

高原鼯鼠血红蛋白基因鉴定、表达及其分子进化分析*

鲁松松, 赵夏下, 祁重霞, 杜雨薇

甘肃农业大学林学院, 甘肃 兰州 730070

摘要: 结合比较基因组、转录组、分子进化和蛋白质物理化学性质及结构等分析方法, 探讨高原鼯鼠 *Eospalax baileyi* 血红蛋白 (Hb, hemoglobin) 的 α 、 β 基因簇结构、基因表达模式及正选择位点对蛋白质结构的潜在影响等, 旨在研究高原鼯鼠 Hb 氧亲和力升高的机理。结果表明: 高原鼯鼠的 α 和 β 基因簇分别含有 3 个 α ($5'$ - α^E , α^A , α^Q -3') 和 7 个 β ($5'$ - ϵ , γ , $\gamma 1$, γ , σ (pesudo), $\beta 1$, $\beta 2$ (pesudo)-3') 基因; 成体高原鼯鼠除表达常规的成体 Hb 基因 α^A 和 $\beta 1$ 外, 还表达了胚胎 α^E 基因; Hb $\beta 1$ 亚基血红素“口袋”开口处 12 个氨基酸残基的突变, 使得该区域稳定性和亲水性增强, 口袋 (活性中心) 体积也稍有增加。高原鼯鼠较高 Hb 氧亲和力的主要原因可能是其胚胎型 α^E 基因在成体中的表达, 同时 α^E 和 $\beta 1$ 亚基稳定性增加, 有利于高氧亲和力的 α^E 型 Hb ($\alpha^E\beta 1$)₂ 积累到较高的浓度。本文首次发现 α^E 基因在成体哺乳动物中被高表达, 这可能是极端低氧环境诱导的, 也可能是 α^E 基因调控区域遗传基础的变化引起的, 具体机制尚待进一步验证。

关键词: 高原鼯鼠 *Eospalax baileyi*; 血红蛋白; 氧亲和力; 基因表达

中图分类号: Q71 **文献标志码:** A **文章编号:** 2097-0137 (2022) 04-0049-11

Hemoglobin gene identification, expression and molecular evolution analysis in *Eospalax baileyi*

LU Songsong, ZHAO Xiaxia, QI Chongxia, DU Yuwei

College of Forestry, Gansu Agricultural University, Lanzhou 730070, China

Abstract: The study was aimed to reveal the mechanism of elevated hemoglobin oxygen affinity of Plateau zokor (*Eospalax baileyi*). By using a comparative genomic and transcriptomic analysis of the α and β gene clusters in the plateau zokor to characterize the changes in the size and membership composition of the α and β gene families within the Spalacidae, the study evaluated the evolutionary changes in the developmental regulation of gene expression of the plateau zokor and explored the potential impact of positive selection sites on protein structure and function. The results showed that the plateau zokor α and β gene clusters contain three α ($5'$ - α^E , α^A , α^Q -3') and seven β ($5'$ - ϵ , γ , $\gamma 1$, $\gamma 2$, σ (pesudo), $\beta 1$, $\beta 2$ (pesudo)-3') genes respectively; in addition to the conventional adult expression genes α^A and $\beta 1$, adult plateau zokor expressed an extra embryonic α^E gene; the twelve amino acid mutations on heme “pocket” opening of $\beta 1$ subunit might be able to increase the stability and hydrophilicity of the region, and slightly increases the volume of the pocket (active center). The results also suggested that the abnormally high oxygen affinity of plateau zokor might result from the expression of embryonic α^E gene in adults; the increased stability of α^E and $\beta 1$ subunits and the decrease of α^A subunit stability may lead to accumulation of α^E type Hb ($\alpha^E\beta 1$)₂ to a concentration similar to α^A type ($\alpha^A\beta 1$)₂. It was the

* 收稿日期: 2021-05-18

录用日期: 2021-08-23

网络首发日期: 2021-12-13

基金项目: 甘肃农业大学科技创新基金(2017RCZX-23, GAU-XKJS-2018-101);

甘肃省青年科技基金计划(18JR3RA184)

作者简介: 鲁松松 (1986年生), 男; 研究方向: 野生动物生理生态及分子生态; E-mail: luss13@lzu.edu.cn

first report that the α^E gene is highly expressed in adult mammals, this phenomenon may be induced by extreme low oxygen environment, or it may be caused by changes in the genetic basis of the α^E gene regulatory region. The specific mechanism needs further study.

Key words: *Eospalax baileyi*; hemoglobin; oxygen affinity; gene expression

高原土著动物的血红蛋白(Hb, hemoglobin)普遍进化出较高的氧亲和力, 属于生理适应的经典例子^[1]。大量的比较研究结果显示, 高海拔啮齿类动物通常具有较低海拔近缘物种更高的血红蛋白或血液氧亲和力^[2-4]。低氧条件下, Hb氧亲和力的升高使得动、静脉氧分压差值最大化, 从而有效提高机体的氧运输效率^[5]。有颌类脊椎动物的Hb是由2个相同 α 亚基和2个相同 β 亚基组成的异源四聚体, 每个亚基包含一个血红素(Heme)辅基, 亚铁离子(Fe^{2+})位于卟啉环中心位置, 可逆地与1分子氧气结合以发挥Hb的运输功能。四聚体结构赋予Hb与配体结合的协同效应和变构效应。Hb氧亲和力的进化改变主要涉及 α 和 β 亚基的氨基酸突变引起的协同效应(固有氧亲和力)及变构效应(变构因子的敏感度)的改变。

高原鼯鼠 *Eospalax baileyi* 是青藏高原特有的地下鼠, 隶属鼯鼠科 Spalacidae 鼯鼠属 *Myospalax* 凸颅亚属 *Eospalax*。其生存不仅要面临高原和密闭洞穴双重低氧的挑战, 同时还受到洞穴高二氧化碳环境的影响。据报道, 高原鼯鼠栖息地大气氧分压约为海平面水平的 65%, 其居住洞穴内(地下 18 cm)氧含量约为大气的 83%~88%, 且 CO_2 含量最高值可达大气含量的 48.6 倍^[6]。已有的研究表明, 高原鼯鼠长期生活在如此严苛的环境中, 其形态特征、心肺器官、血液循环系统以及分子水平都表现出适应低氧-高二氧化碳环境的特征, 如心脏每搏输出量、肺部单位面积肺泡数、红细胞数和 Hb 含量、动静脉氧分压差等氧气运输的各环节均呈现出协同的改变^[7-9]。最新的研究结果表明, 高原鼯鼠纯化 Hb 的固有氧亲和力显著高于小鼠, 且加入 DPG 和 Cl^- 等变构因子后高原鼯鼠 Hb 的 P_{50} 值(Hb 50% 氧饱和度下的氧分压)依然显著低于小鼠^[10], 说明高原鼯鼠升高的 Hb 氧亲和力可能是遗传基础改变引起的, 是自然选择的结果。值得注意的是, 在相同条件下, 高原鼯鼠 Hb 的固有氧亲和力 [P_{50} 仅为 (706.23 \pm 7.98) Pa], 显著低于其他啮齿类近缘种, 如小鼠 [(960.26 \pm 3.99) Pa]、家鼠(约为 984.20 Pa)、高原鼠类鹿鼠 *Peromyscus maniculatus* (1 010.80~1 170.40 Pa) 及同样营严格

地下生活的 6 种非洲鼯形鼠 (982.87~1 401.82 Pa)^[11-13]。高原鼯鼠成体 Hb 的 P_{50} 值与哺乳动物胚胎型 Hb (532.00~798.00 Pa) 更接近^[14]。此外等电聚焦电泳结果显示, 高原鼯鼠成体红细胞中表达了 2 种主要 Hb 组分, 一种含量稍高的组分具有较低的等电点 (pI, isoelectric point), 另一种含量稍低的组分具有较高的 pI^[10], 这与其他哺乳动物完全不同 (成体红细胞中仅表达 1 种主要 Hb 亚型)^[1]。为维持机体在不同环境中的氧供需平衡, 理论上机体可以通过对具有不同氧合特性的 Hb 亚型 (氧亲和力不同) 的表达调节, 从而可逆地调节血氧亲和力, 但这种调节机制较常见于鸟类, 在哺乳动物中尚未发现^[1, 15-16]。因此, 鉴定两种 Hb 亚型的组成是目前深入理解和阐释高原鼯鼠 Hb 适应低氧-高二氧化碳环境机理的关键。

随着测序技术的发展和运用, 基因组和转录组学研究在非模式动物中广泛应用, 近期高原鼯鼠基因组和多个组织转录组数据使用权限的开放, 极大地便利了对高原鼯鼠基因组 *Hb* 基因的鉴定和表达分析。本文通过比较基因组和转录组学分析了高原鼯鼠基因组中 *Hb* 基因组成及在成体各组织中的表达情况, 初步鉴定了两种成体 Hb 亚型的组成, 并通过分子进化及蛋白质理化性质和结构分析深入探讨了高原鼯鼠 Hb 进化出较高氧亲和力的潜在机制。本研究首次发现哺乳类动物胚胎型 α^E 基因在成体中表达, 这可能是成体高原鼯鼠 Hb 具有异常高氧亲和力的主要原因, 为哺乳动物低氧适应的研究和高原医学的发展提供理论依据和新的思路。

1 材料与方法

1.1 血红蛋白基因鉴定和基因簇结构分析

从数据库中下载小鼠 (NCBI, GRCm39)、银星竹鼠 (NCBI, RhiPru_1.0)、以色列鼯鼠 (NCBI, S. galili_v1.0) 和高原鼯鼠 (NGDC, plateau_zokor_v2.1) 基因组序列及其注解文件 (GFF)。基因注解使用 BLAST 软件 (2.11.0 版)^[17], 比对到的基因组片段通过 GenScan 在线服务 (<https://www.genes.mit.edu/GENSCAN.html>) 进行基因、编码区

序列(CDS, coding sequence)和蛋白质预测。通过Mega软件将对齐后的序列进行手动矫正。本文将CDS序列内提前出现终止密码子和碱基缺失/插入导致的三联密码子错位认定为假基因。对NCBI数据库中存放的基因组、基因簇DNA序列全长使用Nucleotide库根据各基因在基因组片段的位置获取。高原鼯鼠*Hb*基因及基因簇DNA序列全长使用自编Perl语言脚本进行提取。基因簇结构可视化使用GSDS2.0在线服务完成(<http://gsds.cbi.pku.edu.cn/>)。

1.2 转录物组数据和转录本从头组装

本文选用NCBI数据库中高原鼯鼠所有组织样品的转录物组测序数据以评估高原鼯鼠*Hb*基因的表达,包括肾脏、骨骼肌和视网膜组织(PRJNA282349, $N=1$),以及脑(PRJNA342079, $N=3$)和肝脏组织(PRJNA211727, $N=3$)。选用小鼠胚胎及成体组织的测序数据,以验证啮齿类基因组中胚胎型和成体型*Hb*基因的组成,包括3.5 d(PRJNA674788, $N=3$)和5.5 d(PRJNA683789, $N=3$)单囊胚,13~15 d肝脏(PRJNA625751, $N=3$)、心脏(PRJNA602232, $N=3$)和脑(PRJNA670344, $N=3$)组织,成体心脏(PRJNA264588, $N=3$)、肝脏(PRJNA668230, $N=3$)、脑(PRJNA674888, $N=3$)、骨骼肌(PRJDB10551, $N=3$)、肾脏(PRJNA543476, $N=3$)及视网膜(PRJNA435475, $N=3$)。转录物组测序基础数据使用自编Perl语言脚本进行过滤(删除低质量值的双末端测序reads)。使用Trinity软件(2.11.0版)^[18]对过滤后的数据分别进行从头组装,得到转录本后,使用上述BLAST法鉴定*Hb*基因的转录本。从头组装的转录本用于进一步确定基因组注解中预测的假基因及CDS序列的准确性。

1.3 血红蛋白基因系统发生关系重建和分子进化分析

本文分别采用最大似然(maximum likelihood)和贝叶斯(Bayesian)法对*Hb*基因进行系统发生重建。DNA全长序列用MUSCLE软件^[19]对齐后,使用Treefinder软件(<http://www.treefinder.de/>)和MrBayes软件(3.2.7a)^[20]分别构建最大似然树和贝叶斯树。其中,最大似然树使用Propose Model插件选择最优替代模型后,采用bootstrap抽样检验法(1000次)评估分枝节点的支持率。贝叶斯树构建参数设置为广义时间可逆(GTR, general time reversible)模型,8条链同时运行1000000代,采用频率为1000。舍弃前25%的老化样本(burnin)

后,采用多数决定原则统计共有树及其分枝节点的贝叶斯后验概率(Bayesian posterior probabilities)。

本文以转录本中鉴定的CDS序列为基础,使用PAML4软件(4.9j)^[21]的分枝位点模型(branch site)分析了高原鼯鼠和甘肃鼯鼠成体表达*Hb*基因的潜在正选择位点。选用MA模型并指定鼯鼠分枝为前景分枝以检测该分枝中可能的正选择位点,并使用经验贝叶斯法(BEB, bayes empirical bayes)计算正选择位点的后验概率,比较模型选择空模型(null model)。上述模型间差异显著性采用似然比检验(LRT, likelihood ratio test)进行测试。

1.4 血红蛋白基因表达分析

过滤后的转录物组数据使用Bowtie2软件(2.2.9版)^[22]比对到各物种*Hb*基因的CDS序列上,比对参数设置为-D 20 -R 3 -N 0 -L 20 -i S, 1, 0.50。比对结果使用自编Perl语言脚本进行reads数量统计,并计算各组织中*Hb*基因的表达量。基因表达量表示为每百万reads比对到CDS序列每千碱基上的数量(RPKM, reads per kilobase per million mapped reads),RPKM<1时为不表达基因。单样品测序数据仅用于评估基因是否表达。对有生物学重复样品的测序数据($N=3$),使用SPSS软件单因素方差分析(one-way ANOVA)统计差异的显著性($P<0.05$ 时差异显著)。

1.5 蛋白质物理化学性质和结构分析

使用ExPASy在线服务(www.expasy.org)完成蛋白质pI、总平均亲水指数(GRAVY, grand average of hydropathicity)和氨基酸亲水性/极性等物理化学性质分析。血红蛋白四聚体同源建模使用Modeller软件(9.24版)^[23]完成,同源模板选择小鼠去氧合成体血红蛋白X射线晶体衍射结构(3HRW, 0.28 nm),共构建100个模型,使用DOPE和GA341方法进行评估并选择最优模型用于结构比较。同源结构比较和3D结构显示使用VMD软件(1.9.3版)^[24]。血红素“口袋”(heme pocket)体积计算使用GHECOM 1.0在线服务(<https://pdj.org/ghecom/>)完成。

2 结果

2.1 鼯鼠科动物 α 和 β 基因簇结构

本文对高原鼯鼠、银星竹鼠和以色列鼯鼠3个鼯鼠科物种基因组中的*Hb*基因进行了重新鉴定,与已有研究结果相符^[25],3种鼠类*Hb*基因中,

α 和 β 两个亚家族基因分别位于两个不同基因组片段且家族成员间相互聚集成簇(图 1)。比较分析结果显示, 3 种鼠类 α 基因簇长度不尽相同, 其中高原鼯鼠(GWHABJZ00000029: 10 108 817~10 119 436 bp)和银星竹鼠(VZQC01006430: 2 529 015~2 541 954 bp)长度相近, 约为 11 kb; 而以色列鼯鼠(NW_008353596: 4 698 608~4 724 458 bp)长度与小鼠和人相近, 约为 26 kb。与小鼠相比(5'- $\alpha^E, \alpha^{A1}, \alpha^{O2}, \alpha^{A2}, \alpha^{O1}$ -3'), 3 种鼯鼠科物种的 α 基

因家族明显收敛, 基因簇中仅保留了 3 个基因, 排列模式为 5'- $\alpha^E, \alpha^A, \alpha^O$ -3'。3 种鼠类的 β 基因簇长度差异较大, 分别为高原鼯鼠 53 kb(165 907~219 796 bp)、银星竹鼠 110 kb(196 599~306 296 bp)和以色列鼯鼠 86 kb(752 279~838 350 bp), 但基因数量相对保守, 均包含 7 个家族成员。与啮齿目其他鼠类的研究相似^[26], 3 种鼯鼠科鼠类 β 基因簇 5' 端为单拷贝的胚胎型 ϵ 基因, 3' 端为至少 2 个胚后表达型 β 基因。

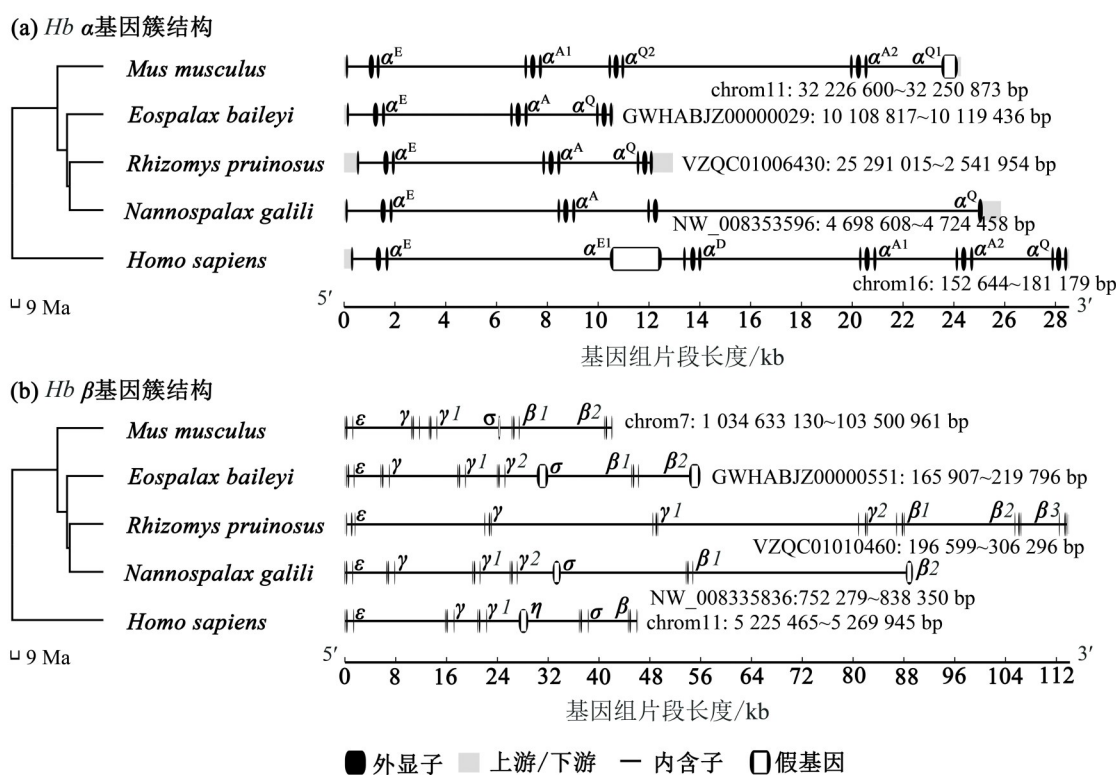


图 1 啮齿类动物 α 和 β 基因簇结构

Fig. 1 Genomic structure of the α - and β -globin gene family in rodents

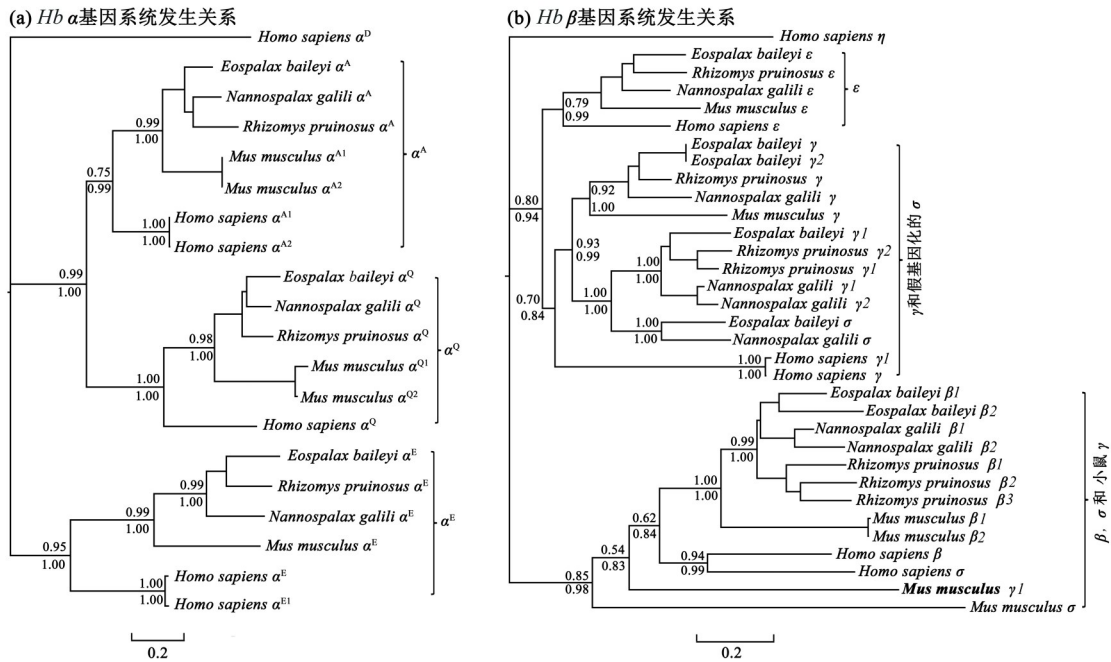
2.2 血红蛋白基因的系统发生关系

基于啮齿类动物 α (图 2a)和 β 基因(图 2b)DNA 序列全长重建系统发生关系, 结果显示最大似然树和贝叶斯树具有相似的拓扑结构。在 α 基因树中, 鼯鼠科鼠类 3 个单拷贝 α 基因分别与小鼠和人类 α^A (MLbs = 0.75, Bapp = 0.99)、 α^O (MLbs = 1.00, Bapp = 1.00) 及 α^E (MLbs = 0.95, Bapp = 1.00) 形成主要分枝, 其中 α^A 和 α^O 两个主要分枝形成姊妹分枝组成一个根分枝 (MLbs = 0.99, Bapp = 1.00)。本文在所有啮齿目物种中均未发现与人类 α^D 基因直系同源的基因。上述结果说明, 鼯鼠科鼠类 3 个 α 基因为 1:1 直系同源基因, 且啮齿类动物基因组中 α^D 基因已被删除。

在 β 基因树中, 鼯鼠科鼠类 β 基因形成 4 个主要分枝: 与小鼠和人类 ϵ 基因 1:1 同源的分枝 (ϵ 分枝, MLbs = 0.79, Bapp = 0.99); 与小鼠 γ 基因同源的分枝 (包括了鼯鼠科鼠类 γ 基因及高原鼯鼠 γ^2 基因, MLbs = 0.92, Bapp = 1.00); 鼯鼠科鼠类 γ^1 、银星竹鼠和以色列鼯鼠 γ^2 基因、高原鼯鼠和以色列鼯鼠两个 σ 假基因 (MLbs = 1.00, Bapp = 1.00); 鼯鼠科鼠类 β^1 、 β^2 和 β^3 基因 (MLbs = 0.99, Bapp = 1.00)。已有的研究表明, 现存啮齿类动物 ϵ/γ 和 σ/β 基因分别起源于真兽亚纲动物共同祖先中原 ϵ (proto ϵ) 和原 β (proto β) 基因的一次基因复制事件^[27]。与上述结果相符, ϵ 分枝与除小鼠 γ^1 基因外的所有 γ 基因聚为一个根分枝 (MLbs = 0.80, Bapp =

0.94), 同时所有 β 基因与除高原鼯鼠和以色列鼯鼠两个 σ 假基因外的所有 σ 基因聚为一支(MLbs = 0.85, Bapp = 0.98)。值得注意的是, 高原鼯鼠和以色列鼯鼠两个 σ 假基因和小鼠 γI 远离了 σ 或 γ 分枝。出现这个结果的原因可能是高原鼯鼠和以色列鼯鼠

γ 基因(供体)与邻近的 σ 基因(受体)及小鼠 σ 基因(供体)与邻近的 γ 基因(受体)间发生过基因转换事件(染色体不平等交换), 使得受体基因在系统发生树中远离原分枝而更接近供体分枝, 这种情况在脊椎动物 β 基因簇中较为常见^[28-29]。



分枝节点上方数字为bootstrap值, 分枝节点下方数字为贝叶斯后验概率。

图2 啮齿类动物 α 和 β 基因系统发生关系

Fig. 2 Maximum likelihood phylogram describing relationships among the α - and β -globin genes of rodents

2.3 小鼠和高原鼯鼠Hb基因的表达

小鼠胚胎和成体组织Hb基因表达结果(图3、图4)表明, 胚胎和成体阶段Hb基因具有不同的表达谱, 即使在胚胎发育的不同时期Hb基因的表达谱也明显不同。在胚胎发育早期(3.5~5.5 d), 仅检测到 α^{A1} 和 α^{A2} 两个基因被表达, RPKM分别为 3.96 ± 2.68 和 2.11 ± 0.68 。在胚胎发育后期(13~15 d), 除 γI 和两个假基因(σ 和 α^{O1})外其他所有基因均检测到被表达, RPKM范围为(1.12 \pm 0.39)~(38 254.00 \pm 3 295.14)。其中 α^E 、 ϵ 和 γ 基因在胚胎脑组织(14 d)中的表达量最高, RPKM分别为 213.98 ± 30.93 、 $1 111.30 \pm 174.57$ 和 31.29 ± 8.99 ; α^{A1} 、 α^{O2} 、 α^{A2} 、 $\beta 1$ 和 $\beta 2$ 基因在胚胎肝脏组织(15 d)中的表达量最高, RPKM分别为 $38 254.00 \pm 3 295.14$ 、 30.15 ± 2.96 、 $26 120.00 \pm 2 278.35$ 、 $30 849.00 \pm 3 071.05$ 和 $24 313 \pm 2 645.87$ 。在成体阶段, α^{A1} 、 α^{A2} 、 $\beta 1$ 和 $\beta 2$ 在心脏、骨骼肌、脑、肝脏、肾脏、视网膜和睾丸中均被表达, RPKM范

围为(4.60 \pm 1.68)~(389.03 \pm 190.56), 其他基因未检测到表达(RPKM < 1)。

在成体高原鼯鼠肝脏和脑组织转录本中, 本研究仅鉴定到 α^A 、 α^E 和 $\beta 1$ 基因的表达及完整的CDS序列, 其他基因RPKM均小于1且CDS序列不完整。如图5所示, 高原鼯鼠肝脏和脑组织中 α^A (RPKM分别为 $2 079.70 \pm 593.66$ 和 $1 655.90 \pm 475.47$)和 $\beta 1$ (RPKM分别为 869.12 ± 147.64 和 $1 361.5 \pm 387.91$)基因的表达量均显著高于小鼠同源基因的表达量(α^{A1} 基因RPKM分别为 157.04 ± 13.45 和 210.83 ± 37.56 ; α^{A2} 基因RPKM分别为 137.15 ± 9.55 和 200.49 ± 34.43 ; $\beta 1$ 基因RPKM分别为 188.88 ± 14.74 和 53.58 ± 8.64 ; $\beta 2$ 基因RPKM分别为 93.29 ± 7.27 和 44.73 ± 8.90)。值得注意的是, 在成体高原鼯鼠脑组织中 α^E 基因被检测到表达(RPKM为 2.58 ± 0.53), 小鼠中该基因仅表达于胚胎期(14和15 d)的肝脏和脑组织。

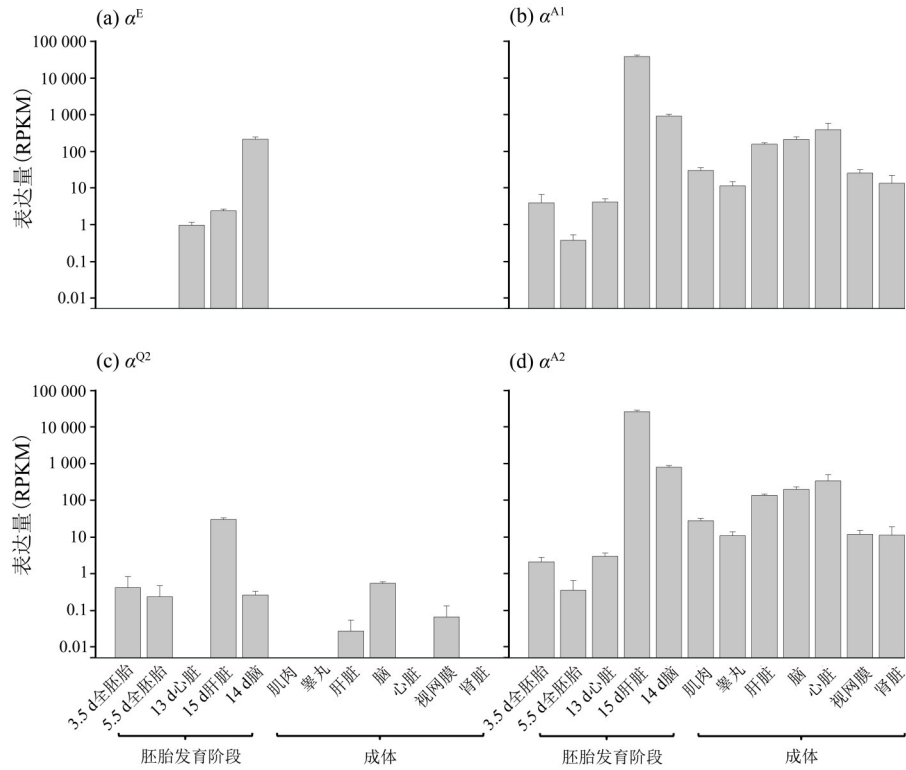


图 3 α 基因在小鼠胚胎和成体多个组织中的表达水平

Fig. 3 Expression level of α -globin genes in multiple tissues of mice embryo and adult stages

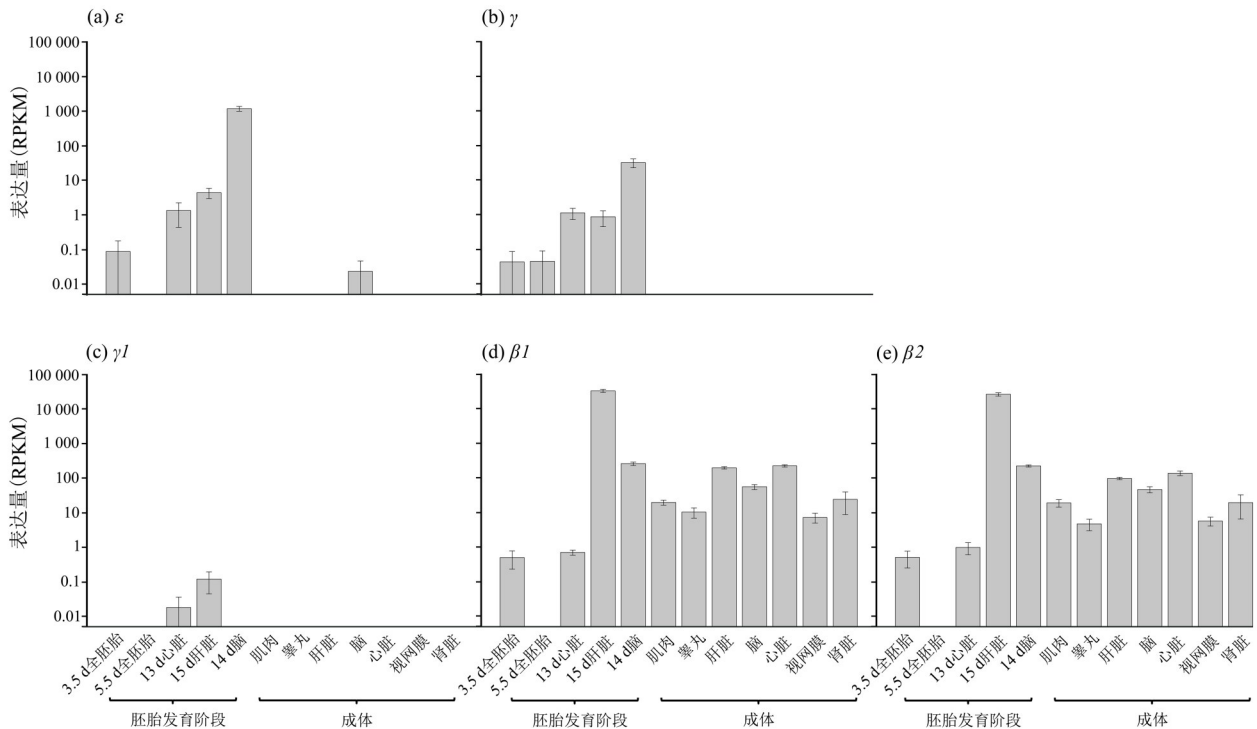


图 4 β 基因在小鼠胚胎和成体多个组织中的表达水平

Fig. 4 Expression level of β -globin genes in multiple tissues of mice embryo and adult stages

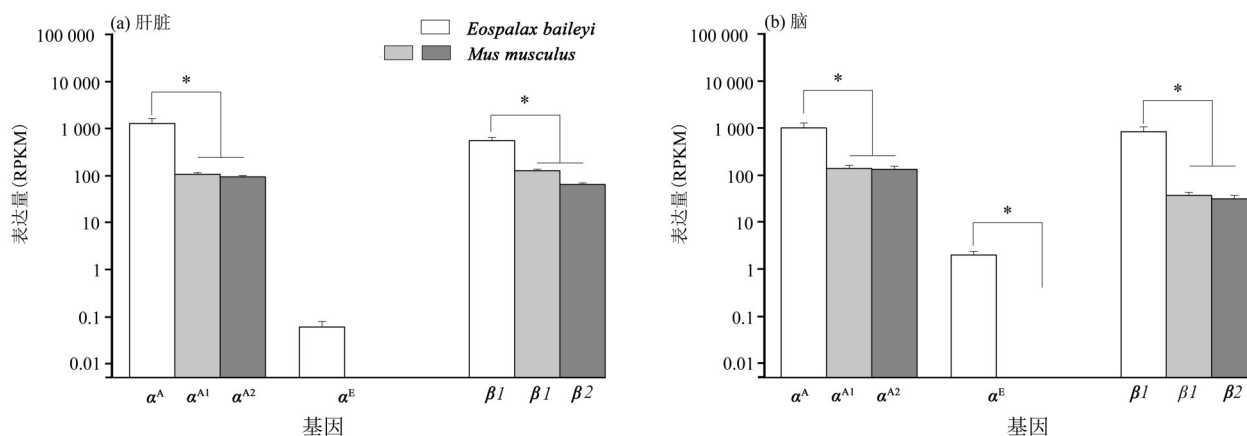


图5 成体小鼠和高原鼯鼠肝脏和脑中血红蛋白基因表达水平 (* $P < 0.05$)

Fig. 5 Expression level of Hb genes in liver and brain of adult mice and plateau zokor (* $P < 0.05$)

2.4 高原鼯鼠成体表达基因的正选择位点

Pu等^[10]的研究结果表明, 高原鼯鼠血红蛋白 α 和 β 亚基分别具有17和24个氨基酸差异, 其中 α 111、 α 131、 β 4、 β 5和 β 115等5个氨基酸突变可能是高原鼯鼠Hb固有氧亲和力升高及与变构因子和 CO_2 敏感度改变的主要因素。本文采用分枝位点模型以高原鼯鼠各基因为前景分枝 (foreground), 分析了上述突变是否是正自然选择的结果 (表1)。 α^A 、 α^E 和 β 1基因 Model A0 的 $\ln L$ 值分别为-1 165.27、-1 261.79和-1 317.58, Model A1 的 $\ln L$ 值分别为-1 161.67、-1 261.79和-1 312.34。LRT 检验结果显示 α^A 和 β 1基因拒绝 Model A0 模型 ($P < 0.01$), 说明高原鼯鼠这两个基因的某些位点可能受到正选择。BEB 分析结果显示, α^A 和 β 1基因分别有1个(24Leu)和3个(43Ser, 61Ser和87His)位点可能具有大于1的 ω 值, 后检验概率分别为0.933、0.639、0.939和0.711。

2.5 高原鼯鼠血红蛋白的物理化学性质和结构

已有研究表明, 成体高原鼯鼠至少表达了2种Hb亚型, 其中低等电点的亚型相对含量稍高于另外一种高等电点亚型; 成体小鼠只表达一种Hb亚型, 等电点与高原鼯鼠含量高的亚型相近^[10]。基于基因表达分析结果, 本文推测成体高原鼯鼠的两种Hb亚型为成体 α^A 型($\alpha^A\beta$ 1)₂和胚胎 α^E 型($\alpha^E\beta$ 1)₂。蛋白质理化性质分析结果 (表2) 进一步证实了该推测, α^A 型Hb是所有已知脊椎动物成体表达的主要亚型, 这种模式在成体高原鼯鼠中依然保守, α^A 型与 α^E 亚型相比具有较低的pI值 (分别为7.17和7.92)。小鼠成体表达基因预测蛋白质序列结果显示, 2个 α (α^{A1} 和 α^{A2})和2个 β (β 1和 β 2)基因最终翻译为完全相同的氨基酸序列, 故非变性电泳条带

只有1条(1个主要Hb组分), 其 α^A 和 β 多肽pI分别为7.97和7.13, 这与高原鼯鼠 α^A 型Hb两个亚基的pI更为接近(7.17和7.95)。此外, 高原鼯鼠所有Hb亚基的GRAVY数值均较低, 说明高原鼯鼠2种Hb亚型的亲水性均较小鼠高。

蛋白质结构比较分析结果 (图6) 显示, 高原鼯鼠成体Hb β 亚基的24个氨基酸突变中有12个(41、43、58、61、68、69、70、72、76、80、86和87位)位于血红素“口袋”的开口处, 3个潜在正选择位点刚好也在其中。分子动力学计算结果未检测到明显影响其T态到R态转变的因素(氢键或盐键形成或丢失), 且这些位点也不是典型的变构因子或 CO_2 的结合位点^[10]。进一步分析结果显示 (图6~图7), 41Tyr-Phe和43Asp-Ser两个氨基酸突变使得该区域亲水性和极性减弱, 该区域由无规则卷曲变为更稳定的 α 螺旋; 70Ala-Ser和86Ala-Ser由非极性氨基酸突变为极性氨基酸, 增强了该区域的亲水性和极性, 这可能是高原鼯鼠 β 亚基具有较小鼠高的亲水性的主要原因, 同时结合87Ser-His突变, 使得86~87位氨基酸区域由无规则卷曲变为 α 螺旋; 与43和86~87位氨基酸突变相反, 80Ser-Asp的反向突变使得原 α 螺旋变为无规则卷曲。上述突变造成的区域构象改变, 尤其是80位氨基酸区域 α 螺旋的消失可能引起血红素“口袋”的夹角增大, 最终使得血红素口袋的体积增大(由小鼠的1.1320 nm³增加到高原鼯鼠的1.1894 nm³), 有利于提高 O_2 的运输效率。

3 讨论

系统发生关系结果表明, 啮齿目动物共同祖先中 α 和 β 基因簇排列模式可能为5'- α^E , α^A , α^O -3'和

表 1 高原鼯鼠成体表达血红蛋白基因的正选择分析

Table 1 Maximum likelihood analysis of adult Hb genes from rodents under branch-site models

分支位点模型	参数估计值	自由度	ln L	测试 (A1 vs A0)	
				χ^2	P
α^A	位点类型: 0/1/2a/2b 比例: 0.80/0.20/0.00/0.00			7.20	0.007
Model A0 ($\omega_2=1$)	背景 ω : 0.03/1.00/0.03/1.00 前景 ω : 0.03/1.00/1.00/1.00 比例: 0.81/0.18/0.01/0.00	15	-1 165.27		
Model A1 ($\omega_2>1$)	背景 ω : 0.03/1.00/0.03/1.00 前景 ω : 0.03/1/219.37/219.37	16	-1 161.67		
α^E	位点类型: 0/1/2a/2b 比例: 0.85/0.12/0.03/0.00			0.00	1.00
Model A0 ($\omega_2=1$)	背景 ω : 0.09/1.00/0.09/1.00 前景 ω : 0.09/1.00/1.00/1.00 比例: 0.85/0.12/0.03/0.00	11	-1 261.79		
Model A1 ($\omega_2>1$)	背景 ω : 0.09/1.00/0.09/1.00 前景 ω : 0.09/1.00/1.00/1.00	12	-1 261.79		
$\beta 1$	位点类型: 0/1/2a/2b 比例: 0.65/0.35/0.00/0.00			10.49	0.001
Model A0 ($\omega_2=1$)	背景 ω : 0.05/1.00/0.05/1.00 前景 ω : 0.05/1.00/1.00/1.00 比例: 0.65/0.33/0.01/0.01	13	-1 317.58		
Model A1 ($\omega_2>1$)	背景 ω : 0.05/1.00/0.05/1.00 前景 ω : 0.05/1.00/998.99/998.99	14	-1 312.34		

表 2 小鼠和高原鼯鼠血红蛋白 α 和 β 多肽物理化学性质分析¹⁾

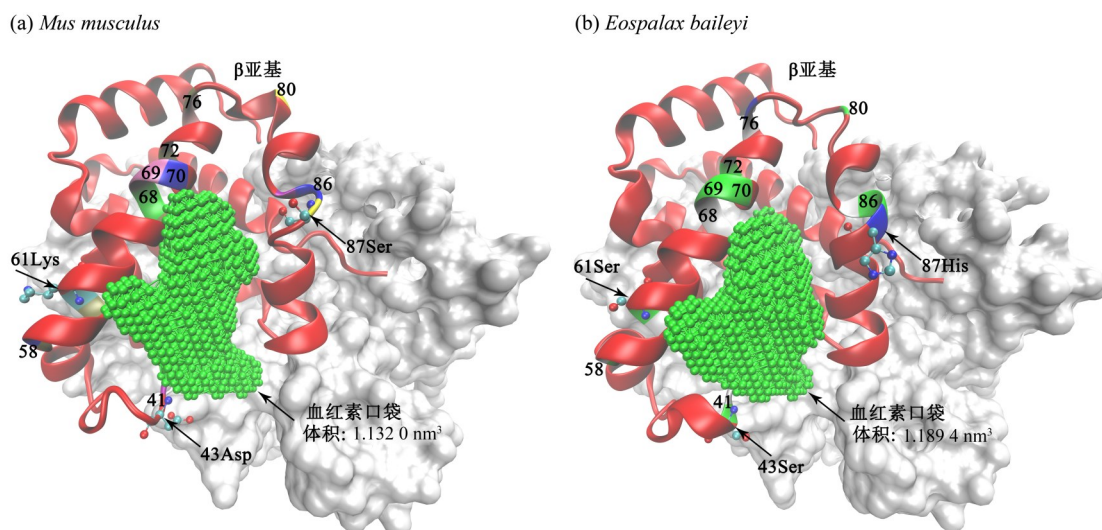
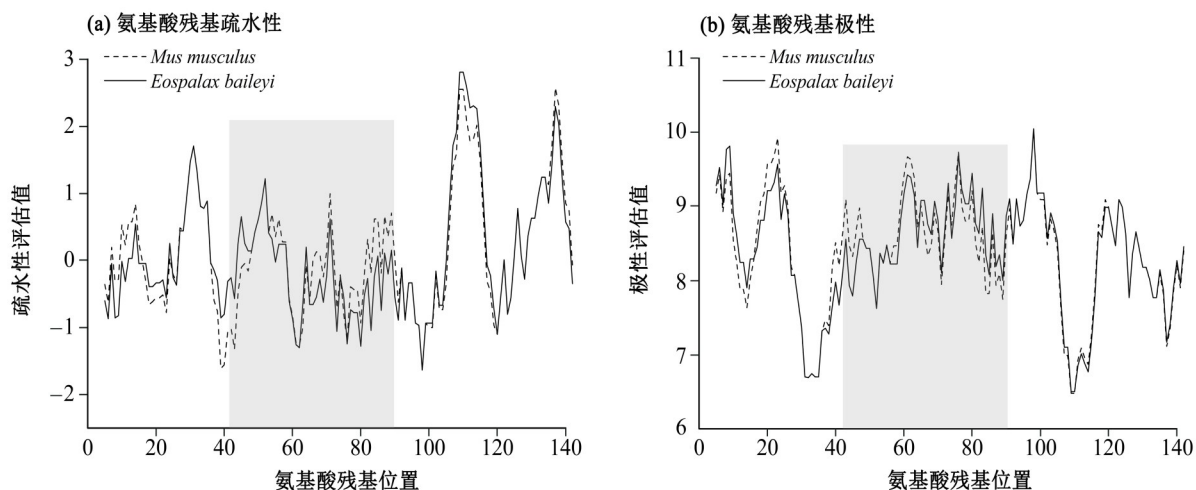
Table 2 Predicted physicochemical properties of hemoglobin α and β polypeptides from mice and plateau zokor

多肽亚基	NAA	MW	pI	GRAVY	不稳定指数
<i>Eospalax baileyi</i> α^A	142	15 374.50	7.17	-0.129	16.4
<i>Mus musculus</i> α^A	142	15 112.20	7.97	-0.054	7.86
<i>Eospalax baileyi</i> α^E	142	16 094.58	7.92	-0.071	33.95
<i>Mus musculus</i> α^E	142	16 234.90	7.06	-0.054	38.05
<i>Eospalax baileyi</i> $\beta 1$	147	15 837.19	7.95	0.056	0.97
<i>Mus musculus</i> β	147	15 748.07	7.13	0.092	7.41

1) NAA: 氨基酸数量; MW: 相对分子质量; GRAVY: 总平均亲水指数。

5'- ϵ , γ , σ (pesudo), β -3', 在现存不同谱系中 α 和 β 基因亚家族经历了较为复杂的基因复制、删除、同源重组和假基因事件^[26,30]。本文研究结果显示, 鼯鼠科物种 α 基因簇高度保守, 均保留了 3 个单拷贝 α^E , α^A 和 α^Q 基因, 但 β 基因簇在 3 个物种中分

别经历了不同的基因复制、删除、同源重组和假基因化事件。基于系统发生关系重建结构, 本文推测鼯鼠科共同祖先中 α 和 β 基因簇中基因排列模式分别为 5'- α^E , α^A , α^Q -3' 和 5'- ϵ , γ , $\gamma 1$, σ (pesudo + $\gamma 1$ 重组), β -3', 其中原 γ 基因经过复制后产生 γ

图6 小鼠和高原鼯鼠 α 型血红蛋白($\alpha^{\beta}1$)₂的三维结构模型Fig. 6 Models of the three-dimensional structure of ($\alpha^{\beta}1$)₂ of mice and plateau zokor图7 小鼠和高原鼯鼠血红蛋白 β 亚基氨基酸残基的疏水性和极性Fig. 7 The hydrophobicity and polarity of amino acid residues in hemoglobin β polypeptides of mice and plateau zokor

和 $\gamma 1$ 、 $\gamma 1$ 基因(供体)与 σ 假基因(受体)发生过同源重组。尽管3种鼯鼠科物种具有相似的 β 基因簇排列模式,但它们的 $\gamma 2$ 基因起源不同,高原鼯鼠起源于共同祖先中 γ 基因的复制,另外两个物种则起源于 $\gamma 1$ 基因的复制。高原鼯鼠、银星竹鼠和以色列鼯鼠 β 基因簇3'端的2~3个 β 基因均为物种特异性基因复制事件的产物,其中高原鼯鼠和银星竹鼠 $\beta 2$ 基因及以色列鼯鼠 $\beta 3$ 基因产生后不久就经历了假基因化事件。与小鼠和人类相比,3个鼯鼠科物种 α 和 β 基因簇仍然保留了5'端胚胎特异性表达基因和3'端成体特异性表达基因的排列模式^[25-26,30]。

在成体高原鼯鼠转录本中,本文检测到 α^{E} 、 α^{F} 和 $\beta 1$ 三个基因的表达。理论上,这3个基因可以

翻译并最终合成成体 α^{A} 型($\alpha^{\text{A}}\beta 1$)₂和胚胎 α^{E} 型($\alpha^{\text{E}}\beta 1$)₂两种Hb亚型。与等电聚焦电泳结果相符^[10],本研究结果显示预测的 α^{E} 亚基具有较高的pI值。已有的研究表明,所有哺乳动物的 α^{E} 基因仅在胚胎发育时期被表达,截至目前尚无证据表明哺乳动物成体阶段表达该基因^[11,25]。小鼠胚胎和成体时期转录物组数据分析进一步验证和细化了这种模式,即 α^{E} 基因仅在胚胎发育后期(13~15 d)被表达,而 α^{A} 基因则在胚胎和成体所有组织中被表达。高原鼯鼠是首个被检测到在成体表达胚胎 α^{E} 型Hb的哺乳动物, α^{E} 型Hb具有较 α^{A} 型更高的氧亲和力,两种氧亲和力不同的Hb形成对血氧亲和力调节的一种级联机制,因此理论上成体高原鼯鼠血氧亲和力的调节范围可达两种亚

型 Hb 固有氧亲和力的极限值^[1]。上述结果可部分解释高原鼯鼠成体表达的两种 Hb 亚型及其具有异常高的固有氧亲和力的原因。根据基因表达分析结果,成体高原鼯鼠脑组织中 α^E 基因的表达式 (RPKM 仅为 2.58 ± 0.53) 远远低于 α^A 基因的表达式 $1\,655.90 \pm 475.47$, 这与电泳结果中蛋白质相对含量的结果差别较大。出现这种结果的原因可能有 2 个: ① α^E 基因可能在其他未被测序的组织中被高表达, 该推测尚需实验进一步验证; ② α^E 基因编码多肽的稳定性可能较高, 使得其可以在红细胞内积累到较高的浓度。蛋白质物理化学性质分析结果显示, 与小鼠相比, 高原鼯鼠 α^A 多肽的稳定性明显降低, 而 α^E 和 β 基因编码多肽的稳定性明显升高, 该结果支持第 2 种推测, 即高原鼯鼠通过提高 α^E 和 β 基因编码多肽的稳定性, 最终使 α^E 型 Hb 积累到与 α^A 型 Hb 相近的浓度。分子进化和蛋白质结构分析结果进一步佐证, 两种类型 Hb 的表达对成体高原鼯鼠适应极端低氧环境是有利的, 其中高氧亲和力的 α^E 型 Hb 有利于肺部毛细血管从低氧环境中装载更多的氧气, 而低氧亲和力的 α^A 型 Hb 则有利于氧气释放到需氧组织的过程。

分子进化分析结果表明, 高原鼯鼠 α^E 基因相

对保守未检测到正选择位点, 但 α^A 和 β 基因均显示出显著的正选择位点。蒲鹏^[31] 已较为深入地研究了氨基酸突变对高原鼯鼠 α^A 型 Hb 结构的影响, 本文主要探讨了 Hb β 亚基血红素“口袋”开口处 12 个氨基酸突变可能对结构造成的影响。结果表明氨基酸突变对 β 亚基结构最突出的影响是 43 和 86~87 位氨基酸突变, 这些突变通过形成两个 α 螺旋使血红素口袋区域更加稳定, 同时 80 位氨基酸突变使得开口夹角处 α 螺旋消失, 部分缓解了该区域的张力使得血红素“口袋”的开口略微增大。此外, 70Ala-Ser 和 86Ala-Ser 两个突变明显增加了 β 亚基潜在的亲水性。因此, 本文推测高原鼯鼠 Hb β 亚基的 12 个氨基酸突变是其稳定性提高的关键因素, 同时氨基酸突变引起的亲水性升高, 均有利于 α^E 型 Hb 组装以及积累到较高浓度。

综上所述, 高原鼯鼠成体表达的两种 Hb 亚型分别为 $(\alpha^A\beta)_2$ 和 $(\alpha^E\beta)_2$; 其异常高的氧亲和力可能主要由 $(\alpha^E\beta)_2$ 高表达引起; Hb β 亚基血红素“口袋”开口处 12 个氨基酸突变对提高 β 亚基稳定性和亲水性起到关键作用。本文对 α^E 型 Hb 最终积累到较高浓度推测及两种 Hb 亚型的实际氧亲和力尚需实验进一步证实。

参考文献:

- [1] STORZ Z, JAY F. Hemoglobin-oxygen affinity in high-altitude vertebrates: Is there evidence for an adaptive trend? [J] *Journal of Experimental Biology*, 2016, 219 (20): 3190-3203.
- [2] JENSEN B, STORZ J F, FAGO A. Bohr effect and temperature sensitivity of hemoglobins from highland and lowland deer mice [J]. *Comparative Biochemistry and Physiology A—Molecular & Integrative Physiology*, 2016, 195: 10-14.
- [3] CHANDRASEKHAR N, HOFFMANN F G, LANIER H C, et al. Intraspecific polymorphism, interspecific divergence, and the origins of function-altering mutations in deer mouse hemoglobin [J]. *Molecular Biology and Evolution*, 2015, 32(4): 978-997.
- [4] OSTOJIC H, CIFUENTES V, MONGE C. Hemoglobin affinity in Andean rodents [J]. *Biological Research*, 2002, 35(1): 27.
- [5] STORZ J F, SCOTT G R, CHEVIRON Z A. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates [J]. *The Journal of Experimental Biology*, 2010, 213(24): 4125-4136.
- [6] ZENG J, WANG Z, SHI Z. Metabolic characteristics and some physiological parameters of mole rat (*Myospalax baileyi*) in alpine area [J]. *Acta Biologica Plateau Sinica*, 1984, 3: 163-171.
- [7] 齐新章, 王晓君, 朱世海, 等. 高原鼯鼠和高原鼠兔心脏对低氧环境的适应 [J]. *生理学报*, 2008, 60 (3): 348-354.
- [8] WEI D B, WEI L, ZHANG J M, et al. Blood-gas properties of plateau zokor (*Myospalax baileyi*) [J]. *Comparative Biochemistry and Physiology A—Molecular & Integrative Physiology*, 2006, 145: 372-375.
- [9] 王晓君, 魏登邦, 魏莲, 等. 高原鼯鼠和高原鼠兔肺细叶的结构特征 [J]. *动物学报*, 2008, 54(3): 531-539.
- [10] PU P, LU S, NIU Z, et al. Oxygenation properties and underlying molecular mechanisms of hemoglobins in plateau zokor (*Eospalax baileyi*) [J]. *American Journal of Physiology—Regulatory Integrative and Comparative Physiology*, 2019, 317(5): R696-R708.
- [11] WEBER R E, JARVIS J U M, FAGO A, et al. O₂ binding and CO₂ sensitivity in haemoglobins of subterranean African mole rats [J]. *Journal of Experimental*

- Biology, 2017, 220(21): 3939–3948.
- [12] STORZ J F, RUNCK A M, MORIYAMA H, et al. Genetic differences in hemoglobin function between highland and lowland deer mice [J]. *Journal of Experimental Biology*, 2010, 213 (15) : 2565–2574.
- [13] STORZ J, WEBER R, FAGO A. Oxygenation properties and oxidation rates of mouse hemoglobins that differ in reactive cysteine content [J]. *Comparative Biochemistry and Physiology A—Molecular & Integrative Physiology*, 2011, 161: 265–270.
- [14] BRITTAIN T. Molecular aspects of embryonic hemoglobin function [J]. *Molecular Aspects of Medicine*, 2002, 23(4): 293–342.
- [15] 马丽英, 蒲鹏, 陈强, 等. 地山雀血红蛋白高原低氧适应的分子机制 [J]. *中国生物化学与分子生物学报*, 2021, 37(1): 88–101.
- [16] NATARAJAN C, HOFFMANN F, WEBER R, et al. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings [J]. *Science*, 2016, 354: 336–339.
- [17] CAMACHO C, COULOURIS G, AVAGYAN V, et al. BLAST⁺: Architecture and applications [J]. *BMC Bioinformatics*, 2009, 10(1): 421.
- [18] GRABHERR M G, HAAS B J, YASSOUR M, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data [J]. *Nature Biotechnology*, 2013, 29: 644.
- [19] EDGAR R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic Acids Research*, 2004, 32(5): 1792–1797.
- [20] HUELSENBECK J P, RONQUIST F. MRBAYES: Bayesian inference of phylogenetic trees [J]. *Bioinformatics*, 2001, 17(8): 754–755.
- [21] YANG Z. PAML 4: Phylogenetic analysis by maximum likelihood [J]. *Molecular Biology and Evolution*, 2007, 24(8): 1586–1591.
- [22] LANGMEAD B, SALZBERG S L. Fast gapped-read alignment with Bowtie 2 [J]. *Nature Methods*, 2012, 9(4): 357–359.
- [23] WEBB B, SALI A. Comparative protein structure modeling using Modeller [J]. *Current Protocols in Protein Science*, 2016, 86(1): 2.9.1–2.9.37.
- [24] HUMPHREY W, DALKE A, SCHULTEN K. VMD: Visual molecular dynamics [J]. *Journal of Molecular Graphics*, 1996, 14(1): 33–38.
- [25] HOFFMANN F G, STORZ J F, GORR T A, et al. Lineage-specific patterns of functional diversification in the alpha- and beta-globin gene families of tetrapod vertebrates [J]. *Molecular Biology and Evolution*, 2010, 27(5): 1126–1138.
- [26] HOFFMANN F G, OPAZO J C, STORZ J F. New genes originated *via* multiple recombinational pathways in the β -globin gene family of rodents [J]. *Molecular Biology and Evolution*, 2008, 25(12): 2589.
- [27] OPAZO J C, HOFFMANN F G, STORZ J F. Genomic evidence for independent origins of β -like globin genes in monotremes and therian mammals [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(5): 1590–1595.
- [28] AGUILETA G, BIELAWSKI J P, YANG Z. Gene conversion and functional divergence in the β -globin gene family [J]. *Journal of Molecular Evolution*, 2004, 59(2): 177–189.
- [29] OPAZO J C, HOFFMANN F G, NATARAJAN C, et al. Gene turnover in the avian globin gene families and evolutionary changes in hemoglobin isoform expression [J]. *Molecular Biology and Evolution*, 2014, 32(4): 871–887.
- [30] STORZ J F, HOFFMANN F G, OPAZO J C, et al. Adaptive functional divergence among triplicated alpha-globin genes in rodents [J]. *Genetics*, 2008, 178(3): 1623–1638.
- [31] 蒲鹏. 高原麝鼠血红蛋白氧合特性及其分子机制 [D]. 兰州: 兰州大学, 2018.

(责任编辑 张冰)