

# 利用转录组数据分析高原鼫鼠和裸鼹鼠基因的趋同进化

邓小弓<sup>1,2</sup> 王堃<sup>3</sup> 张守栋<sup>1</sup> 苏建平<sup>1</sup> 张同作<sup>✉</sup> 林恭华<sup>✉</sup>

(1 中国科学院西北高原生物研究所, 中国科学院高原生物适应与进化重点实验室, 西宁 810001)

(2 中国科学院大学, 北京 100049) (3 兰州大学生命科学院, 草地农业生态系统国家重点实验室, 兰州 730000)

**摘要:** 地下鼠由于长期适应于地下洞道生活, 在诸多生物学特征上都产生了一定的趋同进化现象, 但目前尚未在分子水平分析其趋同进化特征。本研究利用高通量测序数据, 从氨基酸位点变化和基因表达量水平, 分析两种亲缘关系较远的地下鼠——高原鼫鼠和裸鼹鼠的趋同进化。氨基酸位点变化分析显示, 两种地下鼠的 54 个基因都存在至少一个理论趋同进化位点, 其中 13 个基因可以找到相对明确的表型功能注释。基因表达量分析显示, 两种地下鼠在转录组水平的总体表达模式上存在趋同效应; 有 103 个基因在两种地下鼠脑中的表达量是作为对照的地上鼠 (小鼠、豚鼠) 表达量的 3 倍以上, 其中有 20 个基因可以找到相对明确的表型功能注释。本研究的结果表明, 高原鼫鼠和裸鼹鼠为了适应地下生活方式, 在氨基酸序列和基因表达量水平上都产生了一定程度的趋同进化。其中, 低氧适应相关基因 (氧运输、氧代谢、DNA 修复等) 是地下鼠进化适应的最重要组成部分, 挖掘器官 (尤其是肌肉) 功能相关基因的重要性次之, 能量代谢及平衡相关基因也是地下鼠适应性进化的关键要素。

**关键词:** 地下鼠; 趋同进化; 低氧适应; 挖掘器官; 能量代谢

中图分类号: Q346.3

文献标识码: A

文章编号: 1000 - 1050 (2014) 02 - 0129 - 09

## Transcriptomic determination of convergent evolution between plateau zokors (*Eospalax baileyi*) and naked mole rats (*Heterocephalus glaber*)

DENG Xiaogong<sup>1,2</sup>, WANG Kun<sup>3</sup>, ZHANG Shoudong<sup>1</sup>, SU Jianping<sup>1</sup>, ZHANG Tongzuo<sup>✉</sup>, LIN Gong-hua<sup>✉</sup>

(1 Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China)

(2 University of Chinese Academy of Sciences, Beijing 100049, China)

(3 State Key Laboratory of Grassland Agro-ecosystem, College of Life Science, Lanzhou University, Lanzhou 730000, China)

**Abstract:** Subterranean rodents have evolved convergent traits in many of their biological characteristics that allow them to adapt to a similar underground burrowing life style. However, no study has yet been published about this convergent evolution at the molecular level. Here, we use high-throughput next-generation sequencing data (transcriptomes) to identify both amino acid (AA) variations and gene expression levels in order to analyze the convergent evolution of two typical subterranean rodents, the plateau zokor (*Eospalax baileyi*) and the naked mole rat (*Heterocephalus glaber*). AA variation analysis detected 54 genes of which each had at least one convergent adaptation site; of these genes 13 were successfully annotated by their relatively obvious phenotypical functions. Within the 13 genes, four (*Ptpn6*, *Sco1*, *Dhcr 24*, and *Add2*) were related with hypoxia tolerance or adaptation to oxidative stress, four (*Mbtps1*, *Atp1a1*, *Acs13*, and *Ncoa2*) were involved in nutrient transport or energy metabolism, four (*Smoc1*, *Kif1b*, *Tcap*, and *Cryab*) were involved in limb development or muscle health, and the remaining gene (*Myo6*) was involved in auditory function. Gene expression level analyses of the two subterranean rodents detected 103 genes that were significantly upregulated (>3 fold) as compared to mouse (*Mus musculus*) and guinea pig (*Cavia porcellus*); 20 of these genes could be successfully annotated by their phenotypical

基金项目: 国家自然科学基金资助项目 (31372197, 31370405); 中国科学院知识创新工程领域前沿项目 (KSCX2 - EW - J - 26)

作者简介: 邓小弓 (1988 -), 男, 硕士研究生, 主要从事动物生态学研究。

收稿日期: 2013 - 09 - 23; 修回日期: 2013 - 12 - 12

\* 通讯作者, Corresponding authors, E-mail: zhangtz@nwipb.cas.cn; lingonghua@gmail.com

functions. Within these 20 genes, 12 were related with hypoxia tolerance functions, such as oxygen transport (*Dsg2*, *Tab2*, *Hp*, *Rock1*, and *F5*), oxygen metabolism (*Tfb1m* and *Mrpl1*), and DNA repair (*Tdp2*, *Rad50*, *Alkbh2*, *Hltf*, and *Plk1*), 3 were related with the functional integrity of muscle (*Sync* and *Taf3*) and bone (*Acp5*), and the other 5 were involved in immune (*Spon2* and *Sh2d4a*), biological rhythm (*Clock*), starvation tolerance (*Pdk4*), and reproduction (*Clgn*). Our results show that the hypoxia tolerance-related genes were the most important adaptation genes found in the two subterranean rodents, followed by the digging apparatus-related genes (especially the skeletal muscle) in addition to the genes important in energy metabolism and energy balance. This paper provides evidence for convergence of subterranean rodents at the molecular level.

**Key words:** Convergent evolution; Digging apparatus; Energy metabolism; Hypoxia tolerance; Subterranean rodent

地下鼠是指大部分时间生活在地面之下, 并且已经适应了地下环境的一类特殊的啮齿动物 (Lacey *et al.*, 2000)。全世界大约有 250 种, 隶属于 8 科 38 属, 除大洋洲和南极洲之外的所有大陆都有广泛分布 (Begall *et al.*, 2007)。由于其特殊的生活方式, 以及由此带来的丰富的进化适应特征, 地下鼠已经成为国内外动物进化适应的重要研究对象。目前已有大量关于这类动物的生态生理学、感官进化、行为学、生活史等方面的研究报道, 为了解动物进化与适应提供了丰富的信息来源 (Nevo, 1999; Begall *et al.*, 2007; 林恭华, 2010)。近年来, 随着分子生物学研究手段的进步, 许多学者开始从分子进化的角度揭示地下鼠的进化适应机制 (Ashur-Fabian *et al.*, 2004; Avivi *et al.*, 2010; 朱瑞娟等, 2012; Tian *et al.*, 2013)。地下鼠由于长期适应地下洞道生活, 在诸多生物学属性上都产生了一定的趋同特征, 如都具有强壮的挖掘器官、退化的视觉系统和较低的基础代谢率 (Nevo, 1999)。然而, 这方面的研究绝大多数仅限于宏观水平 (形态、生理、行为等) 的分析, 而在分子水平揭示趋同进化现象的研究, 至今未见报道。可能的原因是, 在分子水平产生趋同进化的基因座位在基因组中所占比例较小, 而目前绝大部分相关研究着眼于单个或少数几个基因, 不足以发现趋同进化现象。

以基因组和转录组测序为代表的高通量测序技术是近年来发展起来的新一代测序技术 (Chan and Ragan, 2013), 其优势是产出数据量大、数据较为可靠, 一次测序过程可得到巨量的准确度较高的序列信息, 这就为我们全面了解一个物种的分子进化提供了充足的数据基础。当前, 高通量测序分析技术已被应用到地下鼠的分子进化研究中。裸鼯鼠 (*Heterocephalus glaber*) 隶属于啮齿目 (Rodentia),

豪猪亚目 (Hystricomorpha), 滨鼠科 (Bathyergidae), 是典型的地下鼠 (Jarvis and Sherman, 2002)。Kim 等 (2011) 对裸鼯鼠进行了全基因组和若干个组织的转录组测序, 检测到其肿瘤抑制、体温调节、低氧适应等相关基因有适应性进化的趋势, 而视觉、味觉等相关基因则出现一定程度的退化现象。相关数据的发表, 迎来了地下鼠研究的基因组时代, 也为我们进行物种间比较, 从而挖掘趋同进化信息提供了重要的可能性。高原鼯鼠 (*Eospalax baileyi*) 隶属于啮齿目 (Rodentia), 鼠形亚目 (Myomorpha), 鼯鼠科 (Spalacidae) (Norris *et al.*, 2004), 主要分布于青藏高原东部的草甸、灌丛、农田等生境类型, 终生营地下生活, 是一种典型的地下鼠 (Zhang *et al.*, 2003)。高原鼯鼠作为青藏高原地区特有的地下鼠, 已成为国内探讨动物进化与适应机制的重要研究对象 (Wang and Zhang, 2012; Sun *et al.*, 2013)。然而, 与其他大部分地下鼠的状况类似, 目前对于高原鼯鼠的研究也仅止于单个或少数几个基因; 同时, 这些研究也没有通过种间比较, 进而分析高原鼯鼠与其他地下鼠的分子进化与适应机制上有无异同点。这在一定程度上制约了我们对这类动物进化适应研究的认识水平。本研究对高原鼯鼠脑和肝脏的转录组进行测序, 同时结合裸鼯鼠的数据, 从序列位点选择和基因表达量两个方面系统分析两种地下鼠的分子趋同进化特征。通过对数千个直系同源基因的分析, 筛选高原鼯鼠和裸鼯鼠分子进化的共同点, 进一步阐明地下鼠进化适应特征的分子机制。

## 1 研究方法

### 1.1 转录组测序和组装

2012 年 10 月在青海省大通县向化乡 (北纬 37°7'30", 东经 101°48'42") 用自制鼯鼠活捕箭

(专利号: ZL201120429596.9) 捕捉雄性高原鼯鼠 1 只, 乙醚麻醉后迅速解剖, 取脑及肝脏组织保存于液氮中。在实验室用 Trizol 试剂盒 (Invitrogen, CA, USA) 提取总 RNA, 用干冰送样至华大基因公司构建片段化的 cDNA 文库。用 Illumina Hiseq 2000 进行高通量测序, 得到 raw reads 文件并对其进行数据评估、过滤 (去除接头、低质量区域等), 最终得到用于生物信息学分析的 clean reads (已提交至 GenBank 中的 SRA 数据库, 登录号: SRX320115 and SRX321868)。在 Linux 服务器中将脑和肝脏的数据合并, 用 Trinity 软件 (Grabherr *et al.*, 2011) 对 clean reads 进行拼接组装, 后用 CD-HIT-EST 软件 (Li and Godzik, 2006) 进行进一步组装和去重, 得到 unigenes 片段。用 Trinity 软件中的 Perl 脚本 transcripts\_to\_best\_scoring\_ORFs.pl 提取基因编码区序列 (cds) 并翻译其对应的多肽序列 (pep)。

### 1.2 直系同源序列提取和比对

直系同源序列 (orthologs) 是指位于不同物种间, 在物种形成过程中源自某一共同祖先的基因序列, 一般具有相同的功能, 是研究物种间序列进化的基础。本研究首先提取高原鼯鼠、裸鼯鼠、小鼠 (*Mus musculus*, 地上鼠, 与高原鼯鼠亲缘关系最近的模式动物)、豚鼠 (*Cavia porcellus*, 地上鼠, 与裸鼯鼠亲缘关系最近的模式动物) 和人 (*Homo sapiens*, 作为外类群) 的 orthologs 序列。从 GiGaDB 数据库 (<http://gigadb.org/dataset/100022>; version 1.7-2) 中下载裸鼯鼠基因 cds 序列和 pep 序列; 从 Ensembl 数据库下载豚鼠的 cds 和 pep 数据; 从 GenBank 的 CCDS 数据库 (Pruitt *et al.*, 2009) 下载小鼠和人的 cds 和 pep 数据。

用 Inparanoid (O'Brien *et al.*, 2005) 软件对上述 5 个物种的 pep 序列提取两两物种的 orthologs 序列对应信息, 之后用 Multiparanoid 软件 (Alexeyenko *et al.*, 2006) 对 Inparanoid 软件得到的结果进行归总, 得到 5 个物种 orthologs 的对应序列号。利用这个序列号和 cdbfasta 软件 (<http://sourceforge.net/projects/cdbfasta/>) 提取每个基因对应的编码区 cds 序列, 分别写入单独的文本中 (fasta 格式)。依次用 MACSE (Ranwez *et al.*, 2011) 和 prank (Loytynoja and Goldman, 2005) 两个软件对每个基因对应的 orthologs 序列进行比

对。用 trimAl 软件 (Capella-Gutierrez *et al.*, 2009) 清除未比对上的区域, 得到每个基因对应的 5 个物种等长的直系同源 cds 序列 (只采纳最终长度大于 200 bp 的基因), 并用 EMBOSS 软件包 (Rice *et al.*, 2000) 中的 transeq 程序翻译得到 pep 序列。为了排除 orthologs 提取、比对过程中的错误, 以及排除同源异构体 (isoform) 带来的影响, 我们用 Phyml 软件对每个基因的 orthologs 构建系统发育树, 排除与物种树 (图 1) 不符的基因。由于涉及数千个基因, 所有以上步骤都以 Perl 脚本调用外部命令的形式批量进行。

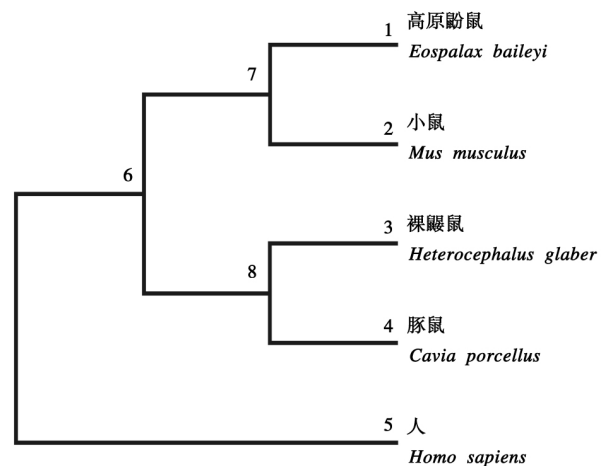


图 1 本研究涉及的物种树(数字表示节点)

Fig. 1 Phylogenetic tree used in this study (the numbers show the nodes)

### 1.3 趋同进化位点检测

利用自编的 Perl 脚本提取每个基因的 pep 序列中潜在的地下鼠趋同进化位点 (2 种地下鼠的氨基酸相同, 但与 2 种地上鼠及人的氨基酸都不相同)。为了排除随机因素带来的“假阳性”, 我们同时考察位点的上下游各 5 个氨基酸的序列 (包括考察位点共 11 个氨基酸), 只采纳所有物种之间这 11 个氨基酸序列都保守的位点。用 Ancestor 软件 (Zhang and Nei, 1997) 预测每个具有至少一个潜在趋同进化位点的基因在系统树节点上 (7 和 8, 图 1) 的祖先序列, 分析节点 7→1 和 8→3 的氨基酸变化, 判断是否存在理论上的趋同进化位点。为了提高模型的准确度, 在节点氨基酸的推测过程中, 对后验概率低于 90% 的结果不予考虑。利用 Converge 软件 (Zhang and Kumar, 1997) 检

测这些趋同进化的判断是否显著,  $P > 0.05$  的予以舍弃。以上所有条件都满足的位点定义为理论趋同进化位点 (theoretical convergent site, TCS)。最后, 对存在至少一个 TCS 位点的基因, 在 Genecards 数据库 (Rebhan *et al.*, 1997) 中搜索鉴定 (结合 Entrez Gene summary 和 UniProtKB/Swiss-Prot 的描述进行综合判断) 其功能。

#### 1.4 基因表达量分析

转录组测序的对象是特定组织或细胞在某一发育阶段或功能状态下转录出来的绝大部分 RNA (主要包括 mRNA 和非编码 RNA), 因此, 转录组数据除了可以通过组装得到大量的 unigenes 片段外, 还可通过统计相关读段 (reads) 数计算出不同 RNA 的表达量 (祁云霞, 2011)。现有的转录组研究中, 脑组织作为功能最复杂的器官往往成为首选研究对象。我们利用地上鼠和地下鼠脑组织转录组数据, 提取各转录本的表达量信息, 检验地下鼠在基因表达模式上是否有趋同特征。由于裸鼯鼠的采样个体为雌性 (Kim *et al.*, 2011), 而本研究中的高原鼯鼠个体为雄性, 为了确定性别对基因表达的影响程度, 我们同时选取两种性别的豚鼠和小鼠转录组。

从 GenBank 的 SRA 数据库中下载裸鼯鼠 (SRR306401)、豚鼠 (雄性: ERR266376, 雌性: ERR266394) 和小鼠 (雄性: SRR306762, 雌性: SRR306757) 转录组 short reads 数据, 利用 bowtie、express 软件和 CDS 序列, 从转录组中提取各个物种对应基因的读段数 (counts)。利用 SPSS 20.0 软件对所有基因的 counts 数据进行转录组样品间分层聚类 (选用非参数的 Counts 计数变量及卡方测度方法, 其余设置为软件默认值), 用以比较物种或个体之间基因表达模式的总体异同。最后, 将每个基因的 counts 信息读入 EBSeq 软件, 检测每个基因在两个地下鼠转录组和 4 个地上鼠转录组数据之间的差异表达基因及其后验差异倍数 (posterior fold change, PosteriorFC) 和 FDR 矫正 (Benjamini and Hochberg, 1995) 后验概率 (PPDE)。最后, 对在两种地下鼠中的表达量都大于地上鼠 3 倍以上 (PosteriorFC > 3) 且 FDR 矫正显著 (PPDE > 0.95) 的基因, 在 Genecards 数据库中鉴定其功能。

## 2 结果

### 2.1 转录组组装和直系同源序列提取

对高原鼯鼠的脑和肝脏转录组组装后, 一共得到 138 872 条 unigenes 序列, 经 transcripts\_to\_best\_scoring\_ORFs.pl 识别, 提取 36 031 条 cds 序列, 平均长度为 975 bp, 50% 分位点 (N50) 为 1 322 bp。

通过 5 个物种的比对, 共找到 7 195 个基因的 orthologs。经 MACSE 和 prank 软件比对和 trimAl 软件清理, 排除 57 个 200 bp 以下的序列; 经 Phyml 建树比较, 进一步排除了 2 204 个系统发育树与图 1 所示树形不完全一致的基因。最后, 得到 4 934 个满足所有筛选条件的可用基因, 长度范围 201 - 7 569 bp, 平均长度为 1 269 bp。

### 2.2 地下鼠位点趋同进化

共检测到 54 个基因存在满足所有筛选条件的 TCS 位点, 其中 2 个基因各存在 2 个 TCS 位点。所有这些位点上, 3 种非地下鼠 3 个节点 (node6 - 8) 的氨基酸完全相同, 且不同于 2 种地下鼠的氨基酸。包括 18 种氨基酸变异方式, 其中最多的 3 种是 VI (前一个字母表示地下鼠的氨基酸, 后一个字母表示地上鼠及节点) 或 IV (12 个位点)、AT 或 TA (9 个位点)、NS 或 SN (7 个位点), 其他 15 种变异方式对应的 TCS 位点数都  $\leq 3$ 。

经 Genecards 搜索, 其中 43 个基因功能 (或部分功能) 已知, 其中 30 个基因属于常规催化或上游调控类, 参与的功能过程很多, 与表型的关系不易确定, 本文未做详细探讨。其余 13 个基因的功能与表型对应关系相对明确, 分别与肌肉健康 (*Kif1b*、*Tcap*、*Cryab*)、血液循环 (*Scol*、*Dhcr24*)、氧代谢 (*Ptpn6*、*Add2*) 和能量平衡 (*Acsl3*、*Ncoa2*) 等存在密切关联。另有 4 个分别与眼睛和四肢发育 (*Smoc1*)、脂类代谢 (*Mbtps1*)、营养物质运输 (*Atp1a1*) 及听觉 (*Myo6*) 功能有密切联系。具体突变信息和功能描述在表 1 中列出。

### 2.3 高原鼯鼠与裸鼯鼠的基因表达特征

共检测到 103 个在两种地下鼠中上调表达的基因 (PosteriorFC > 3 且 PPDE > 0.95), 这些基因的平均上调倍数 (即 PosteriorFC 值) 为 7.07, 其中上调最大的 3 个基因分别是 *Hp* (43.46 倍)、*Rnase10*

表 1 具有地下鼠趋同进化氨基酸位点的基因及功能注释 (基因功能不明确的未列出)

Table 1 Genes which have convergent evolutionary amino acid (AA) site and their functions (those with unclear functions are not listed)

基因流水号 Gene No.	小鼠 CCDS CCDS sequence of mouse	位点变异 AA variation*	基因名称 Gene name	功能简述 Gene function
gene248	CCDS26017.1	A342V	<i>Smoc1</i>	眼睛和四肢发育 Plays essential roles in both eye and limb development
gene385	CCDS51908.1	M514I	<i>Ptpn6</i>	造血功能 Plays a key role in hematopoiesis
gene829	CCDS22707.1	S544N	<i>Mbtps1</i>	酯类代谢 Has a central role in the regulation of lipid metabolism in cells
gene1169	CCDS17683.1	V109I	<i>Atp1a1</i>	营养物质主动运输 Providing the energy for active transport of various nutrients
gene1743	CCDS18956.1	S32N, A388T	<i>Kif1b</i>	肌肉健康 Mutations in this gene cause Charcot-Marie-Tooth disease, type 2A1
gene1829	CCDS23368.1	V130I	<i>Myo6</i>	听觉相关 Required for structural integrity of inner ear hair cells
gene1915	CCDS36183.1	I208V	<i>Sco1</i>	有氧呼吸 Catalyzes the transfer of reducing equivalents from cytochrome c to molecular oxygen and pumps protons across the inner mitochondrial membrane
gene2927	CCDS15087.1	A458T	<i>Acs13</i>	能量代谢 Has mainly an anabolic role in energy metabolism
gene3061	CCDS18420.1	Q316R	<i>Dhcr24</i>	抗氧化应激 Protects cells from oxidative stress by reducing caspase 3 activity during apoptosis induced by oxidative stress
gene5411	CCDS35515.1	T1070A	<i>Ncoa2</i>	能量平衡 Required with NCOA1 to control energy balance between white and brown adipose tissues
gene5432	CCDS25346.1	S145P	<i>Tcap</i>	肌肉健康 Mutations in this gene are associated with limb-girdle muscular dystrophy type 2G
gene6656	CCDS23172.1	M70L	<i>Cryab</i>	肌肉健康 A missense mutation co-segregated in a family with a desmin-related myopathy
gene7012	CCDS20308.1	N178S	<i>Add2</i>	血压调节 Associated with the regulation of blood pressure in an animal model of hypertension

\* 数字表示对应于小鼠 CCDS 序列上的氨基酸位点, 数字前的字母表示地下鼠氨基酸, 数字后的字母表示其他 3 个物种 (小鼠、豚鼠、人) 及节点 6-8 上 (图 1) 的氨基酸

\* The numbers show the AA sites according to the corresponding mouse CCDS sequences, the letters ahead the number are the AAs of the subterranean rodents, the letters behind the numbers are the AAs of the other three species (mouse, guinea pig, and human) as well as the ancestor nodes 6-8 in Fig. 1

(33.72 倍) 和 *Spon2* (29.52 倍)。SPSS 分层聚类分析显示, 雌、雄小鼠聚成一枝, 雌、雄豚鼠聚成另一枝, 之后小鼠和豚鼠聚成一个大枝; 而高原鼯鼠和裸鼯鼠则聚成另一大枝 (图 2)。

经 Genecards 搜索, 其中 80 个基因功能 (或部分功能) 已知, 但大部分 (60 个) 基因属于常规催化或上游调控类, 与表型的关系不易确定。其余 20 个有相对明确表型对应特征的基因功能如表 2 所示, 其中有 5 个基因 (*Tdp2*、*Rad50*、*Alkbh2*、*Hltf*、*Plk1*) 与 DNA 修复相关, 3 个 (*Rock1*、*Hp*、*F5*) 与红细胞功能相关, 3 个 (*Sync*、*Acp5*、*Taf3*) 与肌肉或骨骼功能相关, 其他基因则分别与

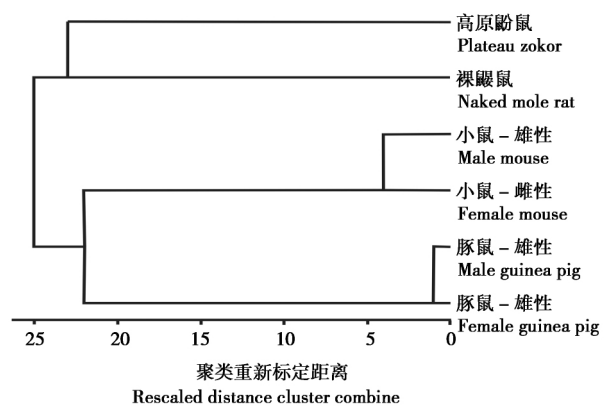


图 2 基于 4 种动物脑转录组基因表达量的分层聚类  
Fig. 2 Hierarchical clustering of transcriptomes of the four rodents

心脏 (*Tab2*)、线粒体 (*Tfb1m*、*Mrpl1*)、免疫 (*Clock*) 和饥饿耐受 (*Pdk4*) 等功能相关联 (表 2)。  
(*Spon2*、*Sh2d4a*)、生殖 (*Clgn*)、生物节律 (2)。

表 2 地下鼠趋同上调表达的基因及其功能注释 (基因功能不明确的未列出)

Table 2 Genes which had expression levels in the two subterranean rodents at least three times greater than those in the two above-ground rodents (those with unclear functions are not listed)

流水号 Gene No.	小鼠 CCDS CCDS sequence of mouse	表达量差异倍数 PosteriorFC	基因名称 Gene name	功能简述 Gene function
gene4874	CCDS40470.1	43.46	<i>Hp</i>	游离血红蛋白回收 Bind and help to degradate free plasma hemoglobin, preventing loss of iron through the kidneys and protecting the kidneys from damage by hemoglobin
gene153	CCDS19200.1	29.52	<i>Spon2</i>	免疫 Functions as an opsonin for macrophage phagocytosis of bacteria
gene6212	CCDS35754.1	11.25	<i>F5</i>	止血 Central regulator of hemostasis
gene4420	CCDS22452.1	8.94	<i>Clgn</i>	生殖 Probably plays an important role in spermatogenesis
gene436	CCDS24684.1	8.32	<i>Rad50</i>	DNA 修复 Double-strand break (DSB) repair of DNA
gene5630	CCDS52564.1	7.08	<i>Sh2d4a</i>	免疫 May play a role in T-cell development and function
gene4080	CCDS19902.1	6.95	<i>Pdk4</i>	饥饿耐受 Plays an important role in maintaining normal blood glucose levels under starvation
gene1222	CCDS19559.1	6.21	<i>Alkbh2</i>	DNA 修复 Dioxygenase that repairs alkylated DNA and RNA
gene5040	CCDS22923.1	4.90	<i>Acp5</i>	骨骼功能 Involved in osteopontin/bone sialoprotein dephosphorylation
gene6004	CCDS15675.1	4.00	<i>Taf3</i>	肌细胞分化 Required in complex with TBPL2 for the differentiation of myoblasts into myocytes
gene6405	CCDS19449.1	3.99	<i>Mrpl1</i>	线粒体蛋白合成 Help in protein synthesis within the mitochondrion
gene2829	CCDS19360.1	3.75	<i>Clock</i>	生物节律 Activate E-box element transcription of a number of proteins of the circadian clock
gene1238	CCDS17261.1	3.54	<i>Hlf</i>	DNA 修复 Plays a role in error-free post-replication repair (PRR) of damaged DNA
gene236	CCDS26381.1	3.52	<i>Tdp2</i>	DNA 修复 Involved in both DNA repair and the activation of transcription factors
gene4192	CCDS28363.1	3.37	<i>Tfb1m</i>	线粒体 DNA 转录 Required for basal transcription of mitochondrial DNA
gene1158	CCDS29053.2	3.30	<i>Rock1</i>	红细胞分化 Plays a role in terminal erythroid differentiation
gene5627	CCDS23691.1	3.28	<i>Tab2</i>	心脏发育 Involved in heart development
gene3381	CCDS21812.1	3.26	<i>Plk1</i>	DNA 修复 Also required for recovery after DNA damage checkpoint and entry into mitosis
gene3010	CCDS18687.1	3.13	<i>Sync</i>	肌肉功能 May have a structural role in striated muscle
gene3198	CCDS29084.1	3.12	<i>Dsg2</i>	心脏健康 Mutations is associated with arrhythmogenic right ventricular dysplasia, familial, 10

### 3 讨论

本研究利用高通量测序数据分析了高原鼯鼠和裸鼯鼠的趋同进化, 涉及数千个功能基因的分析。

在氨基酸位点变异水平上, 检测到 43 个基因存在理论趋同进化位点。需要说明的是, 本研究所使用的 Converge 软件, 其作者将氨基酸位点的趋同变化分为趋同 (convergent) 和平行 (parallel) 两种

进化模式 (Zhang and Kumar, 1997)。以本文的图 1 为例, 如果节点  $7 \rightarrow 1 = A \rightarrow S$  且  $8 \rightarrow 3 = T \rightarrow S$ , 则定义为趋同进化; 而  $7 \rightarrow 1$  和  $8 \rightarrow 3$  都为  $A \rightarrow S$  时, 定义为平行进化。根据这一定义, 本研究中检测到的所有 56 个位点都应属于平行进化位点, 这可能与本研究中所涉及的物种树有关。值得注意的是, 趋同进化和平行进化本来就是两个界限比较模糊的概念 (Arendt and Reznick, 2008)。按照经典的定义, 当具有相似生物学性状的 2 个物种, 其各自祖先如果也有相似性状则定义为平行进化, 而其 2 个祖先的相关性状很不相同时则定义为趋同进化。然而, 由于这个定义比较模糊, 在实际应用中通常演变为两类判定方式: 方式 1, 基于物种之间的亲缘关系判定, 即两种亲缘关系近的物种具有相似性状的称为平行进化, 而亲缘关系远的物种具有相似性状的称为趋同进化; 方式 2, 基于遗传机制判定, 即两个物种通过相同遗传性状 (相同基因或相同调控通路等) 的改变而产生相似表型的称为平行进化, 而通过不同遗传性状的改变而产生相似表型的称为趋同进化 (Arendt and Reznick, 2008)。在以位点改变为分析对象的研究中, 其数据基础为直系同源基因, 一般认为具有相同或类似的功能, 因此方式 2 在比较同一大类的不同物种间 (如本例的地下鼠和地上鼠之间) 的进化差异上不适用。换句话说, 地下鼠是否存在位点上的趋同或平行进化, 只能通过方式 1 来界定。本研究涉及的两种地下鼠——高原鼯鼠和裸鼯鼠, 分别隶属于鼠形亚目和豪猪亚目, 两者在中始新世 (约 5 千万年前) (Rose, 2006) 即已分化, 应该视为亲缘关系较远的物种。且这两种地下鼠都面临着低氧适应、挖掘和能量代谢等方面的共同压力。因此, 本研究中检测到的所有位点可以认为是趋同进化位点。

本研究检测到的有趋同进化位点的基因中, 其中 13 个基因与表型性状有相对明确的关联 (表 1)。其中 4 个基因分别与造血功能 (*Ptpn6*)、有氧呼吸 (*Sco1*)、抗氧化应激 (*Dhcr24*)、血压调节 (*Add2*) 相关, 这些功能本质上都与有氧呼吸代谢相关。作为典型的地下鼠, 高原鼯鼠和裸鼯鼠长期生活在缺氧的地下洞道中, 势必都需要在氧气的运输、利用及对抗氧自由基的损伤应激方面产生分子水平的进化适应。有 4 个基因分别与营养物质的代谢 (*Mbtps1*)、运输 (*Atp1a1*) 及能量代谢 (*Ac-*

*sl3*) 与平衡 (*Ncoa2*) 相关。地下鼠所处的环境决定其取食难度远高于地面动物, 一方面, 其觅食范围有限导致觅食效率严重不足, 另一方面, 在觅食过程中要进行挖掘活动, 耗费大量的能量。这些胁迫导致了高原鼯鼠和裸鼯鼠都需要在营养物质和能量的高效利用上产生趋同进化。另有 4 个基因与四肢发育 (*Smoc1*) 和肌肉健康 (*Kif1b*、*Tcap*、*Cry-ab*) 相关, 充分显示了挖掘器官在地下鼠生存方面起到关键作用。地下鼠的挖掘过程中所消耗的能量损耗率是静息代谢时的 3–5 倍以上 (Jarvis and Sherman, 2002), 而挖掘一段距离的洞道其能耗可以达到地上奔跑同样距离的 360–3 400 倍 (Vleck, 1979)。因此, 挖掘器官尤其是肌肉组织的健康对地下鼠的生存尤为重要, 也是产生地下鼠趋同进化的关键因子之一。此外, 还有 1 个与听觉相关的基因产生趋同进化, 这可能与地下洞道环境中其听阈范围需要产生适应性的改变有关。

本研究中的基因表达量分析检测到 103 个基因同时在高原鼯鼠和裸鼯鼠中发生上调, 其中有 20 个基因与表型有相对明确的对应关系 (表 2)。与位点趋同进化的分析结果相似, 其中有 7 个基因分别与心脏功能 (*Dsg2*、*Tab2*)、红细胞功能 (*Hp*、*Rock1*、*F5*)、有氧呼吸 (*Tfb1m*、*Mrpl1*) 关联, 这些都体现了氧的输送与代谢在地下鼠生存上的关键作用; 有 3 个基因与肌肉 (*Sync*、*Taf3*) 和骨骼 (*Acp5*) 的功能相关, 体现了挖掘相关组织的进化适应对地下鼠生存的重要意义。有趣的是, 另有 5 个基因 (*Tdp2*、*Rad50*、*Alkbh2*、*Hltf*、*Plk1*) 与 DNA 的修复相关。组织低氧通常导致细胞产生大量的氧自由基 (reactive oxygen species, ROS) (Clanton, 2007), 从而对 DNA 产生损坏, 这一点已在裸鼯鼠身上得到验证 (Andziak *et al.*, 2006)。而有效的 DNA 修复功能则可以在一定程度上减少细胞损伤, 从而挽救细胞命运。Kim 等 (2011) 对裸鼯鼠的研究找到一个受到正选择的 DNA 修复基因 (*APEX1*)。Shams 等 (2013) 对另一种地下鼠——鼯形鼠 (*Spalax ehrenbergi*) 的研究显示, 低氧可以上调其 DNA 修复基因 (*p53R2*) 的表达, 提示 DNA 修复功能对地下鼠在低氧下的生存具有重要意义。此外, 另有 5 个基因分别与免疫、生物节律等相关, 由于每项功能涉及的基因数较少, 在此不做进一步讨论。

对 4 934 个基因表达量的聚类分析显示, 尽管高原鼯鼠和小鼠、裸鼯鼠和豚鼠亲缘关系更近 (图 1), 但总体基因表达特征上, 两类地上鼠聚成一枝, 两类地下鼠聚成另一枝 (图 2), 充分体现地上/地下生活方式在基因表达上的趋同效应。另外, 雌、雄小鼠和雌、雄豚鼠分别聚成单独小枝, 证明了性别对转录组表达的影响不至于超过物种间的差异, 这在一定程度上显示我们的分析方法的可靠性。当前, 高通量测序技术的发展, 产生了巨量的动物基因序列数据, 趋同进化的研究也开始进入组学水平。如对回声定位物种蝙蝠 (*Rhinolophus ferrumequinum*, *Megaderma lyra*) 和海豚 (*Tursiops truncatus*) 基因组进行分析, 找到大量的基因涉及趋同进化 (Parker *et al.*, 2013)。然而, 基于基因表达量水平的趋同进化研究, 目前仍鲜见报道。本研究利用转录组数据提取表达量信息, 进行物种间聚类, 同时筛选在表达水平具有趋同进化的基因, 可为趋同进化研究提供新的数据支持和研究思路。

综上所述, 高原鼯鼠和裸鼯鼠为了适应地下生活方式, 在氨基酸序列和基因表达量水平上都产生了一定程度的趋同进化。其中, 低氧适应相关基因 (氧运输、代谢、DNA 修复等) 为地下鼠进化适应的最重要组分, 挖掘器官或组织 (尤其是肌肉健康) 相关基因次之, 能量代谢及平衡的相关基因也是地下鼠适应性进化的关键要素。如前所述, 地下鼠是一个较为复杂的生物类群, 分散在多个不同的分类单元 (8 科 38 属), 显示地下生活方式具有在不同地区不同时间多次起源的特点。尽管如此, 所有地下鼠都需面临低氧、挖掘洞道的能耗和食物匮乏等一系列生存胁迫, 这就导致地下鼠在形态、生理和行为等水平上发生趋同进化以适应相同的生存压力 (Nevo, 1999)。本研究的结论则首次证明, 地下鼠的趋同进化特征在分子水平上也可以得到显著体现。与此同时, 我们的研究发现, 低氧适应、挖掘和能量代谢相关基因占趋同进化基因的绝大部分, 这在一定程度上表明这三类基因对应的环境压力是地下鼠趋同进化的关键因素。需要指出的是, 为了保证有足够多的直系同源基因进行分析, 本研究仅涉及两种地下鼠, 不足以揭示地下鼠的所有趋同进化特征。而随着高通量技术的发展, 越来越多的地下鼠基因组将被测序和注释, 关于地下鼠

趋同进化规律的结论也将会愈加丰富和可靠。

#### 参考文献:

- Alexeyenko A, Tamas I, Liu G, Sonnhammer E L. 2006. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics*, **22** (14): e9 – e15.
- Andziak B, O' Connor T P, Qi W, DeWaal E M, Pierce A, Chaudhuri A R, Remmen H V, Buffenstein R. 2006. High oxidative damage levels in the longest-living rodent, the naked mole-rat. *Aging Cell*, **5** (6): 463 – 471.
- Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, **23** (1): 26 – 32.
- Ashur-Fabian O, Avivi A, Trakhtenbrot L, Adamsky K, Cohen M, Kajakaro G, Joel A, Amariglio N, Nevo E, Rechavi G. 2004. Evolution of p53 in hypoxia-stressed *Spalax* mimics human tumor mutation. *Proc Natl Acad Sci USA*, **101**: 12236 – 12241.
- Avivi A, Gerlach F, Joel A, Reuss S, Burmester T, Nevo E, Hankeln T. 2010. Neuroglobin, cytoglobin and myoglobin contribute to the hypoxia-adaptation of the subterranean mole rat *Spalax*. *Proc Natl Acad Sci USA*, **107**: 21570 – 21575.
- Begall S, Burda H, Schleich C E. 2007. Subterranean Rodents: News from Underground. Berlin: Springer Berlin Heidelberg.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*, **57**: 289 – 300.
- Capella-Gutierrez S, Silla-Martinez J M, Gabaldon T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25** (15): 1972 – 1973.
- Chan C X, Ragan M A. 2013. Next-generation phylogenomics. *Biol Direct*, **8**: 3.
- Clanton T L. 2007. Hypoxia-induced reactive oxygen species formation in skeletal muscle. *Journal of Applied Physiology*, **102** (6): 2379 – 2388.
- Grabherr M G, Haas B J, Yassour M, Levin J Z, Thompson D A, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q D, Chen Z H, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma F D, Birren B W, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, **29** (7): 644 – U130.
- Jarvis J U M, Sherman P W. 2002. *Heterocephalus glaber*. *Mammalian Species*, **706**: 1 – 9.
- Kim E B, Fang X D, Fushan A A, Huang Z Y, Lobanov A V, Han L J, Marino S M, Sun X Q, Turanov A A, Yang P C, Yim S H, Zhao X, Kasaikina M V, Stoletzki N, Peng C F, Polak P, Xiong Z Q, Kiezun A, Zhu Y B, Chen Y X, Kryukov G V, Zhang Q, Peshkin L, Yang L, Bronson R T, Buffenstein R, Wang B, Han C L, Li Q Y, Chen L, Zhao W, Sunyaev S R, Park T J, Zhang G J, Wang J, Gladyshev V N. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*,



- 479: 223 – 227.
- Lacey E A , Patton J , Cameron G. 2000. Life Underground: the Biology of Subterranean Rodents. Chicago: University of Chicago Press.
- Li W Z , Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* , **22** ( 13 ) : 1658 – 1659.
- Lin G H. 2010. Morphological adaptations of digging apparatus and population genetics of the subterranean rodent *Myospalax cansus*. Ph. D dissertation of Graduate University of the Chinese Academy of Sciences , Beijing. ( in Chinese )
- Löytynoja A , Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* , **102** ( 30 ) : 10557 – 10562.
- Nevo E. 1999. Mosaic Evolution of Subterranean Mammals: Regression , Progression and Global Convergence. Oxford: Oxford University Press.
- Norris R W , Zhou K Y , Zhou C Q , Yang G , Kilpatrick C W , Honeycutt R L. 2004. The phylogenetic position of the zokors ( Myospalacinae ) and comments on the families of muroids ( Rodentia ) . *Molecular Phylogenetics and Evolution* , **31**: 972 – 978.
- O' Brien K P , Remm M , Sonnhammer E L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* , **33** ( Suppl 1 ) : D476 – D480.
- Parker J , Tsagkogeorga G , Cotton J A , Liu Y , Provero P , Stupka E , Rossiter S J. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* , **502** ( 7470 ) : 228 – 231.
- Pruitt K D , Harrow J , Harte R A , Wallin C , Diekhans M , Maglott D R , Searle S , Farrell C M , Loveland J E , Ruff B J , Hart E , Sunner M M , Landrum M J , Aken B , Ayling S , Baertsch R , Fernandez-Banet J , Cherry J L , Curwen V , DiCuccio M , Kellis M , Lee J , Lin M F , Schuster M , Shkeda A , Amid C , Brown G , Dukhanina O , Frankish A , Hart J , Makedak B L , Mudge J , Murphy M R , Murphy T , Rajan J , Rajput B , Riddick L D , Snow C , Steward C , Webb D , Weber J A , Wilming L , Wu W Y , Birney E , Haussler D , Hubbard T , Ostell J , Durbin R , Lipman D. 2009. The consensus coding sequence ( CCDS ) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* , **19** ( 7 ) : 1316 – 1323.
- Qi Y X , Liu Y B , Rong W H. 2011. RNA-Seq and its applications: a new technology for transcriptomics. *HEREDITAS* , **33** ( 11 ) : 1191 – 1202. ( in Chinese )
- Ranwez V , Harispe S , Delsuc F , Douzery E J. 2011. MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS One* , **6** ( 9 ) : e22594.
- Rebhan M , Chalifa-Caspi V , Prilusky J , Lancet D. 1997. GeneCards: integrating information about genes , proteins and diseases. *Trends in Genetics* , **13** ( 4 ) : 163.
- Rice P , Longden I , Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* , **16** ( 6 ) : 276 – 277.
- Rose K D. 2006. The Beginning of the Age of Mammals. Baltimore: Johns Hopkins University Press.
- Shams I , Malik A , Manov I , Joel A , Band M , Avivi A. 2013. Transcription pattern of p53 – targeted dna repair genes in the hypoxia-tolerant subterranean mole rat *Spalax*. *Journal of Molecular Biology* , **425**: 1111 – 1118.
- Sun S Z , Wei L , Wei D B , Wang D W , Ma B Y. 2013. Differences of glycolysis in skeletal muscle and lactate metabolism in liver between plateau zokor ( *Myospalax baileyi* ) and plateau pika ( *Ochotona curzoniae* ) . *Acta Physiologica Sinica* , **65** ( 3 ) : 276 – 284.
- Tian X , Azpurua J , Hine C , Vaidya A , Myakishev-Rempel M , Ablaeva J , Mao Z , Nevo E , Gorbunova V , Seluanov A. 2013. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*.
- Vleck D. 1979. The energy cost of burrowing by the pocket gopher *Thomomys bottae*. *Physiol Zool* , **52**: 122 – 135.
- Wang Z , Zhang Y. 2012. Predicted structural change in erythropoietin of plateau zokors – adaptation to high altitude. *Gene* , **501** ( 2 ) : 206 – 212.
- Zhang J , Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution* , **14** ( 5 ) : 527 – 536.
- Zhang J , Nei M. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony , likelihood , and distance methods. *Journal of Molecular Evolution* , **44** ( 1 ) : S139 – S146.
- Zhang Y , Zhang Z , Liu J. 2003. Burrowing rodents as ecosystem engineers: the ecology and management of plateau zokors *Myospalax fontanieri* in alpine meadow ecosystems on the Tibetan Plateau. *Mammal Review* , **33** ( 3 – 4 ) : 284 – 294.
- Zhu R J , Rao X F , Wei D B , Wang D W , Wei L , Sun S Z. 2012. Functional difference of malate-aspartate shuttle system in liver between plateau zokor ( *Myospalax baileyi* ) and plateau pika ( *Ochotona curzoniae* ) . *Acta Physiologica Sinica* , **64** ( 2 ) : 177 – 186. ( in Chinese )
- 祁云霞 , 刘永斌 , 荣威恒. 2011. 转录组研究新技术: RNA-Seq 及其应用. *遗传* , **33** ( 11 ) : 1191 – 1202.
- 朱瑞娟 , 饶鑫峰 , 魏登邦 , 王多伟 , 魏莲 , 孙生祯. 2012. 高原鼢鼠和高原鼠兔肝脏苹果酸天冬氨酸穿梭系统的功能差异. *生理学报* , **64** ( 2 ) : 177 – 186.
- 林恭华. 2010. 地下啮齿类——甘肃鼢鼠挖掘器官形态适应与种群遗传学分析. 中国科学院研究生院博士学位论文.