (Supplementary Tables 4 and 5, Supplementary Figs 5–7). The predicted NMR gene set included 22,561 genes (Table 1 and Supplementary Table 6), which is comparable to other mammals (22,389 in human, 23,317 in mouse, and 22,841 in rat). Of these, 21,394 (94.8%) genes were transcribed (based on the RNAseq data for seven organs). More than 98% of NMR genes could be functionally annotated using homology approaches (Supplementary Table 7), and the quality of predicted genes was comparable to that of well-annotated mammalian genomes (Supplementary Tables 8 and 9 and Supplementary Fig. 8).

Most of the NMR genome (93%) showed synteny to human, mouse or rat genomes (Supplementary Table 9), and pairwise comparisons suggested a relatively low rate of NMR genome rearrangements after the split from the murid common ancestor. We defined common syntenic blocks in human, mouse, rat and NMR genomes and identified segmental duplications and lineage-specific insertions and deletions (Supplementary Tables 10 and 11 and Supplementary Fig. 9). By analyzing single-copy orthologous groups, we constructed a phylogenetic tree involving the NMR and other mammals (Fig. 1). As expected, the NMR placed within rodents and its ancestor split from the ancestor of rats and mice approximately 73 million years ago, whereas the ancestor of NMR, mouse and rat split from rabbits approximately 86 million years ago. Thus, in spite of some exceptional traits, the overall properties of the NMR genome appeared to be similar to those of other mammals.

Lineage-specific gene family expansions may be associated with the emergence of specific functions and physiology. Compared to other mammals, the NMR showed a moderate number of gene families under expansion and contraction (Fig. 16), including 96 NMR lineage-specific gene families (Fig. 2). Analysis of syntenic regions identified 376 and 320 lost NMR genes (Supplementary Tables 12–14). At least 75.3% of genes gained showed evidence of transcription, and the lost genes were enriched for ribosome and nucleoid biosynthesis functions (Supplementary Table 15). We also identified 244 pseudogenes, containing 183 frameshift and 119 premature termination events (Supplementary Tables 16 and 17). Functional categories enriched for pseudogenes included olfactory receptor activity (GO:0004894, P < 0.001, Fisher's exact test), genes involved in the perception (GO:0007001, P < 0.015, GRI, CRYR5, GN5, KGR7, GCGR and FDERG), and speech intelligibility genes (GO:0007585, P < 0.001, GRH1, CRYR3, CRYR5, CRYR7, GNAI3, CAV1 and FDR6G).

Table 1 | Global statistics of the NMR genome

| Sequencing | Insert size (bp) | Total data (Gbp) | Sequence coverage (fold)
| --- | --- | --- | ---
| Paired end | 170–800 | 126.52 | 47 |
| Libraries | 2–20 × 10^6 | 120.66 | 45 |
| Total | 247.18 | | 92 |
| Assembly | N50 (kb) | Longest (kb) | Size (Gb) |
| Contigs | 19.3 | 179 | 2.45 |
| Scaffolds | 1.885 | 7.787 | 2.66 |

| Annotation | Number | Total length (Mb) | Percentage of the genome |
| --- | --- | --- | ---
| Repeats | 3,090,116 | 666.7 | 25 |
| Genes | 22,561 | 722.8 | 27.1 |
| CDS | 181,641 | 32.5 | 1.2 |

1Department of Behavioral Sciences, Emory University, Seoul, 120-720, Korea. 2Department of Ophthalmology, Shanghai Jiao Tong University, Shanghai, P.R. China. 3Department of Anatomy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. 4Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. 5Department of Ophthalmology and Visual Sciences, University of Texas Health Science Center, San Antonio, Texas, 78229, USA. 6Department of Health and Social Policy, University of Helsinki, P.O. Box 6007, FIN-00014, Helsinki, Finland. 7Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark. 8Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.

*These authors contributed equally to this work.
traits. 45 genes (0.4%) were identified as positively selected in the NMR lineage at the false discovery rate of 0.01 and 141 genes (1.2%) at the false discovery rate of 0.05 (Supplementary Table 18). 12 out of the 45 genes (corresponding to the false discovery rate of 0.01) passed a strict manual inspection for alignment quality. In comparison, 0.7% of genes were predicted to be positively selected in the human lineage from high-quality alignments and using Rm correction for multiple testing11. Interestingly, our set included TEP1, encoding a telomerase component, and TERFI, a telomeric repeat binding factor identified at the false discovery rate of 0.05 (Supplementary Fig. 10). The TERFI gene product is one of six proteins contributing to the shelterin complex, which shapes and protects telomeres12 and has been proposed to regulate telomere length13.

To gain further insights into biological processes that underlie the exceptional traits of the NMR, we identified 39 NMR proteins containing 45 amino acid residues unique among orthologues present in 36 vertebrate genomes (Supplementary Table 19). This gene set included cyclin E1 (CCNE1), uncoupling protein 1 (UCPI) and γ-cristallin (CRYGS), which are associated with the G1/S transition during the cell cycle, thermogenesis and visual function, respectively. Other noteworthy genes were APEX1, a multifunctional DNA repair enzyme, RFC1, replication factor C, and TOP2A, a DNA topoisomerase that controls the topologic states of DNA during transcription. This set also contained eight genes designated as cancer-related14. Finally, TOP2A, along with TEP1 and TERFI, originated from a positively selected gene, are part of a five-protein complex of alternate lengthening of telomere pathway15. Overall, these analyses point to altered telomerase function in the NMR, which may be related to its evolution of extended lifespan and cancer resistance.

We also identified 1.87 million heterozygous single-nucleotide polymorphisms (SNPs). This results in an estimated nucleotide diversity (mean per nucleotide heterozygosity) of 7 × 10^-4, which is much lower than in mouse and rat populations and is comparable to the nucleotide diversity observed in humans. Transition nucleotide changes were observed twice as often as transversions, indicating that variant calls reproduce the expected properties of natural variation in other mammals. This low level of nucleotide diversity may reflect a low effective size of NMR population, but may also be due to a high level of inbreeding, a reduced mutation rate or high efficiency of the repair systems. The variation of diversity along the genome was consistent with inbreeding in the NMR population. In protein-coding regions of the genome, our analysis identified 10,951 non-synonymous and 8,616 synonymous SNPs. Their ratio is much higher than in other studied organisms, including humans, which appears to signal relaxation of purifying selection in the NMR, potentially as a consequence of reduced effective population size. Finally, we analysed the context dependency of NMR SNPs (Supplementary Fig. 11). Relative rates of nucleotide changes and nucleotide context dependencies were similar to those observed in human polymorphism, with the exception of a relative reduction of SNPs due to CpG mutations. This was caused by a combination of the relatively low CpG density in the NMR genome and a higher fraction of CpG dinucleotides within CpG islands compared to the human genome. CpG density was only 0.19 of that expected on the basis of the GC content, which is lower in humans, dog and panda genomes, but is similar to the mouse genome. However, in comparison to mouse, a higher fraction of CpG dinucleotides was concentrated in CpG islands. CpG dinucleotides within CpG islands contribute less to genetic variation because of their lower methylation rate and possibly also due to selection.

Long lifespan is a key feature of the NMR. To study ageing and longevity, we obtained RNA-seq data for brain, liver and kidney of newborn, young adult (4-year-old) and old adult (20-year-old) NMRs (Supplementary Table 20). In contrast to other mammals, few genes showed differential expression between 4- and 20-year-old NMRs, especially in the brain (Supplementary Tables 21-23). A recent study identified 33 underexpressed and 21 overexpressed genes in the human brain during ageing16. Of these, 32 genes did not show
consistent expression changes with ageing in NMRs, including 30 genes that had stable expression and two genes that changed in the opposite direction compared to human brain (Supplementary Table 21). For example, CPY66A1 and SMAD3 were downregulated in the human brain, but showed elevated expression in the NMR brain. The product of the CPY66A1 gene is a mediator of cholesterol homoeostasis that influences the tendency of Aβ to aggregate. The product of SMAD3 is a modulator of TGF-β signalling, playing a role in cancer development by slowing down the rate of cell proliferation. Elevated expression of SMAD3 in the NMR during ageing may help optimise the rate of cell death, protecting NMRs from cancer.

A previous meta-analysis of age-related gene expression in mice, rats and humans revealed 56 consistently overexpressed and 17 underexpressed genes. However, many of these genes did not show the same expression changes, suggesting that different regulatory mechanisms may underlie NMR longevity (Supplementary Tables 22 and 23). For example, genes related to degradation of macromolecules, such as GSTA1, DRELI and GNS, were not upregulated with age in NMRs. We also found that genes encoding mitochondrial proteins (NDUFBI1, ATP5G3 and UQCRK2) were not downregulated, consistent with stable maintenance of mitochondrial function during ageing. It is also of interest that TERT (telomerase reverse transcriptase) showed stable expression regardless of age (Supplementary Fig. 12). This finding is consistent with the role of the telomerase complex, highlighted by positive selection on TEP1 and TERR1. Overall, transcriptome and sequence data revealed different (compared to humans, mice and rats) patterns of NMR genes, which may underlie longevity mechanisms in this animal.

Non-shivering thermogenesis is a major heat production process in mammals that mainly depends on the action of UCP1, one of the 39 vertebrate genes that changed uniquely in the NMR (Supplementary Table 19). UCP1 featured changes in amino acids Gin 146, Arg263, Thr264 and Thr303, with the latter two residues being subject to positive selection (P < 0.001, likelihood ratio test for the branch-site model, n = 30) and Arg263 and Thr264 located in the conserved nucleotide binding motifs (Fig. 3a). With Arg→Trp instead of the rigid Gly→Pro in the key regulatory site, UCP1 is expected to lose the tight regulation by purine nucleotides as inhibitors and fatty acids as activators (Fig. 3b and c). The same loop also features two positively charged Lys residues followed by a negatively charged residue (also a unique combination), that should markedly affect the local electrostatic potential of UCP1. In addition, Gin 146 replaced a conserved His involved in proton transport, and the same mutation was shown to decrease proton conductance of UCP1 in vitro40. The 303 is located in the carboxy-terminal motif (RqTxDCxT) required for binding purine nucleotides41. Taken together, these observations indicate a tight association of UCP1 function with the unique thermoregulation of the NMR.

In mammals, switches between light and dark periods affect synthesis of the hormone melatonin, which modulates sleep and circadian rhythms. NMRs live in a naturally dark habitat and their pineal glands, where melatonin is synthesized, are atrophied42, but we found that the genes involved in melatonin synthesis (TPH1, TPH2, DDC, AANAT and ASMT) are intact. Interestingly, the expression of genes involved in the final two steps of melatonin synthesis was very low (AANAT) or undetectable (ASMT) in the NMR brain regardless of age (Supplementary Table 24 and Supplementary Fig. 13). Moreover, two major mammalian melatonin receptors (MTNR1A and MTNR1B, encoding MT1 and MT2, respectively) were inactivated by mutations that introduce premature stop signals (Supplementary Fig. 14). Synteny analyses showed that these pseudogenes corresponded to mouse MTNR1A and MTNR1B. Although melatonin signalling appears to be disrupted in the NMR, its circadian rhythms were maintained in terms of locomotor activity and body temperature when exposed to periodic light/dark changes43. Our finding is consistent with a previous report that MT1/MT2 knockout mice maintained essentially normal circadian rhythms44. These mice also showed decreased insulin secretion45. Likewise, our transcriptome analysis of the NMR revealed decreased expression of genes involved in insulin/GF-1 signalling in the liver compared to mice (Supplementary Fig. 15).

To explain the extraordinary resistance of the NMR to cancer, a two-tier protective mechanism involving contact inhibition mediated by p1616-18 and p2719-21 was proposed. The involvement of p1616-18 is unusual, since humans and mice show only contact inhibition mediated by p2722-24. We analysed the gene locus and the transcript reads corresponding to tumour suppressors p1616-18 and p19-21. As in mice, the p1616-18 transcript consists of three exons (Supplementary Fig. 16). However, sequence similarity in the last exon is low, and two early stop codons in the second exon were predicted to result in a shorter, 14 kDa protein (Supplementary Fig. 17). The four stop codons were, however, intact and Thr69, a residue important for CDK6 binding, was conserved, so the function of the protein may be partially preserved (Supplementary Fig. 18). The p19-21 transcript consists of two exons, but four stop codons in the second exon should lead to a shorter, 10 kDa protein (Supplementary Figs 19–21). The NMR is also unique in that its skin and cutaneous C-fibres lack the neuroepithelial Substance F, making this animal insensitive to certain types of pain11. Our analysis revealed the presence of intact TAC1 encoding Substance P. However, the NMR had a deletion in the core promoter region highly conserved among mammals (Supplementary Fig. 22). Thus, this neurotransmitter appears to be functional but may be under unique regulation.

We further examined the molecular basis for poor visual function and small eyes in the NMR. Of the four vertebrate opsin genes (RHO, OPINL1, OPINL2 and OPINL3), two (OPINL1 and OPINL2) were missing (Table 2); this distinguishes the NMR from other rodents with dichromatic colour vision, such as mice, rats and guinea pigs. However, the NMR has intact RHO (rhodopsin) and OPN4 (melanopsin), supporting the presence of rod-dominated retinae and the capacity to distinguish light/dark cues. Of about 200 genes associated with visual perception (GO:0007061) in humans and mice, almost 10% were inactivated or missing in the NMR (Table 2 and Supplementary Fig. 23). These mammalian genes participate in crystallin formation, photoreduction in the retina, retinal development, dark adaptation, night blindness and colour vision. For at least ten of these
genes, we observed relaxation of the functional constrin on NMR sequences by estimating the ratio of non-synonymous to synonymous substitutions, which corroborated the dysfunction of these genes. Inactivation of CRYBA4, a microphthalmia-related gene, may be associated with the small-sized eyes, whereas inactivation of CRYB4 and CRYB3 and a NMR-specific mutation in CRYGS (Supplementary Table 19) may be associated with abnormal eye morphology. Thus, while some genes responsible for vision are preserved in the NMR, its poor visual function may be explained by deterioration of genes coding for various critical components of the visual system.

Further analysis revealed substantial divergence of the NMR nuclear receptor coactivator Hairless from other mammalian orthologues and the presence of amino acid replacements associated with the hairless phenotype, which is consistent with the lack of fur in NMRs (Supplementary Fig. 24). In addition, we found substantial sequence variation in the sweet taste receptor and lack of many bitter taste receptors common to other mammals (Supplementary Fig. 23 and 26). In particular, the NMR appears to lack the phenylthiocarbamide taste, a dominant genetic trait in humans, as well as several other common bitter tastes.

Air in NMR burrows is low in O₂ (~8%) and high in CO₂ (>10%) overgrowing the large air supply and poor gas exchange through soil. To cope with the low O₂ conditions, the NMR has developed adaptive circulatory (altered haemoglobin oxygen affinity) and metabolic functions, reducing metabolic rate and slowing down development. To obtain insights into this adaptation, we examined gene expression changes in several tissues of NMR subjected to 8% O₂ for one week (Supplementary Tables 25-31 and Supplementary Fig. 27-30). Many changes associated with energy metabolism and redox control were observed. Sequence analysis of NMR hyposxia-induced factor 1x (HIF1α) revealed a T407I exchange unique among mammals and located in the VHL-binding domain. Under normal oxygen conditions, VHL mediates ubiquitin-dependent degradation of HIF1α. In addition, NMR VHIs harbours VHL61I exchange at a functionally important site. These amino acid changes are consistent with relaxation of ubiquitin-dependent degradation of HIF1α, and, thus, with adaptation to low oxygen conditions.

To summarize, sequencing and analysis of the NMR genome revealed numerous insights into the biology of this remarkable animal. In addition, this genome and the associated data sets offer the research communitys working in ageing, cancer, eusociality and many other areas a rich resource that can be mined in numerous ways to uncover the molecular bases for the extraordinary traits of this most unusual mammal. In turn, this information provides unprecedented opportunities for addressing some of the most challenging questions in biology and medicine, such as mechanisms of ageing, the role of genetic makeup in regulating lifespan, adaptations to extreme environments, hyposia tolerance, thermogenesis, resistance to cancer, circadian rhythms, sexual development and hormonal regulation.

**METHODS SUMMARY**

The NMR genome was sequenced on the Illumina HiSeq 2000 platform. The sequenced individual male NMR was from a captive breeding colony located at the University of Illinois, Chicago. The genome was assembled using SOAPdenovo. We obtained 2.5 Gb (gigabase pairs) contig sequences with N50 19.3 kb (kilobase pairs) and N90 744 kb with scaffold sequences with N50 1.6 Mb (megabase pairs) and N90 0.3 Mb. (The N50 (or N90) contig size is the length of the smallest contig in the sorted list of all contigs where the cumulative length from the largest contig to the smallest is at least 50% (or 90%) of the total assembly length.) RNA-seq data (aging and low O₂ experiments) were done for animals from the same colony. See Supplementary Information for data analysis and additional details.

Received 14 April; accepted 5 September 2011.

Published online 12 October 2011.

---


Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. C. Lüttke for help in preparing the manuscript. We acknowledge financial support from the MRC Program (R31-2008-006-10017-0), the NIH (1R01DA012410, 1R01DA15642 and 1R28DA003477A), the National Natural Science Foundation of China (30772008) and the National Science Foundation (0744979).


Author Information The NMR whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AFS00000000. The version described in this paper is the first version, AFS00000000. The mitochondrial sequence has been deposited at GenBank under the accession number JN242981. All short-read data have been deposited into the Short Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRX304668. Raw sequencing data of the transcriptomes have been deposited in the Gene Expression Omnibus with the accession number GSE30337. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike license, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.Z. (dshang@genomics.org.cn) or J.Z. (wangj@genomics.org.cn) or V.N.G. (luuddhuyne@nus.edu.sg).
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,2 Dingding Fan,1 Sun Hee Yim,3 Xiaoming Yao,1 Siming Ma,3 Lan Yang,1 Sang-Goo Lee,3,4 Eun Bae Kim,5 Roderick T. Bronson,5 Radim Sumbera,6 Rochelle Buffenstein,7 Xin Zhou,1 Anders Krogh,2 Thomas J. Park,8 Guojie Zhang,1,2 Jun Wang,1,2,9,* and Vadim N. Gladyshev3,4,*

1 BGI-Shenzhen, Shenzhen 518083, China
2 Department of Biology, University of Copenhagen, Copenhagen, 2200 Copenhagen N, Denmark
3 Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
4 Department of Bioinspired Science, Ewha Womans University, Seoul 120-750, South Korea
5 Rodent Histopathology Laboratory, Harvard Medical School, Boston, MA 02115, USA
6 University of South Bohemia, Faculty of Science, Ceske Budejovice 37005, Czech Republic
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
8 Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA
9 King Abdulaziz University, Jeddah 21441, Saudi Arabia
10 Co-first author

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

Published in Cell Report, 2014, 8(5):1354-64.
Adaptations to a Subterranean Environment and Longevity Revealed by the Analysis of Mole Rat Genomes

Xiaodong Fang,1,2,3 Inge Seim,4,6,8 Zhiyong Huang,1 Maxim V. Geraschenko,2 Zhizhong Xiong,1 Anton A. Taranov,2 Yabing Zhu,9 Alexei V. Lobanov,2 Dongfeng Fan, Sun Hee Yim, Xiaoming Yao,9 Shiming Ma,9 Lan Yang,9 Song-Guo Lai,1,10 Eun Bae Kim,10 Roderick T. Bronson,11 Radim Sumbera,12 Rochelle Bueffenstein,13 Xin Zhou,14 Anders Krogh,14 Thomas J. Park,15 Guojie Zhang,14,16 Jun Wang,14,16 and Vadim N. Gladyshev1,4,16

1Department of Biology, University of Copenhagen, Copenhagen, 2200 Copenhagen N, Denmark
2Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
3Department of Bioinspired Science, Ewha Womans University, Seoul 120-750, South Korea
4Rodent Hepatopathology Laboratory, Harvard Medical School, Boston, MA 02115, USA
5University of South Bohemia, Faculty of Science, Ceske Budejovice 37005, Czech Republic
6Department of Physiology and The Sam and Ari Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, TX 78245, USA
7Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA
8King Abdulaziz University, Jeddah 21441, Saudi Arabia
9Co-first author
10Correspondence: wang@genomics.org.cn (J.W.), vgladyshev@rics.bwh.harvard.edu (V.N.G.)
http://dx.doi.org/10.1016/j.cellrep.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 Demarasland 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 melanin 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 (Segall et al., 2007). African mole rats (hysteroclid 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, 2003), 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 (Kn, 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 Damara land 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 (~78-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 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 (~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% 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 ecophys-iology and exceptional longevity of underground-dwelling African mole rats.

Figure 1. Relationship of the NMR and DMR
(A) The ~35 g NMR and the ~140 g DMR.
(B) Species range map of African mole rats. DMR and NMR occurrence is shown in blue and green, respectively.
(C) Phylogenetic tree constructed using 4-fold degenerate sites from single-copy orthologs, with branch lengths scaled to estimated divergence time (with error range shown in parentheses). Distances are shown in millions of years (Myr). See also Figure S1.

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, 2003). However, their eyes can still serve to alert the colony to invasion by predators by detecting light entering their tunnels (Nemec et al., 2006; Kott et al., 2013). 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 (MDC2) was lost and 13 were pseudogenized in the DMR (Table S2). Three visual genes (CIB1, GRIK7, and GU410) were inactivated or missing in the DMR and the new NMR genome assembly (Table S3). 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 catarracts 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 (Kasikina 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 (Faddy et al., 2001; Wang et al., 2009; Wolf et al., 2005).
Table 1. General Features of the DMR (Fukomys damarensis) Genome and the Improved NMR (Heterocephalus glaber) Genome Assembly

<table>
<thead>
<tr>
<th></th>
<th>Fukomys damarensis</th>
<th>Heterocephalus glaber</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insert Size (bp)</td>
<td>151.5</td>
<td>170-800</td>
</tr>
<tr>
<td>Total Data (Gb)</td>
<td>50.5</td>
<td>128.5</td>
</tr>
<tr>
<td>Seq Coverage (%)</td>
<td>2-20 x 10^3</td>
<td>2-20 x 10^3</td>
</tr>
<tr>
<td>Insert Size (bp)</td>
<td>77.5</td>
<td>120.7</td>
</tr>
<tr>
<td>Total Data (Gb)</td>
<td>25.8</td>
<td>44.7</td>
</tr>
<tr>
<td>Assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A (ko)</td>
<td>259.0</td>
<td>247.2</td>
</tr>
<tr>
<td>Longest (ko)</td>
<td>76.4</td>
<td>91.6</td>
</tr>
<tr>
<td>Size (Gb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>scaffold</td>
<td>29.9</td>
<td>44.7</td>
</tr>
<tr>
<td></td>
<td>22.8</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>3.46</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>18.3</td>
<td>18.9</td>
</tr>
<tr>
<td></td>
<td>178.9</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>2.45</td>
<td>2.45</td>
</tr>
<tr>
<td>Annotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>4,995</td>
<td>21,347</td>
</tr>
<tr>
<td>Total Length (Mb)</td>
<td>22.3</td>
<td>80,826</td>
</tr>
<tr>
<td>Percentage of Genome</td>
<td>2.51</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeats</td>
<td>4,478,125</td>
<td>3,090,116</td>
</tr>
<tr>
<td>Percentage of Genome</td>
<td>28.6</td>
<td>688.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24.2</td>
</tr>
<tr>
<td>Genes</td>
<td>22,179</td>
<td>22,591</td>
</tr>
<tr>
<td>Percentage of Genome</td>
<td>28.7</td>
<td>723.3</td>
</tr>
<tr>
<td>Coding DNA sequence</td>
<td>187,927</td>
<td>181,541</td>
</tr>
<tr>
<td></td>
<td>33.6</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>1.31</td>
<td>1.18</td>
</tr>
</tbody>
</table>

See also Figures S1A and S1B.

Adaptations to Hypoxia and a High Carbon Dioxide and Ammonia Environment

The DMR, NMR, and other subterranean rodents nest 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, 2003). Ammonia is a potent irritant that arises from nitrogren and methane accumulation in lattines and nests (Burda et al., 2007; Laflinika et al., 2008). 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 Leu214, which is present in 38 other vertebrate species (Figure 2A). This amino acid change was also detected in the distinctly related subterranean conuro (Spelaeomys cyamus) and the semi-subterranean degu (Octodon degus) of South America. The common ancestor of Octodontidae (conuro and degu) and Caviomorphs (guinea pig) diverged ~35 Mya, when African and South American rodents diverged (~41 Mya (Antonie et al., 2012; Moretti et al., 2011; Figure 2B). His254 is located immediately downstream of a conserved motif required for binding manganese and Arg1 function (Dovling et al., 2008; Figure 2C). Moreover, ARG1 is a homotrimmer, with the salt bridges formed by Arg255 and Glu256 being critical for its assembly (Laxolo et al., 2001; Sabol 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 (SLC9A15), which is essential for the urea cycle (Piermonte 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 nociceptor 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 Maashona mole rat [Fukomys darlingi, FD]), the South American subterranean conuro 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 galil also harbors the same amino acid changes (Fang et al., 2014). These animals are exposed to chronic hypoxia and hypercapnia in burrows or caves (Figure 2E), suggesting that convergent evolution resulted in similar 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 conuro 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 (HBA1 and HBA2,
Figure 2. Subterranean Adaptations in Hypoxia-tolerant Rodents
(A) Subterranean hypoxia-tolerant rodents share a charged residue at position 254 of arginase 1 (ARG1). The manganese-binding site, residues critical for enzyme trimer assembly (Arg255 and Gln256), and unique His254 changes are highlighted in purple, orange, blue, and red, respectively. Identical residues in vertebrates are shaded in black.
(B) Phylogenetic relationship of hypoxia-tolerant rodent lineages examined in this study. Approximate divergence times (Myr) are indicated.
(C) Structural model of human ARG1 monomer. Residues are highlighted as in (B).
(D) Schematic representation of the roles of components of the urea cycle with altered sequence (purple box) or expression in NMR and DMR (green boxes). CPS1, carbamyl phosphate synthetase 1; OTC, ornithine transcarbamylase; OAT, ornithine transporter 1; ASL, argininosuccinate lyase; ARS1, arginase 1; ARS2, arginase 2.
(E) Species of hypochoeran habitats share a negatively charged residue at position 254 of the beta 1-7 sodium channel protein. Acidic amino acid residues in the motif, corresponding to amino acids 1716 and 1720 of the human sequence, are shown in red. Identical residues in vertebrates are shaded in black.
(F) Heatmap of globin expression in normoxic rodent brains. Scaled log2 transformed normalized read counts (centered around the row mean) are plotted in beige-blue color, with blue indicating high expression and beige indicating low expression. B. guinea pig, Brazilian guinea pig; MBH1/2, hemoglobin x; NGB, neuroglobin; CYGB, cytoglobin. Red stars indicate differentially expressed genes in subterranean rodents.
(G) Western blot of hemoglobin x in normoxic rodent brains with antibodies against the mouse protein.
(H) Comparison of globin gene expression under normoxia (21% O2) and hypoxia (8% O2, over 8 h). Annotated as in (F).

See also Figure S2.
associated with improved oxygen delivery in hypoxia-tolerant seals (Schnee et al., 2012) and shellfish (Kruus and Colozzo, 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 (Edery et al., 2011).

We compared the transcriptomes of the liver (a relatively homogenous organ) of the NMR and DMR (Hystricognathi) with those of the short-lived rat and mouse (Mundidae) (Table S4). Compared with the mouse and rat, which spend considerable time underground, 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 (Krutiski et al., 2009), 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., 2014) and reduced oncogenic potential of cultured human cancer cells (Zhu 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., 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 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; Kramar and Sutterstein, 2004). We found that the NMR, DMR, and other hystricognathous rodents harbor a divergent insulin β-chain sequence (Figure 3A). This finding is consistent with the observation that insulin in the South American hystricognaths is rapidly evolving (Ozur et al., 2006), in the guinea pig and other South American hystricognaths, the regions encoding the α-chain and, in particular, the β-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., 1993; Ozur et al., 2006). Mutations in the human β-chain result in reduced insulin processing, mistitling, and less effective insulin (based on receptor binding) (Lu et al., 2010). Interestingly, residue 22 of the β-chain, whose mutation (Arg22Gln) is associated with mistitling of insulin and diabetes (Lu et al., 2010), is uniquely changed in both African and South American hystricognaths (Figure 3A). In the African created porcupine, this residue was previously linked to an altered insulin structure with reduced affinity for insulin receptors (Ronik et al., 1985). 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 (Edery et al., 2011) and South American hystricognaths (Ozur 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 (Quayle et al., 1987). 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 (Lang 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, IGFBP3, in the liver (Figure 3B; Table S4). We hypothesized that autocrine/paracrine production of IGF2 in the liver substitutes for insulin and may partly mediate a fetal-like mode of glucose handling in hystiocogyn rodent (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 (Boggs et al., 2012). Interestingly, molecular innovations of this axis may contribute to the lifespan of the long-lived Brand's bat (Selim 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 IGFR1 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 hystiocogyn (e.g., porcupine and guinea pig). Our findings support the hypothesis that hystiocogyn rodents have evolved a distinct insulin peptide.

Cancer Resistance

Studies of the NMR (Lang 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; Nased et al., 2000; Manov et al., 2013) suggest that many species of long-lived mole rats escape cancer, and even if they do develop pathology, will present a milder phenotype in comparison with short-lived rodents (e.g., mouse) (Aznau and Selu, 2012; de Magalhães, 2013).

A recent study suggested that a 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 are 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 A476Ser is unique to the NMR (Figure S2B). 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 pathways to cancer resistance in the blind mole rat (Aznau and Selu, 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 (Giu et al., 2013). Analysis of nonsynonymous-to-synonymous substitution (Ka/Ks) ratios of 9,867 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 S9). 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 36°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 (Buffy and Yahav, 1981). 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 (Hulst and Buffonstein, 1946), 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., 1982). 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 knock out animal (Figure S3B and S3C). Both the DMR and NMR lost MTNR1b, although the inactivating mutations are located in different positions (Figure S3B). MTNR1a is also a pseudogene in the distantly related Siberian hamster (Phodopus sungorus) (Plendergast, 2013) 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 (Plendergast, 2013). 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 preprocalcitonin 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 calciactorin gene CALCA is responsible for the synthesis of two distinct preprohormones by means of alternative splicing (Flandel et al., 1998). Splicing into exon 4 results in calciactorin (CT), while splicing into exon 5 and the 3’ untranslated exon 6 encodes the sensory neuropeptide CGRP (Figure S3B). 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 (Juin Cers 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

Figure 4. Adaptive Evolution in the NMR and DMR Genomes
(A) Accelerated evolution of the NMR and DMR genomes. GO categories with putatively accelerated (p ≤ 0.05, t-test): nonhomonymous divergence in the NMR lineage (turquoise) and the DMR lineage (orange) are highlighted.

(b) Conserved gene syntax of the β-actin gene (ACTB) region among human, DMR, and NMR. Boxes represent genes.

(c) Schematic of the intron and exon structure of ACTB along with the location of amino acid changes found in the NMR ACTB protein. Turquoise boxes represent exons.

(d) NMR has unique amino acid changes highlighted in red in the highly conserved (turquoise) ACTB, including the redox-sensitive Cys27/27 residue. See also Figure S3.

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 (GRN/2B), the TRP channel TRPM3, and proenkephalin (PENK) (Table S2). 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, 1976). During oxidative stress, cysteine residues of actin can be oxidized, which is associated with depolymerization and altered regulatory protein interactions (Temman and Kasahara, 2013). Increased ROS levels and actin overoxidation are symptomatic of senescence and diseases such as Alzheimer’s (Alexeenko 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 Ala250 of β-actin are converted to serine in the NMR (Figures 4C and 4D). Cys272 is highly redox sensitive and may serve as a “redox sensor” (Lassig 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 Hysterognath 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-tuco (Chironomus) and the degu (Octodontomys pirroles) (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 Talaus tuco-tuco (Chironomus taurus) (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 hystriognath, the South American guinea pig and the African DMR, do not harbor the cryptic intron (Figure 5B). The data suggest that the cryptic intron and the resulting “broken” 28S rRNA were present in a common ancestor prior to the trans-Atlantic migration of small African hystriognath rodents to South America ~41

Figure 5. 28S rRNA in the NMR
(A and B) Denaturing agarose gel electrophoresis of total RNA from (A) mouse liver and NMR liver, and (B) NMR brain and DMR brain, liver, kidney, and testes.

(C) A cryptic intron in the divergent region D6 of 28S rRNA is conserved between the NMR and tuco-tuco. Tuco-tuco splice donor and acceptor sites are indicated in bold. Dashes (-) indicate gaps and asterisks (*) indicate identical nucleotides.

(D) Proposed biogenesis of 28S-rRNA subunits by “splicing” of the cryptic intron in the D6 domain. Two 28S subunits are produced in NMR organs. No conventional 28S-rRNA exists in the DMR.

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, Azurua et al. (2019) 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 transcriptsomes of these subterranean rodents were also sequenced. African mole rats have 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 melanin 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., 2009). The animals were sacrificed and DNA and RNA were isolated for subsequent sequencing. The genome sequenced was that of a 9-year-old male DMR liver and brain transcripts 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 DNA. In brief, 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. In brief, 250 bp (76X) high-quality data, including 151 Gbp (259G) short-insert size reads, were generated [Table 1]. The genome was de novo assembled by SOAPdenovo [Li et al., 2010a]. Then, 121 Gbp (35X) data from short-insert-size libraries (300-800 bp) were split into 60-mers and contigs, with unambiguous connections in the de Brujin graphs retained. All reads were aligned onto contigs for scaffold building using PS information. We used k-mer analysis [Li, et al., 2010b] to estimate the genome size of the DMR. In this study, K = 17, K/n = 61,035,145,821, and K = depth = 22.5. Therefore, the DMR genome size was estimated to be 2.73 Gb (Figure S1A).

**Assembly of the NMR Genome**

To develop an improved assembly of the NMR genome, we used LASTZ [Zhang, 2009], with the parameter “-k 4” [Kumar et al., 2003] to search for alignments among NMR reads. The primary assembly 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 SAM format, we selected BAMtools, which is based on the Bayesian model for variant analysis [Li et al., 2008]. After selecting alignments by theeth distriburtion of read counts and removing potential PCR duplicates, we used BAMtools to call SNPs and short InDels. We rejected SNPs and InDels within the first 100 bp of the read and used the remaining read for alignment, since the depth of coverage varies greatly between SNPs and InDels. We used the depth from the original read for alignment when the depth was low. We then used the software SAMtools to generate the SAM format, which was used for the pipeline with bp tool with pipe tool. The sequence reads Q 20 = 20-4 C 150-8 25-120 M 5-18 N 5-1.

**Repeat Annotation**

RepeatMasker and RepeatProteinMask (Tesar and Chen, 2002) were used to identify and classify transposable elements by aligning the DMR genome sequences against a library of known repeats. RepeatMasker, 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 mammalian genomes, including the NMR.

**ACCESSION NUMBERS**

The DMR genome whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under accession code AY220200000. The version described in this paper is the first version, AYU22011000.0. All short-read data have been deposited in the Short Read Archive under accession code SRX928405. Fast sequencing data of the transcriptome have been deposited in the DDBJ/EMBL/GenBank under accession code GSE65726.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, four figures, and six tables and can be found with this article online at http://doi.org/10.1016/s0096-8964(14)00759-0.

**AUTHOR CONTRIBUTIONS**


**ACKNOWLEDGMENTS**

This work was supported by the NHM (N0471740, N0472200, A020118, and A018603), the WCU Program (R31-2008-000-10030-5), and the National Nature Science Foundation of China (31171198), NMR and DMR photos were provided by Joel Sartore (http://www.joelsartore.com).

Received: October 3, 2013

Revised: May 11, 2014

Accepted: July 11, 2014

Published online August 29, 2014

**REFERENCES**


Cell Reports 8, 1–11, September 11, 2014 ©2014 The Authors
Melan, G.J., Pesos, C.G., Rossi, M.T., and Kormbllitl, A.R. (1999). Novel pro-
cessing in a mammalian nucleus 28R pre-mRNA: tissue-specific elimination of an intron 'hiding' a broken bad site. EMBO J. 18, 3107–3116.


Park, T.J., Lu, Y., Jüttner, S., Smith, E.S., Hu, J., Brand, A., Watzel, C., Mile-

Prendergast, B.J. (2010). MT1-Metastasis receptors mediate somatic behav-
ioral, and reproductive neuroendocrine responses to photoperiod and mela-
nin in hamster females (Phodopus sungorus). Endocrinology 151, 714–721.

Qiu, G., Zhang, G., Ma, T., Glen, W., Wang, J., Ye, Z., Cao, G., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yam genome and adaptation to life at a high alti-


Genome-wide adaptive complexes to underground stresses in blind mole rats *Spalax*

Xiaodong Fang¹,², Eviatar Nevo³, Lijuan Han¹, Erez Y. Levanon⁴, Jing Zhao²,⁵, Aaron Avivi¹, Denis Larkin⁶, Xuanting Jiang¹, Sergey Feranchuk⁷, Yabing Zhu¹, Alla Fishman⁷, Yue Feng⁴, Noa Sher⁷, Zhiqiang Xiong¹, Thomas Hankeln⁸, Zhiyong Huang¹, Vera Gorbunova⁹, Lu Zhang¹,¹⁰, Wei Zhao¹, Derek E. Wildman¹¹, Yingqi Xiong⁵, Andrei Gudkov¹², Qiumei Zheng⁵, Gideon Rechavi¹³,¹⁴, Sanyang Liu¹, Lily Bazak³, Jie Chen¹, Binyamin A. Knisbacher⁴, Yao Lu¹, Imad Shams³, Krzysztof Gajda⁶, Marta Farre¹⁶, Jaebum Kim¹⁵, Harris A. Lewin¹⁶, Jian Ma¹⁷,¹⁸, Mark Band¹⁹, Anne Bicker⁸, Angela Kranz⁸, Tobias Matthes⁸, Hanno Schmidt⁸, Andrei Seluanov⁹, Jorge Azpuruá⁹, Michael R. McGowen¹¹, Eshel Ben Jacob²⁰, Kexin Li³, Shaoliang Peng²¹, Xiaoqian Zhu²¹, Xiangke Liao²¹, Shuaicheng Li¹⁰, Anders Krogh², Xin Zhou⁵, Leonid Brodsky⁷,²² & Jun Wang¹,²,²³

1 BGI-Tech, BGI-Shenzhen, Shenzhen 518083, China.
2 Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark.
3 Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel.
4 The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel.
5 China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China.
6 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DA, UK.
7 Tauber Bioinformatics Research Center, University of Haifa, Mount Carmel, Haifa 31905, Israel.
8 Institute of Molecular Genetics, Johannes Gutenberg University, 55099 Mainz, Germany.
9 Department of Biology, University of Rochester, Rochester, New York 14627, USA.
10 Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong.
11 Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201, USA.
12 Department of Cell Stress Biology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA.
13 Division of Pediatric Hematology and Oncology, Edmond and Lily Safra Children’s Hospital and Sheba Cancer Research Center, The Chaim Sheba Medical Center, Tel Hashomer 52662, Israel.
14 Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.
15 Department of Animal Biotechnology, Konkuk University, Seoul 143-701, Korea.
16 Department of Evolution and Ecology, University of California, Davis, California 95616, USA.
17 Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.
18 Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.
19 Roy J. Carver Biotechnology Center, University of Illinois, Urbana, Illinois 61801, USA.
20 School of Physics and Astronomy, Maguy-Glass Chair in Physics of Complex Systems, Tel Aviv University, Tel Aviv 69978, Israel.
21 School of Computer Science & State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, China.
22 Department of Evolutionary and Environmental Biology, University of Haifa, Mount Carmel, Haifa 31905, Israel.
23 Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21441, Saudi Arabia.

Correspondence and requests for materials should be addressed to E.N. (email: nevo@research.haifa.ac.il) or to X. Zhou (email: xinzhou@genomics.cn) or to L. Brodsky (email: lbrodsky@research.haifa.ac.il) or to J.W. (email: wangj@genomics.cn).

Published in Nature Communications, 2014, 5:3966.
ARTICLE
Received 24 Oct 2013 | Accepted 28 Apr 2014 | Published 3 Jun 2014 DOI: 10.1038/ncomms4966

Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax

Xiaodong Fang, Eviatar Nevo, Lijuan Han, Erez Y. Levanon, Jing Zhao, Aaron Avivi, Denis Larkin, Xuanting Jiang, Sergey Feranchuk, Yabing Zhu, Alla Fishman, Yue Feng, Noa Sher, Zhiquang Xiong, Thomas Hanksen, Zhiyong Huang, Vera Gorbunova, Lu Zhang, Wei Zhao, Derek E. Wildman, Yingqi Xiong, Andrei Gudkov, Qiumei Zheng, Gideon Rechav, Sanyang Liu, Lily Bazak, Jie Chen, Binyamin A. Krisbacher, Yao Lu, Irrad Shams, Krzysztof Gajda, Marta Farré, Jaeburn Kim, Harris A. Lewin, Jian Ma, Mark Band, Anne Bicker, Angela Kranz, Tobias Matteus, Hanno Schmidt, Andrei Seluanov, Jorge Azpuru, Michael R. McGowen, Eshel Ben Jacob, Kexin Li, Shaoliang Peng, Xiaqian Zhu, Xiange Liao, Shaicheng Li, Anders Krogh, Xin Zhou, Leonid Brodsky & Jun Wang

The blind mole rat (BMR), Spalax golliei, is an excellent model for studying mammalian adaptation to life underground and medical applications. The BMR spends its entire life underground, protecting itself from predators and climatic fluctuations while challenging it with multiple stressors such as darkness, hypoxia, hypocapnia, energetics and high pathogenicity. Here we sequence and analyse the BMR genome and transcriptome, highlighting the possible genomic adaptive responses to the underground stressors. Our results show high rates of RNA/DNA editing, reduced chromosome rearrangements, an over-representation of short interspersed elements (SINEs) probably linked to hypoxia tolerance, degeneration of vision and progression of photoperoxidation perception, tolerance to hypocapnia and hypoxia and resistance to cancer. The remarkable traits of the BMR, together with its genomic and transcriptomic information, enhance our understanding of adaptation to extreme environments and will enable the utilization of BMR models for biomedical research in the fight against cancer, stroke and cardiovascular diseases.

1 BGI-Tech, BGI-Shenzhen, Shenzhen 518083, China. 2 Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark. 3 Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel. 4 The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel. 5 China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China. 6 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DA, UK. 7 Tauter Bioinformatics Research Center, University of Haifa, Mount Carmel, Haifa 31905, Israel. 8 Institute of Molecular Genetics, Johannes Gutenberg University, 55099 Mainz, Germany. 9 Department of Biology, University of Rochester, Rochester, New York 14627, USA. 10 Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong. 11 Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201, USA. 12 Department of Cell Stress Biology, Roswell Park Cancer Institute, Buffalo, New York 14203, USA. 13 Division of Pediatric Hematology and Oncology, Edmond and Lily Safra Children’s Hospital and Sheba Cancer Research Center, The Chaim Sheba Medical Center, Tel-Hashomer 52662, Israel. 14 Sackler School of Medicine, Tel Aviv University, Tel-Aviv 69978, Israel. 15 Department of Animal Biotechnology, Konkuk University, Seoul 143-701, Korea. 16 Department of Evolution and Ecology, University of California, Davis, California 95616, USA. 17 Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. 18 Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. 19 Roy J. Carver Biotechnology Center, University of Illinois, Urbana, Illinois 61801, USA. 20 School of Physics and Astronomy, Megapix-Glass Chair in Physics of Complex Systems, Tel Aviv University, Tel Aviv 69978, Israel. 21 School of Computer Science & State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, China. 22 Department of Evolutionary and Environmental Biology, University of Haifa, Mount Carmel, Haifa 31905, Israel. 23 Princess Al Jawhara Center for Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21441, Saudi Arabia. Correspondence and requests for materials should be addressed to E.N. (email: nevo@research.haifa.ac.il) or to X. Zhou (email: xinzhuo@genomics.cn) or to L. Brodsky (email: lbrodsky@research.haifa.ac.il) or to J.W. (email: wang@genomics.cn).
The subterranean blind mole rat (BMR, Spalax gubeli), belongs to the Spalax ehrenbergi superspecies in the family Spalacidae, a clade of subterranean murid rodents distributed in the Eastern Mediterranean and North Africa. BMRs are solitary mammals that spend their lives in underground burrows, which shelter them from most predators and climatic fluctuations. However, in this subterranean environment they are challenged by multiple stressors such as darkness, hypoxia, hypercapnia, energetic challenges during digging and increased exposure to pathogens. Consequently, BMRs evolved multiple genetic adaptations to cope with these stresses. This has provided researchers with a unique evolutionary model for investigating stressful life underground. The BMR has been extensively studied genomically, proteomically and phenomically (see for full list of publications please refer to the online Spalax publication list: http://evolution.haifa.ac.il/index.php?8-people/publications/152-publication-nevo-spalax).

The lack of light in subterranean burrows presumably has triggered a complex mosaic of traits involving light perception, including severe ocular regression with a subcutaneous minute degenerated eye, coupled with the elaboration of phototrophic perception. The BMR has evolved a larger brain volume compared with that of rodents of similar body size because of the development of an expanded neocortex with developed vibrational, tactile, vocal, olfactory and magnetic (spatial) orientation systems replacing sight as well as the enlargement of the motor structures related to digging activity.

Most remarkably, the BMR has evolved physiological strategies to survive and carry out intense activities in a highly hypoxic and hypocapnic environment with wide fluctuations in oxygen concentration (O2 as low as 2.2% and CO2 as high as 6.1% were recorded in underground burrows). These adaptations include improved myocardial oxygen delivery and function, adaptive heart and breathing frequencies, high blood haemoglobin and haematocrit concentrations, respiratory adaptations, enlarged alveolar surface area and alveolar capillary volume, and increased tissue mitochondrial and capillary densities. At the molecular level, a growing list of genes has been reported to underlie the BMR’s hypoxic tolerance. For example, constitutively increased mRNA and protein expression levels of hypoxia-regulated genes were observed. The p53 tumor suppressor gene is a critical tumor suppressor in mammals and is known to mediate tumor suppression in response to hypoxia. Moreover, in an in vitro system BMR fibroblasts were shown to inhibit growth and even kill human cancer cells. Recent findings suggest that the unique BMR anti-cancer mechanism is mediated by an induction of the necrotic cell-death mechanism in response to hypoxia.

In addition, the BMR shows a striking resistance to cancer: not a single case of spontaneous tumor development was recorded among thousands of captive animals over a 40-year period, indicating animals over 20 years old that may be more vulnerable to cancer. Experiments using chemical carcinogens to induce tumor growth had negative results in BMR, while 100% of mice, rats and hamsters (33 of 33 wild rodent species) developed tumors. Moreover, in an in vitro system BMR fibroblasts were shown to inhibit growth and even kill human cancer cells. Recent findings suggest that the unique BMR anti-cancer mechanism is mediated by an induction of the necrotic cell-death mechanism in response to hypoxia.

Here we report the sequencing and analysis of the BMR genome and transcriptome. Our results reveal the unique adaptive genomic features of the BMR, including high rates of DNA and higher RNA editing compared with that of the mouse and rat, lower rates of chromosome rearrangements, as well as an over-representation of SINE (short interspersed element) transposable elements, which is likely because of a novel response to hypoxia tolerance. In addition, our analysis of the BMR genome and transcriptome identifies the evolution of placenta-specific genes and revealed molecular adaptations involved in response to darkness, tolerance of hypercapnia and hypoxia, for example, by modifications of respiratory proteins, and resistance to cancer. The extreme adaptations and characteristics of the BMR, together with the reported genome and transcriptome, will empower future use of the BMR model for biomedical research in the fight against cancer, stroke and cardiovascular diseases.

Results

Genome assembly and annotation. The DNA obtained from the brain of a female Spalax gubeli (a diploid with the chromosome number of 2n = 52) was sequenced using a whole-genome shotgun strategy utilizing Illumina sequencing technology. Various insert size libraries were used to generate 392 Gb of raw data, of which 259 Gb (66% coverage) of high-quality data were retained for assembly (Supplementary Table 1). Genome assembly using SOAPdenovo36 as described produced a final assembly of 3.06 Gb, consistent with the kmer-based genome size estimation (~3.04 Gb, Supplementary Table 1, Supplementary Figs 1 and 2); the contig and scaffold N50s were 27.5 kb and 3.6 Mbp, respectively (Supplementary Table 3). The average GC content is 41.23%, comparable to the human genome. To assess the quality of the assembly, BMR and rat transcriptome data generated for this study were mapped on our BMR assembly and the rat genome, respectively; on average, 81.5% of reads mapped to the BMR assembly, comparable to 82.1% that mapped to the rat. Transcriptome assembled contigs using Trinity were mapped to the BMR genome. Of the 63.38 Mb contigs, 98.9% were covered by more than 90% coverage, suggesting the completeness of transcript representation in our assembly. Moreover, randomly selected 20-fold of paired-end reads from short insert size libraries were aligned to the assembly, and 92.75% were successfully mapped using BWA (Burrows-Wheeler Aligner, version 0.7.12-r19) giving a genome coverage of ~99.57%, suggesting our assembly covered most of the genome (Supplementary Tables 4 and 5). We identified 3.26 M single-nucleotide polymorphisms (SNPs) and 627 K short InDels, the heterozygosity of SNP and InDel were estimated to be 0.11% and 0.02%, respectively, or 0.13% combined together. This is comparable to the human, but is higher than its relative, another subterranean rodent called the naked mole rat (NMR, Heterocephalus glaber), which is proposed as naturally inbred because of its unusual behavior. By analyzing the context dependency of BMR SNPs, we found heavy reduction in SNP because of CpG mutations compared with other mammals (Supplementary Fig. 3). BMR also has a lower CpG observed/expected value compared with other mammals (Supplementary Fig. 4) and a higher fraction of CpG dinucleotides concentrated in CpG islands as compared with the mouse or rat (Supplementary Table 6). More details shown in Supplementary Note 1.

Reference-assisted chromosome assembly reconstruction of predicted chromosome fragments (PCFs) of the BMR genome was performed using a threshold of 50 and 80 kb to include syntetic fragments (STs) in two independent reconstruction experiments, respectively (Supplementary Note 2 and Supplementary Table 7). At 50 kb resolution, 41 PCFs were recovered, covering 82.1% of the BMR genome. At 80 kb resolution, 36 PCFs were reconstructed, giving an overall genome coverage of 76.8%. These 36 PCFs contain 18 interchromosomal rearrangements of which 17 occurred in the Muridae lineage (mouse and rat) and one in the BMR lineage. The number of BMR scaffolds containing >1 SF was 25 (3.6%), 80 kb-
resolution). These mainly represent BMR-specific chromosomal rearrangements, several chimeric scaffolds or, in some rare instances, misalignments between the BMR, mouse and human sequences. These scaffolds contain 27 potential evolutionary breakpoint regions. At 80 kbp resolution, our data indicate that the BMR genome evolved with the rate of at least 0.56 rearrangements after the split from a common ancestor with the mouse ~47.6 million years ago (MYA; www.timetree.org; Supplementary Fig. 5). This rate is lower than the ~2.1 and ~1.9 rearrangements per million years (300 kbp resolution) in artiodactyl and primate genomes, suggesting that a striking stability of the BMR chromosomal arms in evolution could compensate for a large number of chromosomal fusions and fissions observed in BMR chromosomal species and speciation.

Repeats constitute 43.9% of the BMR genome, with retrotransposons being the most prevalent transposable elements (32.5% of the genome). Interestingly, the proportion of SINEs (11.8%) in BMR is more similar to that in the human genome (13.7%), and much higher than in other sequenced rodents (6.5%, 5.6% and 5.6% in the mouse, rat and NMR, respectively), B1, B2 and B4 are the three most abundant SINEs (10% of the genome) in the BMR assembly, indicating a BMR-specific expansion of these transposable elements compared with mouse, rat, NMR and human (Supplementary Figs. 6–8, Supplementary Table 8, Supplementary Data 1–3). The functional role of B1, B2 and B4 SINEs in mouse has been intensively investigated. B1 and B2 were shown to be stress-inducible factors in the ischaemic brain44. In addition, upregulation of B2 SINE RNA downregulates transcription in response to heat shock in mouse45,46. B2 can induce polymorphic expression of the 5-aminolevulinic acid synthase 1 (ALAS1) gene, the key to nonerythroid haem biosynthesis, and is involved in respiration, drug metabolism and cell signalling47. Therefore, the over-representation of B1, B2 and B4 SINEs in the BMR genome might reflect exposure to hypoxic stresses in the underground environment. In particular, we found that the B1 SINE repeat showed high upregulation (see below) under severe hypoxia in the BMR brain. A similar mode of regulation was reported for human Alu SINEs that were shown to be transcriptionally upregulated under hypoxic conditions, presumably contributing to genomic instability in tumours48.

Gene annotation and molecular phylogeny. We obtained a reference gene set that contained 22,168 coding genes in the BMR genome by combining homologous searching, transcriptome evidence and de novo prediction (Supplementary Note 3; Supplementary Tables 9 and 10). Of the predicted genes, 19,730 (89%) were recovered by the RNA-seq data. Among the reference genes of BMR, orthologues with other animals showed highest similarity between the BMR and Chinese hamster (Cricetulus griseus; Supplementary Table 11, Supplementary Fig. 9). We inferred 12,767 gene families (Supplementary Fig. 10) and constructed a phylogenetic tree from 1,583 single-copy orthologues. As shown in Fig. 1, the BMR was placed within Rodentia and diverged from the ancestor of rats, mice and Chinese hamster (C. griseus) ~47 MYA, consistent with a previous study37. The common ancestor of the BMR, mouse, rat and Chinese hamster separated from a lineage leading to NMR ~71 MYA. In comparison with several other sequenced mammalian genomes, we estimated that 139 gene families showed expansion and 50 gene families showed contractions in the BMR (Supplementary Data 4 and 5; Supplementary Fig. 11). BMR- and NMR-specific genes are shown in Supplementary Data 6 and 7 and Supplementary Table 12. We also detected 35 lost genes (Supplementary Table 13) and 259 pseudogenes (Supplementary Data 8, Supplementary Tables 14 and 15), as well as 48 positively selected genes (Supplementary Data 9, Supplementary Tables 16 and 17) during BMR evolution, which altogether may underlie the physiological adaptations of the BMR to its stressful underground niche. Copy numbers of certain cancer-related genes in BMR, NMR, mouse and rat are shown in Supplementary Data 10.

One outstanding issue in the evolution of murid reproduction is the timing of diversification of murid-specific genes expressed solely in the placenta (cathepsins, prolactins, pregnancy-specific glycoproteins and syncytiotrophoblast proteins)35. We sequenced the transcriptome of the placenta in BMR and queried the BMR genome using rodent transcripts (Supplementary Note 4; Supplementary Table 18). We identified 13 prolactins, 6 pregnancy-specific glycoproteins and 4 placental cathepsins. These numbers are intermediate between murids and NMR (Supplementary Table 19 and Supplementary Figs 12 and 13), implying that these families had started to diversify before the origin of Muroidea ~47
MYA5. In addition, we identified two murid-specific synctins (Syna and Synb), both of which are not present in the NMR. This is potentially because of differences in placental morphology.

Extensive DNA and RNA editing increases adaptive potentials. Increase in genomic diversity may reinforce the adaptation to subterranean lifestyle. DNA and RNA editing of retroelements enhances intra- and interspecies diversity. We screened the BMR genome and transcriptome for these, making BMR one of the first organisms to be comprehensively analysed for both types of editing, preceded only by human and mouse.

First, we detected DNA editing of retroelements across the assembly. The BMR genome contains one APOBEC3 (AD) gene, whose product can introduce a series of C to U mutations into the negative strand of nascent retroelement DNA. Generation and analysis of paired alignments within LTR retrotransposon families revealed numerous retroelements containing clusters of G to A mutations, signs of DNA editing by A3 (2.459 elements, 23,853 edited nucleotides). BMR AD, identical in protein sequence to its murine counterpart, is also identical in its preference for the GxA motif in LTR retroelement-editing sites. Another trait shared by the BMR and murine genomes is a strong signal of editing in the active Intracisternal A-type particle retroelements. For detailed information please refer to Supplementary Note 5, Supplementary Tables 20–23, Supplementary Data 11–15 and Supplementary Figs 14–20.

Next, we detected RNA editing using a combination of transcriptome reads obtained from BMR and rat hypoxia samples (Supplementary Figs 21–33). A-to-I RNA editing is more prevalent in primates than in rodents and catalysed by adenosine deaminases acting on RNA (ADAR) enzymes. Abundance of similar SINEs in a genome increases the chance of two similar and reversely oriented elements to reside next to each other. When transcribed, they are likely to form a double-stranded RNA structure, the preferred substrate for ADAR.

We mapped RNA reads to the assembled genome and found mismatches in 237,094 SINE elements. As RNA editing tends to come in clusters, we searched for these and detected 19,714 SINEs with at least four mismatches of the same type, with a total of 113,338 AG/TC-editing sites, containing the ADAR sequence motif, supporting their authenticity. Strikingly, the number of editing sites was >107 times greater than the number of the control mismatch clusters (G to A/C to T). A similar comparison to rat yielded only a 1.41 ratio between editing and control sites. Together with previous analyses of mouse RNA editing, we conclude that the BMR's RNA-editing occurrence is higher than those of both the rat and mouse. We predict that a relatively large number of lineage-specific RNA-editing events take place in many BMR genes, contributing to its unique adaptation, without necessitating mutations in the genome.

Adaptation to darkness. The lack of visible light underground triggered a complex degradation of the eye in the BMR involving drastic ocular regression with a small degenerate subcutaneous eye and a simultaneous elaboration of circadian rhythm perception. Of the 259 pseudogenes found in the BMR genome, there are 22 genes involved in the visual system (Supplementary Table 15). Fifteen of these visual system pseudogenes contain no alternative splicing forms to avoid the mutation sites and are likely complete pseudogenes. A comparison of the BMR gene families with those of other species (human, mouse, rabbit, rat, mouse, NMR and dog) indicated a contraction in the beta/gamma crystalline gene family (P value = 0.047), potentially linked with the BMR's degliation of vision. This confirms our previous results that transgenic mice carrying the BMR cristallin gene selectively lose lens activity after 13 days of embryogenesis.

The circadian rhythm of the BMR has been previously studied and genes involved were identified. Multi-alignments of the protein sequences of circadian genes revealed that both BMR and NMR CLOCK proteins have an expanded Q-rich region compared with that of the human and mouse, and are different in amino-acid composition from that of the rat (Fig. 2a). The phylogenetic tree indicates that the BMR and the NMR CLOCK proteins display a higher similarity in amino-acid composition, despite being phylogenetically distant (Fig. 2b). Since the glutamine-rich area is assumed to be involved in circadian rhythmicity, the similarity in amino-acid composition of the BMR and NMR Clock genes may indicate convergent evolution by these subterranean animals.

In addition, the glial cell line-derived neurotrophic factor family has expanded in BMR because of multiple duplications of the gene Gfra1. As the BMR brain is twice as large as that of a rat with similar body size, and since Gfra1 can serve as a potent neuronal survival factor, its over-representation may have contributed to the enlargement of the motor structures and somatosensory system in the BMR brain, enabling highly developed digging activity and sense physiology to replace sight.

Adaptation to hypoxia and hypoxia. Transcriptome analysis (Supplementary Note 6, Supplementary Data 16 and 17, Supplementary Figs 34–42 and Supplementary Tables 24 and 25) revealed an upregulation of BMR genes after experimental hypoxia, including known HIF-1 targets, such as hexokinase (Hki), adenominulin (Admn), macrophage migration inhibitory factor (Mif), R étape repeat domain-containing protein 57 (Akrnl57), secretogogin (SCGN), neuromedin-B (NMB) and cocaine-and-amphetamine-regulated transcript (CART). GO analysis (Supplementary Data 18) indicated that hypoxia-upregulated genes were enriched in terms related to cellular defense, response to steroid hormone stimulus, ribosome biogenesis, mitochondrial ribosome, RNA splicing, cytoskeleton, regulation of circulation and blood pressure, regulation of appetite, and regulation of circadian rhythms. At the same time, hypoxia-downregulated genes were enriched in terms related to neuron morphogenesis, intracellular protein transport, transmission of nerve impulses, phosphate metabolic processes, regulation of cell motion, regulation of protein kinase activity and nuclear protein import. Some of these processes have not been reported before as being enriched in BMR under conditions of hypoxia.

As BMR p53 has a substitution, enabling cells to escape hypoxia-induced apoptosis in favour of a reversible cell cycle arrest, we compared expression levels of 57 BMR and rat orthologues involved in the p53 signal pathway (Supplementary Data 19). Eight p53 target genes were found to be differentially regulated by hypoxia in BMR and rat (Supplementary Data 20). Sestrin (Sesn1), cyclin G (Ccn1g), Mdm2, CystC and Casp9 were downregulated in BMR but upregulated in rat, while Cenbl, cyclin D (Ccnd2) and CDK 4/6 (Cdk4) were upregulated in BMR but downregulated in rat (Fig. 3). Since the induction of CystC is a reflection of oxidative stress and the upregulation of Casp9 may point to the activation of mitochondrial apoptosis, we assume that p53 in BMR downregulates apoptosis to avoid an excessive cell loss under hypoxic conditions. Our analysis results contradict previous results on the upregulation of Mdm2 in the brain instead, we found downregulation. This discrepancy might be explained by the difference in platforms and applied analysis techniques.
The B1 SINE repeat showed significant transcriptional upregulation in hypoxia in BMR, but not in rat (Supplementary Note 7; Supplementary Table 26; Supplementary Fig. 43). Interestingly, it has recently been shown that p53 negatively controls B1 SINE repeat expression in normal mammalian cells and that weak p53 control of this expression can lead to massive transcription of B1 and B2 SINEs accompanied with activation of an interferon response (TRAIN) 'transcription of repeats activates interferon' phenomenon. Consistently, our RNA-Seq data indicate that several interferon regulatory factors, implicated in the TRAIN response, are upregulated in BMR but not in rat under hypoxic conditions (Fig. 3). As p53 in BMR is known to have a tumour-like substitution, it may be that it has a weaker SINE-inhibitory effect than rat p53, thus enabling activation of the TRAIN.

Several types of L1 elements demonstrate significant upregulation of expression in BMR under 3% of oxygen (Supplementary Fig. 44). Together with B1 upregulation, this fact could be interpreted as follows: the L1 elements facilitate the genomic propagation and insertion of non-autonomous SINE elements because L1 (LINE-1) retrotransposons are encoding reverse transcriptase (RT) proteins. This L1-derived RT is used by SINE B1 retrotransposons (that do not encode RT) for their retrotransposition. The combination of 'weak' p53, permitting transcription of SINEs and L1 (source of RT) under hypoxic conditions, and frequent exposure to hypoxic stress creates
Figure 3 | BMR adaptive complex related with hypoxia tolerance and cancer resistance. The colours of elements Ccd2, Ccd1, Cdt1, Cdc4, Cyc, Casp9, Hif1 and B1 SINEs depict expression fold changes according to RPKM, while colours of Mx1, Birc3, Irf3, Ahr, Ifn1, Tfrna and FemB represent gene copy number amplification. The green check marks on Tfrna and Ifn1 indicate evidence of positive selection.

conditions facilitating amplification of SINEs and provides a plausible explanation for high abundance of SINEs in the genome BMR viv-i-vis rodents living under normal conditions.

We inferred the genomic organizations of alpha and beta hemoglobin (HB) gene clusters from BMR (scaffolds 104 and 635, respectively; Supplementary Note 6; Supplementary Figs 45 and 46). BMR features a mutation in the proton-gated nociceptor sodium channel Nav1.7 (gene name Scn9a), which results in the replacement of a highly conserved positively charged amino-acid motif (KKV) in domain IV of Nav1.7 by the negatively charged EKD motif. Interestingly, a related mutation, resulting in a negatively charged EKE motif, was found in the NMR Nav1.7. This mutation is thought to attract protons, thereby blocking the channel and thus protecting the NMR tissue from hypercapnia-induced acid pain20. A phylogenetic interpretation of Nav1.7 sequence evolution in mammals (Fig. 4) suggests that the pain-blocking mutation is an adaptive trait, which has arisen independently in hypercapnia-exposed species by convergent evolution.

Analyses of HB-coding sequences appear to confirm the uniqueness of BMR with respect to its embryonic haemoglobin gamma (HBG) component by revealing a comparatively fast rate of HBG sequence evolution and evidence of an elevated non-synonymous to synonymous substitution ratio, indicative of positive Darwinian selection (Supplementary Figs 47 and 48). Three additional potentially reactive Cys residues in helices A, B and D of the BMR HBG T1 and T2 paralogues may in fact protect the embryonic globin from oxidation to the non-functional methHb (Fe^3+^) form, as recently reported for Cys-enriched HBB haplotypes in mouse27. The cysteine content could also affect redox reactions and HB-mediated oxidative or nitrosative stress response28.

Figure 4 | Evolution of the adaptive amino-acid sequence motif from the sodium channel nociceptor protein Nav1.7 in mammals. Note the negative net charge of the motif (as indicated by + or −) in the hypercapnia-exposed BMR (Spatica), NMR (Heterocephalus) and the cave microbat (Myotis). The adaptive trait, which confers resistance to acidosis pain, has evolved by convergence in the three distantly related hypcapnia animal lineages.

We found that NMR, not BMR, adult HBA features a mutation (Pro(44) > His(44)), which is located in the switch region of the globin and affects interaction with His(97) of the HBB chain. It is therefore tempting to speculate that the His(44) amino-acid replacement in HBA, commonly observed in the two hystricomorphs NMR and Caro porcellus, has facilitated an evolutionary adaptation of these taxa to hypoxic conditions, one in underground dwellings and the other in high-altitude habitats.
Increased mRNA and protein expression levels observed in potentially hypoxia-adaptive genes in BMR versus rat are complemented by changes on the genetic level for example, Vegf6 (Vascular endothelial growth factor B) has four copies in BMR. As growth factors regulate cell growth and division, Vegf plays a significant role in the survival of blood vessels and may therefore contribute to BMR’s adaptation to hypoxia.

Cancer resistance. BMR is resistant to both spontaneous cancer as well as the induction of tumours by chemical carcinogens. This may be because of the unique tumour suppression mechanism in which necrosis plays a major role as opposed to apoptosis normally used in many organisms. Necrosis in BMR cells is triggered by a release of interferon-beta in response to the overproliferation of cells. Our analysis of the BMR genome showed that divergent evolutionary regions, for example, darkness caused the BMR to undergo repressive (blindness) and progressive (photoperiodic) evolution. Convergent adaptations for hypercapnia and acidosis tolerance in BMR, NMR and bat were observed, while BMR and NMR followed divergent strategies in oxygen supply by respiratory proteins (Supplementary Fig. 49) and in resistance to cancer. We observed that various types of transposable elements underwent expansions in different mammalian lineages, such as the increased copy number of B1/B2 in BMR, ID elements in rats and 34 in Jaculus jaculus. Whether these expansions were adaptive and what the evolutionary pressures associated with them were, is poorly understood. Our finding that SINE B1/B2 elements underwent expansion in Spalax, and the increase in these elements’ activity during hypoxia associated with the unique anticancer adaptation provides an insight into evolutionary forces shaping the transposon landscape in the Spalax genome. The most interesting story was shown in Fig. 3. It is possible that the unique mutation on BMR p53 leads to different transcriptomic regulation patterns of several p53 target genes upon hypoxia. The expansion of B1 SINE repeats in BMR, and the significant upregulation of these expression upon hypoxia, coupled with the upregulation of Irf7 in BMR but not in rat, indicate that the BMR p53 is ineffective in suppressing B1 SINE expression, leading to activation of the TRA1N mechanism. Since normal p53 plays the key role in the accumulation of senescent cells which contribute to the aging phenotype, we speculate that weakimitated BMR p53, does not efficiently send cells to senescence and fails to suppress SINE expression and subsequent TRA1N. This results in the eradication of those cells, which in normal mouse/rat would become senescent. In addition, in the downstream of JNK, multiple genes involved in the regulation of necrosis and inflammation, including genes (Ikb, Mcl, Nhk, Tnfsfr1a, Birc3, Fm1b and Aifm1), underwent duplication events in BMR compared with mouse, rat and NMR, and genes such as Tnfsfr1a and Tnfsfr1b also show evidence of positive selection, further emphasizing that the BMR may have evolved a unique mechanism heightening necrosis and inflammatory responses to partly replace apoptosis, which contribute to its remarkable hypoxia tolerance. Cancer resistance and anti-aging. The analysis of the BMR genome and transcriptome here could open vast vistas of theoretical and applicative research programmes that will highlight how system evolution operates in nature and how it could be harnessed to cure urgent human medical challenges to support life.

Methods
Sample collection. All animal protocols were approved by the Institutional Ethics Committee. Spalax gubbi were captured in the field and housed under ambient conditions in individual cages in the Animal Facility of the Institute of Evolution, University of Haifa, with free access to vegetable and fruit at 21–25°C in a 12:12 light-dark cycle. Animals were killed with a lethal inhalation anesthesia agent (isoflurane). DNA used for genome sequencing was isolated from the whole brain of an adult female individual. The hypoxia experiments were conducted in a closed cage using the precise oxygen percentage in a balcony with the appropriate percentage from a reliable company. RNA used for hypoxia transcriptome sequencing was extracted from the whole brain.

Genome sequencing. We applied whole-genome shotgun sequencing using the Illumina HiSeq 2000 to sequence the genome of BMR. In order to reduce the read of non-randomness, 14 paired-end libraries, with insert sizes of about 250, 500, 800 bp, 2, 5, 10 and 20 kbp, were constructed. In total, we generated about 392.70 Gb of data, which 259.65 Gb (86% coverage) were retained for assembly after filtering out low quality and duplicated reads.

Genome assembly. The BMR genome was assembled de novo by SOAPdenovo v2.04.4. First, by splitting the reads from short insert size libraries (250–500 bp) into 5-mers and then merging the 31-mers, we constructed the de Bruijn graph. Second, contigs that exhibit unambiguous connections in the de Bruijn graphs were collected. Third, the paired-end information was subsequently used to link contigs into scaffolds, step by step, from short insert sizes to long insert sizes. In the last step, some intra-scaffold gaps were filled by local assembly using the reads in a read-pair, where one end uniquely aligned to a contig, while the other end was located within the gap. The total contig size and N50 were 2,911 Gp and 27,224 bp, respectively. The total scaffold size and N50 were 3,966 Gp and 3,686 bp, respectively.

Chromosome fragments. We applied reference-assisted chromosome assembly to reconstruct PAC of the BMR genome. The BMR scaffolds larger than 10 kbp in size were aligned against the mouse (mm9; reference) and human (hs37; outgroup) genomes using homoloSearch system programme. We used BWA to map the BMR mate-pair reads to BMR scaffolds. A minimum size of 30 and 80 kbp for SfA was used in two independent reconstruction experiments.

Gene annotation. We predicted the protein-coding genes in BMR using a combination of homology-based and de novo methods, as well as transcript evidence. For the homology-based prediction, human (Ensembl release 64), mouse (Ensembl release 64), rat (Ensembl release 64) and NMR protein (44) were collected and mapped on the genome using TblastN. Then, homologous genome sequences were aligned against the matching proteins using Genowin to define gene models. For de novo predictions, Augustus, GlimmerHMM, SNAP and Genematch were employed to predict coding genes. Parameters were trained based on the predicted gene from searching homologous and high-confidence transcriptomes. In addition, RNA-seq data generated for this study were mapped to the genome using TopHat and transcriptome-based gene structures were obtained by cufflinks (http:// cufflinks.cbcb.umd.edu/). Finally, all lines of gene evidence were combined together using GREAT (http://sourceforge.net/projects/great-gene/). For genes that were predicted solely as initio, only those with transcriptome coverage more than 50%...

Acknowledgements
This project was primarily financed by the Lehrs N. Tauerl Family Foundation and secondarily by the National Natural Science Foundation of China (11711109). E.N. thanks the Ansell Thrasher Research Foundation for Genetics and Molecular Evolution for financial support. L. Brodyak and S.F. were partly sponsored by the Defense Advanced Research Projects Agency [DARPA] Grant HRNR11-11-0-004. D.L., K.G., M.F., J.K., H.A.L. and J.M. were supported in part by BBiEGC grant SB0101107/1 and PL-Grid Infrastructure, T.H. and H.S. acknowledge the support by the University of Mainz Center for Computational Sciences (Irin) and the Stiftung Innovation Rheinland-Palatinate. A.A. acknowledges partial support by the Israel Cancer Association (ICA) and M.B. acknowledges partial support by the US-Israel Binational Foundation (BSF). We thank the Tel-Aviv research and development teams of the National University of Defense Technology (NUDES) for testing, optimizing and deploying the software on Tera-scale supercomputers.

Author contributions

Additional information
Accession codes: The blind mole rat (Spalax galei) whole-genome shotgun projects have been deposited in DDBJ/EMBL/GenBank nucleotide database under the accession code AK00000000. All short read data have been deposited in DDBJ/EMBL/GenBank Short Sequence Read Archive under the accession code SK099441. Raw sequencing data of transcriptomes have been deposited in the Gene Expression Omnibus under the accession code GSE49485.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://www.nature.com/reprints

How to cite this article: Yang, X. et al. Genome-wide adaptive complex to underground stress in blind mole rats Spalax. Nat. Commun. 3:9366 doi: 10.1038/ncomms9366 (2014).