



畜牧科学研究

## 基于全基因组重测序的现代马群体遗传结构和选择信号分析

徐媛<sup>1</sup>, 徐继平<sup>2</sup>, 刘小英<sup>2</sup>, 姜雨<sup>1\*</sup>

(1. 西北农林科技大学 动物科技学院, 陕西杨凌 712100; 2. 榆林市榆阳区补浪河区区域畜牧兽医工作站, 陕西榆林 719000)

**摘要:** 本研究基于全基因组重测序数据, 以全球 76 个马品种共 406 匹马为研究对象, 采用主成分分析、系统发育树和祖先成分分析三种方法, 在遗传学层面综合评估了现代马群体的遗传结构。此外, 以核苷酸多样性为指标对现代马群体的遗传多样性进行了评估, 并运用  $F_{st}$  和 XP-EHH 受选择方法检测不同马群体之间的基因组选择信号, 筛选出了 5 个与藏马高原适应性显著相关的候选基因 (*HBB*, *EPAS1*, *TMEM247*, *RHOQ* 和 *ATP6V1E2*), 3 个与中国西南马体型性状显著相关的候选基因 (*HMGA2*, *TBX3* 和 *TBX5*), 以及与欧美高度选育马的马术表演、行为和运动显著相关的基因 (*IGFBP3*, *IGFBP1*, *HSD7A* 和 *ADCY1*)。该研究为进一步保护地方马种的种质资源提供科学依据, 并揭示人工培育马种的选择方向。

**关键词:** 全基因组重测序; 群体结构; 选择信号分析; 候选基因

[中图分类号] S813.23 [文献标志码] A [文章编号] 1004-6704(2024)-05-0001-07

### Modern Horse Population Genetic Structure and Selection Signal Analysis Based on Whole-genome Sequencing

XU Yuan<sup>1</sup>, XU Jiping<sup>2</sup>, LIU Xiaoying<sup>2</sup>, JIANG Yu<sup>1\*</sup>

(1. College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China;  
2. Bulanghe Regional Animal Husbandry and Veterinary Workstation, Yulin, Shaanxi 719000, China)

**Abstract:** This study, based on whole-genome sequencing data, used 406 horses from 76 horse breeds worldwide as the research subjects. Three sophisticated methodologies—principal component analysis, phylogenetic tree, and ancestral component analysis, were used to comprehensively assess the genetic structure of modern horses at the genetic level. Additionally, genetic diversity of modern horse populations was evaluated using nucleotide diversity as an indicator, and the  $F_{st}$  and XP-EHH selection methods were applied to detect genomic selection signals between different horse populations. Five candidate genes (*HBB*, *EPAS1*, *TMEM247*, *RHOQ*, and *ATP6V1E2*) significantly related to the high-altitude adaptation of Tibetan horses, three candidate genes (*HMGA2*, *TBX3*, and *TBX5*) significantly related to body size traits of Chinese Southwest horses, and genes (*IGFBP3*, *IGFBP1*, *THSD7A*, *ADCY1*) significantly related to the equestrian performances, behavior, and movements of highly selected horses in Europe and America were selected. This study provides a scientific basis for further protecting the germplasm resources of local horse breeds and reveals the selection direction of artificially cultivated horse breeds.

**Key words:** whole-genome sequencing; population structure; selection signal analysis; candidate genes

[收稿日期] 2024-06-25

[第一作者] 徐媛(1999-), 女, 博士研究生, 主要从事家养动物基因组学研究工作。E-mail: xuyuanchn@163.com

\* [通信作者] 姜雨, E-mail: yu.jiang@nwfau.edu.cn

人类社会的发展离不开马的驯化, 家马的出现促进了不同文化之间的交流, 通信以及运输的便利; 除此之外, 人类通过骑马获得远超自身的速度, 为扩

展领土从而发动革命性的战争。随着项圈和马蹄铁在农业中的引入,马越来越多地被用于土地耕种,从18世纪开始的养马业对农业生产力的提高和经济发展起到了重要作用,直到20世纪初期,马逐渐被内燃机取代。马极大地影响了人类文明的发展,而人类也通过选择和杂交重塑了马的基因组<sup>[1]</sup>。在1909年,最后一匹欧洲野马(*Equus ferus ferus*)灭绝了<sup>[2]</sup>。随着人类逐渐意识到保护生物多样性的重要性,普氏野马得以在圈养环境中生存下来,但目前的数量仅有2 109只,仍处于濒危状态,约四分之一的普氏野马被重新引进了中国和蒙古的保护区<sup>[3]</sup>。工业革命之后,尽管人们通过改变育种的措施,进而培育出不同表型的家马品种,但也使大量的地方品种数量减少,极大的改变了家马的遗传结构,从而导致家马遗传资源的衰退。

中国作为养马大国,目前已有40多个马品种,其中23个固有品种、17个培育品种、7个引入品种<sup>[4]</sup>。由于中国地域辽阔,气候差异大,某些地方品种为了适应多样的环境条件,表现出了较高的遗传多样性,这种现象可能是自然选择和遗传变异的长期作用而形成的。近年来中国家马遗传资源流失的现象日趋严重,除因外来引种杂交引起遗传资源的衰退,许多地方品种的数量也在减少<sup>[5]</sup>。国际养马业的形势流行于选育遗传性能稳定、血缘纯正、生产能力强的马品种,如英国纯血马,从而导致种质资源极为丰富的地方马品种逐渐被排斥和取代,这表明人工育种对现代马业的发展产生了显著影响。

本研究通过收集全球范围内的现代马品种,探究现代马的群体结构,并检测不同家马群体间的显著选择信号,旨在筛选与环境适应相关的候选基因,为了解现代马群体的遗传结构、加快马品种的育种改良,以及保护地方马种提供了科学依据。

## 1 材料和方法

### 1.1 数据来源与质控

本研究收集全球76个马品种的原始数据,共406匹马。用FASTP v0.23.2软件<sup>[6]</sup>去除接头,利用Slat v1.0.3软件将数据比对到马参考基因组(EquCab 3.0)得到bam文件,使用Slat v1.0.3软件生成gvcf文件,并通过Simgvcf\_gpu v2.0.7软件生成原始的vcf文件。用BCFtool v1.17软件<sup>[7]</sup>进行 $QD < 2.0 \mid FS > 60.0 \mid MQ < 40.0 \mid MQRankSum < -12.5 \mid ReadPosRankSum < -8.0 \mid SOR > 3.0$ 的snp硬过滤,用BCFtools v1.17软件提取maf > 0.01且缺失率小于0.1的二等位点,从而得到

用于分析的vcf文件。

### 1.2 注释变异和统计测序深度

首先在NCBI数据库(<https://www.ncbi.nlm.nih.gov/>)下载马参考基因组的注释文件,利用SnEff v4.3软件<sup>[8]</sup>生成马参考基因组注释库,并对过滤后的变异位点进行注释。使用Qualimap v2.2软件<sup>[9]</sup>统计bam文件的测序深度和覆盖率。

### 1.3 遗传多样性分析

使用VCFtools v0.1.13软件<sup>[10]</sup>计算窗口大小为30 kb、滑动步长为10 kb的现代马群体的核苷酸多样性。

### 1.4 主成分分析

首先使用PLINK v1.90软件<sup>[11]</sup>将vcf文件转为bed格式,之后利用EIGENSOFT v5.0软件<sup>[12]</sup>的Smartpca工具对现代马进行主成分分析,计算各主成分的解释度。

### 1.5 系统发育树分析

使用邻接法构建系统发育树,使用PLINK v1.90软件<sup>[11]</sup>构建成对遗传距离矩阵,然后使用Perl脚本转换成MEGA软件所需的输入格式,分别使用MEGA软件<sup>[13]</sup>和iTOL网站<sup>[14]</sup>构建和美化NJ树。

### 1.6 祖先成分分析

对388匹马进行祖先成分分析,首先使用PLINK v1.90软件去连锁,之后使用ADMIXTURE version 1.3.0软件<sup>[15]</sup>计算了K值从2到9时的遗传结构,统计每次K值对应的CV error值,CV值越小,可信度越高,最终使用最小CV error值对应的K值作为群体遗传结构分析的最终结果。

### 1.7 选择信号分析

1.7.1 群体分化指数(Fst) 利用VCFtools v0.1.13软件,使用窗口大小为30 kb、滑动步长为10 kb的滑动窗口计算藏马与中国地方马、中国北方马与中国西南马,欧美高度选育马和其他家马的Fst加权平均值。取群体遗传分化指数Top 1%作为识别潜在候选阈值。

1.7.2 跨群体扩展单倍型纯合性(XP-EHH) 用BEAGLE v5.4软件<sup>[16]</sup>对vcf集合进行定相填充,使用BCFtools v1.17软件提取对应分组的vcf文件,使用Selscan v1.3.0软件<sup>[17]</sup>计算藏马与中国地方马、中国北方马与中国西南马,欧美高度选育马和其他家养马的跨群体扩展单倍型纯合性(XP-EHH),并利用Perl脚本计算窗口大小为30 kb,步长10 kb的窗口XP-EHH值。取XP-EHH值Top 1%作为识别潜在候选区域的阈值。

## 2 结果与分析

### 2.1 变异检测和注释

本研究共收集世界范围内马重测序样本 406 只。样本测序深度范围从  $1.14\times$  到  $55.95\times$ , 平均测序深度为  $19.4\times$ , 覆盖率范围为 88.5% 到 99.96%。经过变异检测和质量过滤, 共检测到 27 994 945 个 SNP 位点, 对这些位点进行注释, 其中位于基因间区的位点比例为 44.61%。位于内含子区的 SNP 数量为 11 372 764, 占 40.62%, 同义突变的比例为 0.67%, 错义突变的比例为 0.53%, 如表 1。

### 2.2 主成分分析

为探究普氏野马和家马之间的遗传关系, 基于全基因组常染色体所有 SNP 进行主成分分析, 从图 1A 中, 第一主成分 PC1(解释度: 4.84%) 上来看, 普氏野马单独聚为一支, 与家马显著分开。从图 1B 中可以看出, 在家马中, 北欧马、蒙古马和中国地方马在 PC1 中与欧美集中选育的马分隔开, 之后北欧马在 PC2 上与蒙古马和中国地方马区分开。

### 2.3 核苷酸多样性分析

计算不同马群体之间的遗传多样性, 如图 2A, 可以看出蒙古马和藏马的核苷酸多样性最高, 其次是中国北方马和中国西南马, 北欧马的核苷酸多样性仅次于中国地方马, 在家马中, 欧美高度选育的马核苷酸多样性最低。普氏野马在所有现代马中遗传多样性最低。

表 1 变异注释

Table 1 Variation annotation

变异类型	SNP 数量
3'端 UTR 变异	281 469
5'端 UTR 提前起始密码子获得变异	22 152
5'端 UTR 变异	94 011
下游基因变异	1 362 667
起始密码子变异	33
基因间变异	12 488 993
基因内变异	155 223
内含子变异	11 372 764
错义突变	148 993
非编码转录本外显子突变	109 866
剪切受体突变	1 053
剪切供体突变	1 294
剪切区域变异	27 694
起始位点丢失	339
终止密码子获得	2 447
终止密码子丢失	310
终止密码子保留	154
同义突变	187 367
上游基因变异	1 738 115
总数	27 994 945

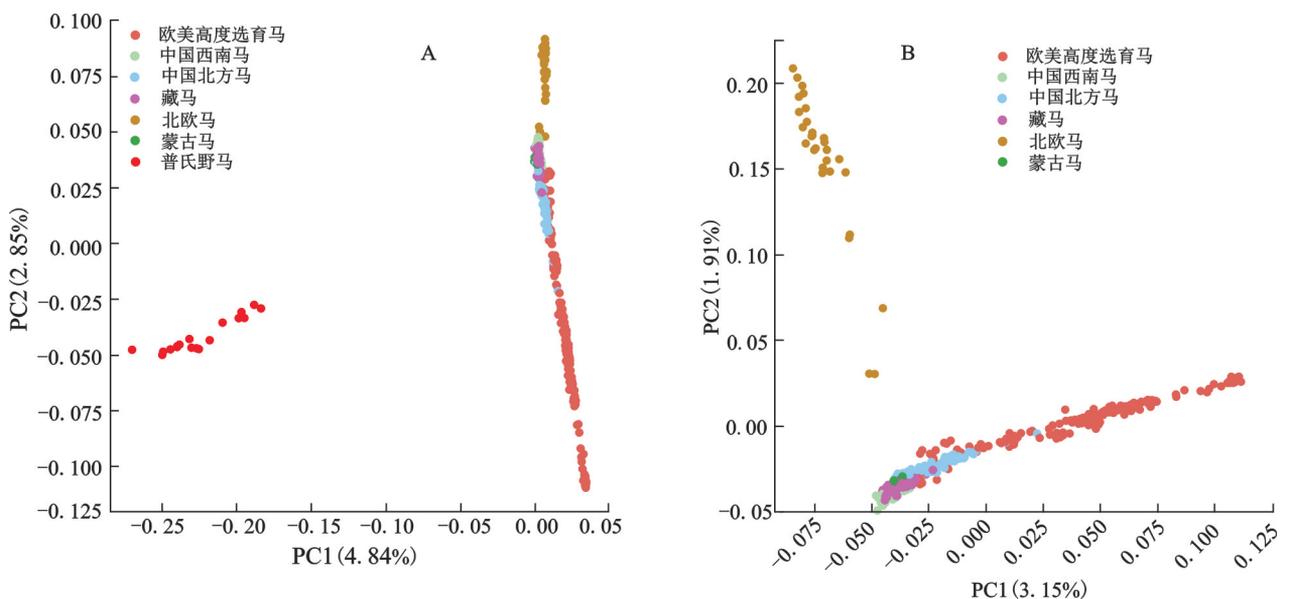


图 1 基于常染色体 SNP 的 PCA 分析

Fig. 1 PCA analysis based on autosomes SNP

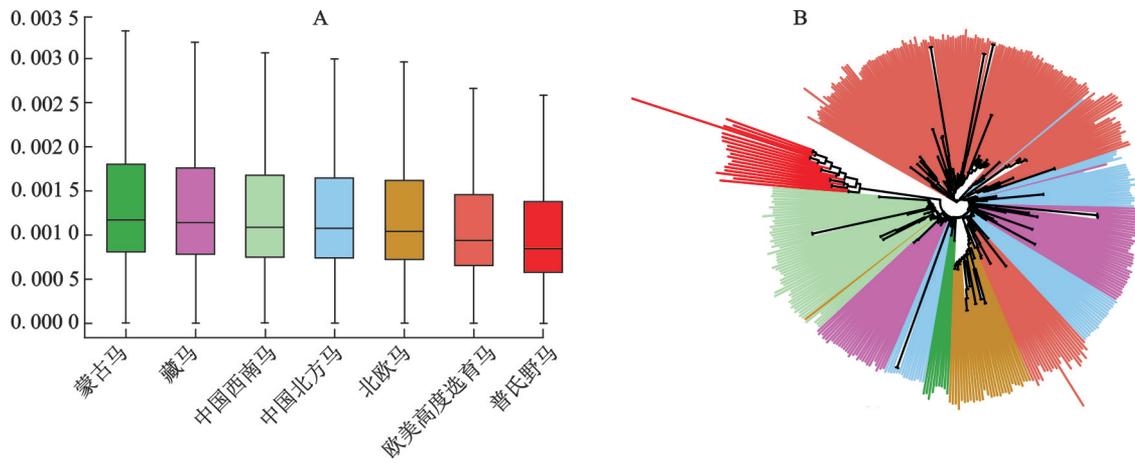


图 2 核苷酸多样性分析和系统发育树分析

Fig. 2 Nucleotide diversity analysis and phylogenetic tree analysis

表 2 不同 K 值对应的 CV error 值

Table 2 CV error values corresponding to different K values

K 值	K=2	K=3	K=4	K=5	K=6	K=7	K=8
CV error 值	0.325 03	0.220 74	0.219 61	0.218 96	0.220 18	0.221 27	0.222 60

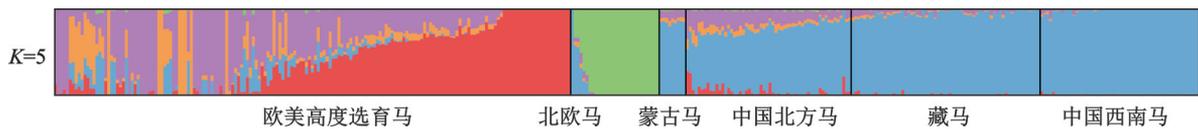


图 3 家马的祖先成分分析

Fig. 3 Analysis of ancestral components of domestic horses

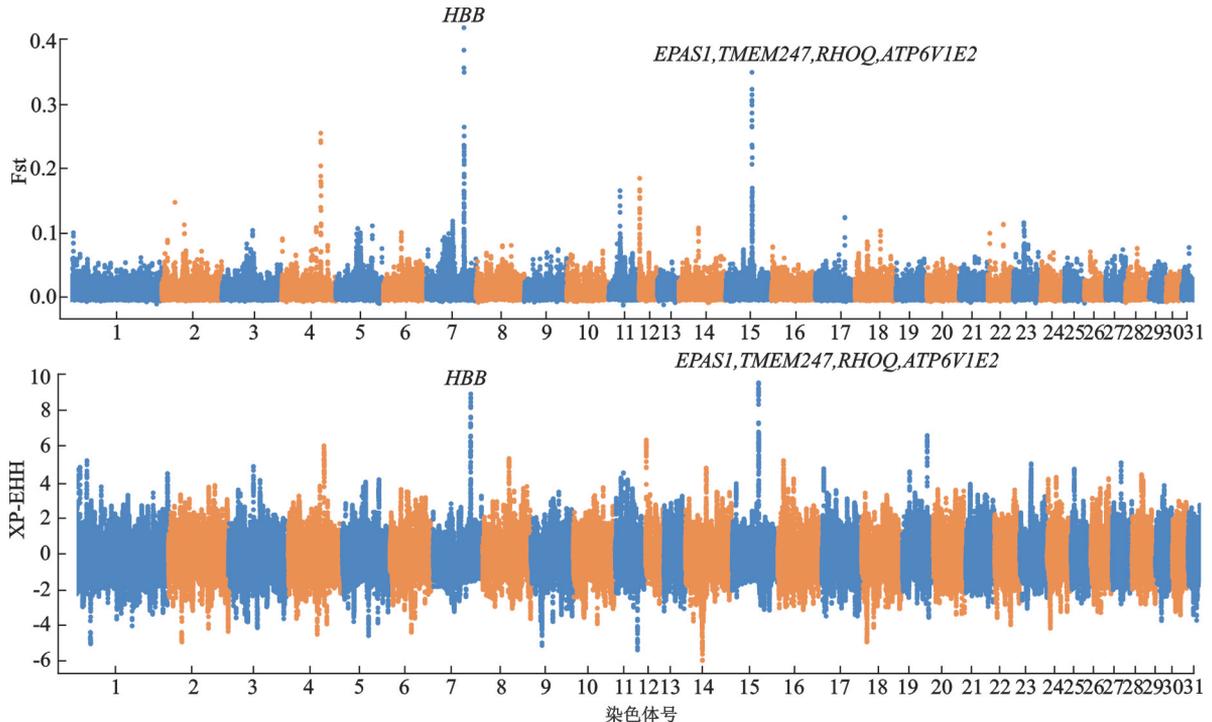


图 4 中国藏马和其他中国地方马的选择信号分析

Fig. 4 Selection signal analysis of Chinese Tibetan horses and other Chinese indigenous horses

## 2.4 系统发育树分析

普氏野马先单独聚成一簇,其他家马都聚在一起,与 PCA 的结果较为一致,如图 2B。北欧马,蒙古马分别聚成一簇,中国地方马内部主要分为西南

马,北方马以及青海—西藏马。

## 2.5 祖先成分分析

对 388 匹马进行祖先成分分析,如表 2 所示,最优分群为  $K=5$ ,即 CV error 值为 0.218 96。从

图 3 可以看出,  $K=5$ , 北欧马有其独特的祖先成分, 中国地方马和蒙古马的祖先成分较为相似, 一部分欧美高度选育的马血统较为纯净, 几乎全为红色。

### 2.6 选择信号分析

对中国藏马和其他中国地方马进行  $F_{st}$  和 XP-EHH 分析, 如图 4 所示,  $F_{st}$  和 XP-EHH 的峰值均

集中于 7 号染色体和 15 号染色体上。将两种方法前 1% 用作识别潜在候选区域的阈值, 两种方法取交集共注释到了 182 个基因。其中 7 号染色体主要注释到了 *HBB* 基因, 15 号染色体上注释到了 *EP-AS1*、*TMEM247*、*RHOQ*、*ATP6V1E2* 等与高原适应性相关的基因。

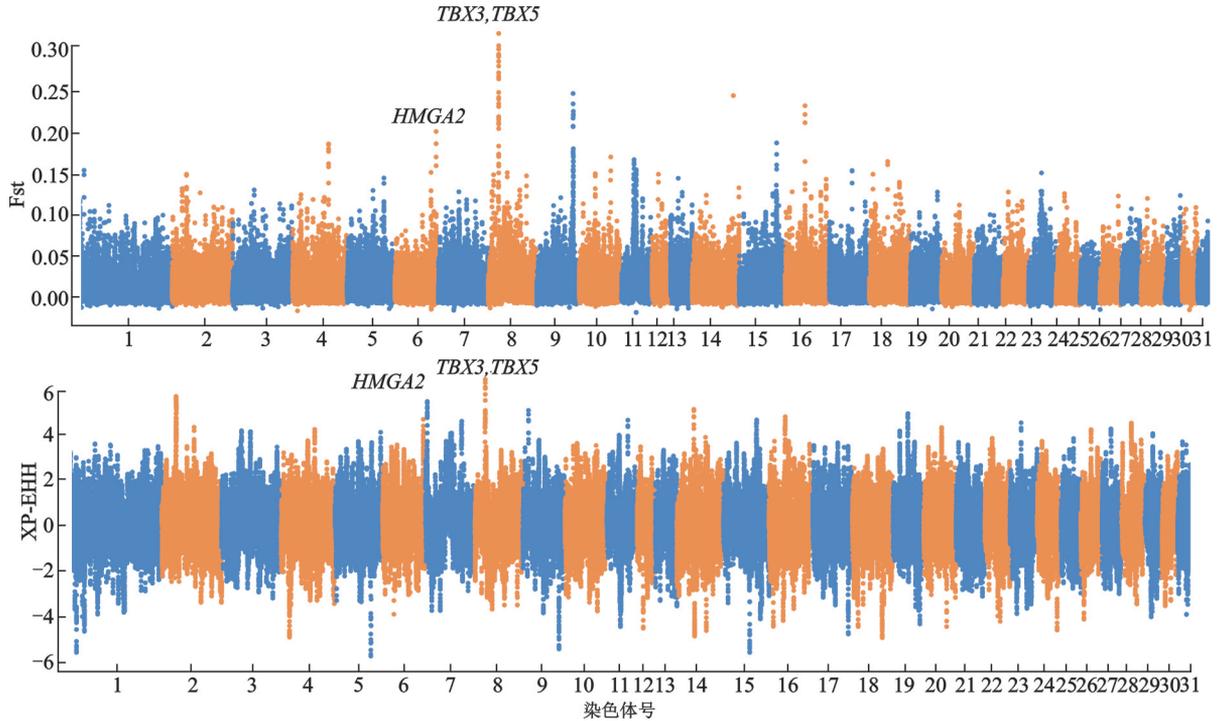


图 5 中国西南马与中国北方马的选择信号分析

Fig. 5 Selection signal analysis of horses in southwest China and horses in North China

*IGFBP3, IGFBP1, THSD7A, ADCY1*

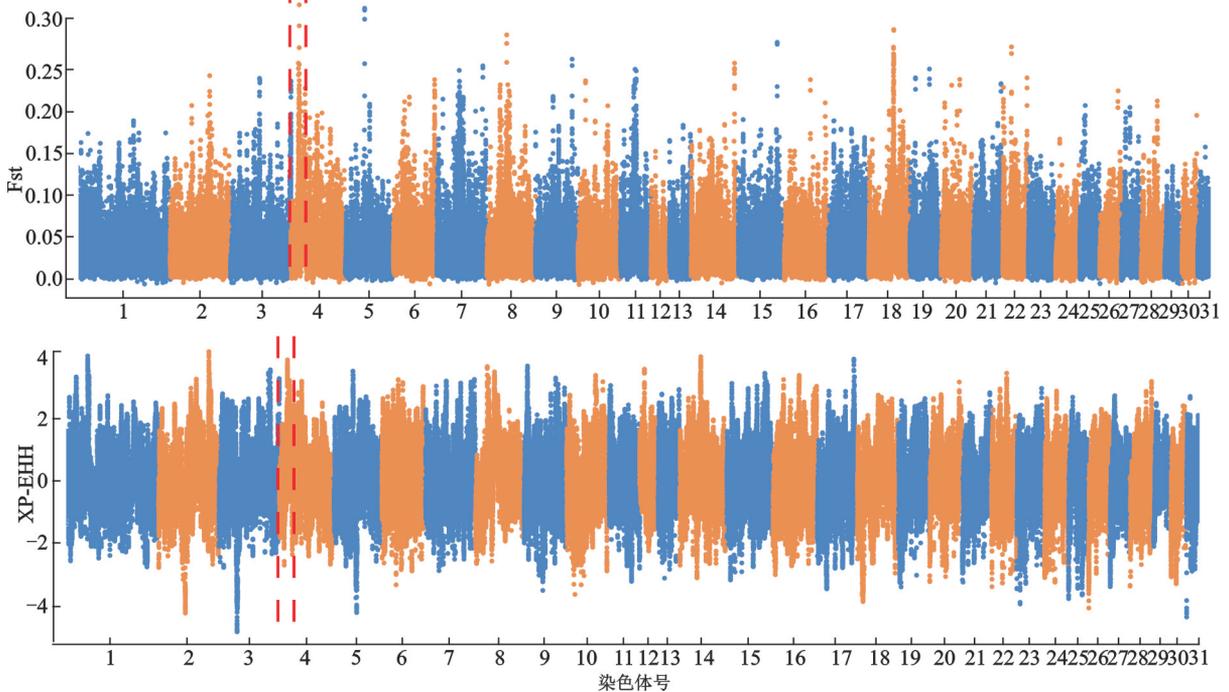


图 6 欧美高度选育马和中国地方马的选择信号分析

Fig. 6 Selection signal analysis of highly bred horses in Europe and America and indigenous horses in China

对中国西南马与中国北方马进行  $F_{st}$  和 XP-EHH 分析,如图 5 所示, $F_{st}$  和 XP-EHH 的峰值均集中于 6 号染色体和 8 号染色体上,两种方法取交集共注释到了 140 个基因。其中 6 号染色体主要注释到了 *HMGA2* 基因,8 号染色体上注释到了 *TBX3* 和 *TBX5* 等与马体型性状相关的基因。

对欧美高度选育马和中国地方马进行  $F_{st}$  和 XP-EHH 分析,如图 6 所示, $F_{st}$  和 XP-EHH 的峰值均集中于 4 号染色体。两种方法取交集共注释到了 166 个基因。其中 4 号染色体主要注释到了 *IGFBP3*, *IGFBP1*, *THSD7A*, *ADCY1* 等与马行为和运动相关的基因。

### 3 讨论

#### 3.1 群体遗传结构分析

群体结构和遗传多样性的表征对于遗传评估、环境适应性理解以及马品种遗传资源的利用和保护至关重要。根据群体遗传分析,可以看出普氏野马先被单独分出来,且聚为一支。普氏野马的遗传多样性是最低的。最初,人们认为生存在野外的普氏野马(*Equus ferus ssp. przewalskii*)是野马的直系后代,然而经过验证发现普氏野马其实是博泰马的后代,并非真正的野马<sup>[18]</sup>。普氏野马目前的数量只有 2 109 只,仍处于濒危状态,种群数量较小,遗传多样性较低。北欧马单独先分成一簇,且含有独特的祖先成分,在家马先单独分出来。其次,中国地方马品种与蒙古马在遗传上展现出较高的相似性,这可能与元朝时期由于战争和贸易导致的蒙古马种的广泛传播有关。

#### 3.2 选择信号分析

对藏马和其他中国地方马进行选择信号分析,其中 7 号染色体主要注释到了 *HBB* 基因。血红蛋白亚基  $\beta$  基因,已知其功能障碍会导致严重的血红蛋白病,例如镰状细胞性贫血和  $\beta$  地中海贫血<sup>[19]</sup>,该基因的正常表达可能有助于藏马适应高海拔地区。15 号染色体上注释到了与高原适应性相关的基因。内皮 PAS 结构域蛋白 1 基因(即 *EPAS1* 基因)编码缺氧诱导因子(HIF)的一个亚基,参与调节血管生成与血红蛋白浓度和红细胞增多的作用<sup>[20]</sup>。*EPAS1* 基因在藏马基因组中显示出明显的正选择信号,其中有两个错义突变与血液生理指数密切相关,这些指数促进了血液循环以及缺氧条件下的氧气运输和消耗。通过功能验证表明,这些突变增加了 *EPAS1* 的稳定性及其对 ARNT 的异源二聚体之间的蛋白亲和力。藏马适应高海拔缺氧环境的关

键进化特征之一是 *EPAS1* 基因发生的错义突变,这种遗传变异可能增强了它们在低氧条件下的生理功能<sup>[21]</sup>。*TMEM247* 基因在多种生物学过程中发挥作用,包括对缺氧、炎症、葡萄糖和脂质代谢、能量平衡、胰岛素受体信号传导、血液凝固以及角蛋白丝的反应。这些功能表明 *TMEM247* 可能在适应高海拔环境的生理机制中扮演重要角色<sup>[22]</sup>。*TP6V1E2* 和 *RHOQ* 基因通过参与 G 蛋白偶联受体和胰岛素受体的信号传导途径,在缺氧条件下发挥关键作用,这些途径对于细胞适应低氧环境至关重要<sup>[22]</sup>。

西南马在地方马种中,体型较小,对中国西南马与中国北方马进行选择信号分析,主要找到了和体型性状相关的基因。其中 6 号染色体主要注释到了 *HMGA2* 基因,*HMGA2* 基因在马和其他物种(包括人类和狗)的身高中具有决定性的作用,该基因先前被报道与设特兰矮马的体型矮小有关<sup>[23]</sup>。8 号染色体上注释到了 *TBX3* 和 *TBX5* 基因。*TBX3* 基因属于 T-box 基因家族,在骨骼发育和肌肉等方面有重要作用,尤其是调控肢体前轴与后轴的发育方面<sup>[24]</sup>。在小鼠中,*TBX3* 缺失会导致小鼠肢体发育迟缓和指头缺失,前肢的尺寸减小比后肢更明显。在中国马中,*TBX3* 基因增强子区域的一个 A/G 单核苷酸多态性变异对体型变化具有显著影响,被认为是一个关键的遗传变异。这一突变能够解释 25% 的身高变异<sup>[25]</sup>。

欧美高度选育的马与其他中国地方马比较找到了 4 个基因。*IGFBP1* 基因和 *IGFBP3* 基因都是胰岛素样生长因子结合蛋白家族的成员,这些蛋白质通过与胰岛素样生长因子 I 结合,调节其生物活性,从而在调控细胞增殖和肌肉组织的生长分布中发挥重要作用<sup>[26]</sup>。*ADCY1* 基因编码的腺苷酸环化酶 1 主要在大脑中表达,对神经可塑性具有重要影响,包括调节长期增强、长期抑郁以及记忆形成等过程<sup>[27]</sup>。在小鼠中,*ADCY1* 基因在恐惧记忆、多巴胺能反应和行为敏化的形成和维持中起重要作用。此外,*ADCY1* 基因参与 G 蛋白偶联受体和 GTP 酶介导的信号转导,这在中枢神经系统中很常见,表明在马表演的学习和反应中起着重要作用。I 型血小板反应蛋白 7A 域(*THSD7A*)参与骨代谢<sup>[28]</sup>,可能影响马的骨骼生长。

中国地方马因其适应性较强,在当地是一种珍贵且应用广泛的遗传资源。鉴于这一遗传资源的重要性,以及马的遗传资源日益枯竭,改善和保护中国本土马的育种计划至关重要。该研究结果为进一步

研究中国地方马的基因组特征与经济重要性状的关系提供了依据。

#### 4 结 论

本研究基于全基因组重测序数据以全球 76 个马品种为研究对象,结果表明,普氏野马遗传多样性最低,属于博泰马的直系后代。在家马中,中国地方马和蒙古马的遗传成分较为相似,北欧马有其独特的祖先成分,欧美高度选育的马在家马中的遗传多样性最低。采用选择信号分析方法,成功筛选出多个与特定马种显著相关的候选基因,为进一步的马匹育种和性能改良提供了潜在的遗传标记。

#### 参考文献:

- [1] LIBRADO P, GAMBA C, GAUNITZ C, et al. Ancient genomic changes associated with domestication of the horse[J]. *Science*, 2017, 356(6 336): 442-445.
- [2] SANDOVAL-CASTELLANOS E, WUTKE S, GONZALEZ-SALAZAR C, et al. Coat colour adaptation of post-glacial horses to increasing forest vegetation[J]. *Nature Ecology & Evolution*, 2017, 1(12): 1 816-1 819.
- [3] DER SARKISSIAN C, ERMINI L, SCHUBERT M, et al. Evolutionary genomics and conservation of the endangered przewalski's horse [J]. *Current Biology*, 2015, 25(19): 2 577-2 583.
- [4] 袁珂蒙, 严达伟. 中国马遗传多样性与起源分化研究进展[J]. *黑龙江畜牧兽医*, 2017(15): 55-58.
- [5] 陶克涛, 韩海格, 赵若阳, 等. 家马的驯化起源与遗传演化特征[J]. *生物多样性*, 2020, 28(6): 734-748.
- [6] CHEN S, ZHOU Y, CHEN Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34(17): 884-890.
- [7] DANECEK P, BONFIELD J K, LIDDLE J, et al. Twelve years of SAMtools and BCFtools [J]. *Giga Science*, 2021, 10(2): 8.
- [8] CINGOLANI P, PLATTS A, WANG L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3 [J]. *Fly*, 2012, 6(2): 80-92.
- [9] GARCÍA-ALCALDE F, OKONECHNIKOV K, CARBONELL J, et al. Qualimap: evaluating next-generation sequencing alignment data [J]. *Bioinformatics*, 2012, 28(20): 2 678-2 679.
- [10] DANECEK P, AUTON A, ABECASIS G, et al. The variant call format and VCF tools [J]. *Bioinformatics*, 2011, 27(15): 2 156-2 158.
- [11] PURCELL S, NEALE B, TODD-BROWN K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses [J]. *The American Journal of Human Genetics*, 2007, 81(3): 559-575.
- [12] PATTERSON N, PRICE A L, REICH D. Population structure and eigenanalysis [J]. *PLOS Genetics*, 2006, 2(12): 190.
- [13] KUMAR S, STECHER G, LI M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms [J]. *Molecular Biology and Evolution*, 2018, 35(6): 1 547-1 549.
- [14] LETUNIC I, BORK P. Interactive tree of life (iTOL) v4: recent updates and new developments [J]. *Nucleic Acids Research*, 2019, 47(1): 256-259.
- [15] ALEXANDER D H, NOVEMBRE J, LANGE K. Fast model-based estimation of ancestry in unrelated individuals [J]. *Genome Research*, 2009, 19(9): 1 655-1 664.
- [16] BROWNING B L, ZHOU Y, BROWNING S R. A One-penny imputed genome from next-generation reference panels [J]. *The American Journal of Human Genetics*, 2018, 103(3): 338-348.
- [17] SZPIECH Z A, HERNANDEZ R D. An efficient multithreaded program to perform EHH-based scans for positive selection [J]. *Molecular Biology and Evolution*, 2014, 31(10): 2 824-2 827.
- [18] GAUNITZ C, FAGES A, HANGHØJ K, et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses [J]. *Science*, 2018, 360(6 384): 111-114.
- [19] CARLICE-DOS-REIS T, VIANA J, MOREIRA F C, et al. Investigation of mutations in the HBB gene using the 1 000 genomes database [J]. *Plos One*, 2017, 12(4): 174 637.
- [20] BEALL C M, CAVALLERI G L, DENG L, et al. Natural selection on EPAS1 (HIF2 $\alpha$ ) associated with low hemoglobin concentration in Tibetan highlanders [J]. *Proceedings of the National Academy of Sciences*, 2010, 107(25): 11 459-11 464.
- [21] LIU X, ZHANG Y, LI Y, et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in tibetan horses [J]. *Molecular Biology and Evolution*, 2019, 36(11): 2 591-2 603.
- [22] SHI J, JIA Z, SUN J, et al. Structural variants involved in high-altitude adaptation detected using single-molecule long-read sequencing [J]. *Nature Communications*, 2023, 14(1): 8 282.
- [23] FRISCHKNECHT M, JAGANNATHAN V, PLATTET P, et al. A non-synonymous HMGA2 variant decreases height in shetland ponies and other small horses [J]. *Plos One*, 2015, 10(10): 140 749.