· 论著 ·

国内北方汉族人群 *HLA-A*、*-B*、*-C* 基因序列多态性与进化分析

宋文倩[1] 余婉珍[2] 白素宁[2] 王丽颖[1] 邵林楠[1] 周世航[1] 梁晓华[1]
(1. 大连市血液中心,辽宁 大连 116001;2. 大连医科大学 检验医学院,辽宁 大连 116044)

摘要:目的 基于长读长测序技术,分析北方汉族人群 *HLA-A*、*-B*、*-C* 基因全长序列多态性特征,评估其选择信号与谱系进化模式。**方法** 纳入 408 名无血缘关系健康献血者,采用 PacBio Sequel II 平台对 *HLA-A*、*-B* 和 *-C* 位点进行长读长测序。通过 Ewens-Watterson 中性检验、滑动窗口核苷酸多样性(Nucleotide diversity,π)与 Tajima's D 分析,以及遗传距离热图、主坐标分析(principal coordinates analysis,PCoA)和最大似然系统发育树,评估三个位点的多态性分布、选择信号及谱系结构。**结果** *HLA-A*、*-B* 和 *-C* 位点分析序列长度分别为 3 257、3 794 和 3 831 bp;位点内序列类型分别为 44、82 和 56 种;位点内序列类型多样性分别为 0.905 2、0.964 9 和 0.939 8。Ewens-Watterson 检验未发现偏离中性模型(P>0.05)。滑动窗口分析显示,*HLA-A* 和 *HLA-B* 在外显子 2、3 及部分内含子区域出现较高 π 值和 Tajima's D 峰值;*HLA-C* 整体多样性相对较低。遗传距离热图、PCoA 和系统发育树表明,HLA-A 和 HLA-C 家族聚类清晰,而 *HLA-B* 谱系结构异质性最高。**结论** 中国北方汉族人群 *HLA-A*、*-B*、*-C* 基因序列具有较高多态性和明显区段特异性。HLA-B 在多样性、选择信号及谱系复杂性均显著高于 HLA-A 和 HLA-C。

关键词:HLA-I 类;长读长测序;位点内序列类型;核苷酸多样性;Tajima's D;Ewens-Watterson 检验
中图分类号:R394 R457.1+1 **文献标识码:**A **文章编号:**1004-549X(2026)6-0743-07
DOI:10.13303/j.cjbt.issn.1004-549x.2026.06.006

Sequence polymorphism and evolutionary analysis of *HLA-A*, *-B*, and *-C* loci in the northern Han Chinese population

SONG Wenqian[1], *YU Wanzhen*[2], *BAI Suning*[2], *WANG Liying*[1], *SHAO Linnan*[1], *ZHOU Shihang*[1], *LIANG Xiaohua*[1].
1. *Dalian Blood Center*, *Dalian* 116001, *China*; 2. *College of Laboratory Medicine*, *Dalian Medical University*, *Dalian* 116044, *China*.
Corresponding author: *LIANG Xiaohua*, *Email*: *liangxiaohua*1968@126.*com*

Abstract: Objective To characterize the sequence polymorphism of *HLA-A*, *-B*, and *-C* loci in the northern Han Chinese population and to evaluate their selection signals and phylogenetic patterns based on long-read sequencing data. **Methods** A total of 408 unrelated healthy blood donors were enrolled. Long-read sequencing of *HLA-A*, *-B*, and *-C* loci was performed using the PacBio Sequel II platform. The Ewens-Watterson neutrality test, sliding-window nucleotide diversity (π) and Tajima's D analyses were conducted, together with genetic distance heatmaps, principal coordinates analysis (PCoA), and maximum likelihood phylogenetic analysis to assess polymorphism distribution, selection signals, and phylogenetic structure across the three loci. **Results** The analyzed sequence lengths for the *HLA-A*, *-B*, and *-C* loci were 3 257 bp, 3 794 bp, and 3 831 bp, respectively; the numbers of locus-specific sequence types were 44, 82, and 56, with corresponding locus-specific sequence type diversities of 0.905 2, 0.964 9, and 0.939 8, respectively. The Ewens-Watterson test showed no significant deviation from neutrality (P>0.05). Sliding-window analysis revealed elevated π and Tajima's D values in exons 2 and 3 as well as in certain intronic regions of *HLA-A* and *HLA-B*, whereas *HLA-C* exhibited relatively lower overall diversity. Genetic distance heatmaps, PCoA, and phylogenetic analysis consistently showed clear clustering in *HLA-A* and *HLA-C*, while *HLA-B* displayed the highest phylogenetic heterogeneity. **Conclusion** *HLA-A*, *-B*, and *-C* gene sequences in the northern Chinese Han population exhibit high levels of polymorphism and marked regional heterogeneity. *HLA-B* shows greater diversity, stronger signals of selection, and more complex phylogenetic structure compared with *HLA-A* and *HLA-C*.

基金项目:大连市科技创新基金项目(2025JJ13PT055)
第一作者简介:宋文倩(1987-),女,副主任技师,Email:wenqian_song@163.com; **通信作者:**梁晓华,Email:liangxiaohua1968@126.com

Key words: HLA class I; long-read sequencing; locus-specific sequence type; nucleotide diversity; Tajima's D; Ewens-Watterson test

人类白细胞抗原(Human Leukocyte Antigen, HLA)系统是基因组中多态性最为丰富的基因簇之一[1]。HLA-I 分子(*HLA-A*、*-B* 和*-C*)负责呈递内源性抗原肽,介导 $CD8^+$ T 细胞的特异性免疫应答。HLA-I 分子的序列多态性不仅直接影响抗原肽结合谱和适应性免疫应答,而且可调控自然杀伤细胞(natural killer, NK)细胞表面的杀伤细胞免疫球蛋白样受体(killer cell immunoglobulin-like receptors, KIR)与 HLA 分子的相互作用[2]。此外研究表明 *HLA-A*、*-B*、*-C* 与*-DRB1* 错配与造血干细胞移植不良结局相关[3]。在人群遗传与进化层面,HLA 呈现典型的"超多态性"与长谱系结构,广泛认为与病原体驱动的平衡选择有关;尤其是与抗原/肽结合功能相关的结构域区域,往往表现出更强的多样性与选择信号[4]。

传统 HLA 群体遗传学研究多聚焦于部分外显子区域(例如外显子 2 和 3),难以全面反映 HLA 基因的多态性特征,且可能遗漏非抗原识别区变异及等位基因顺反相(cis/trans)信息[5]。而长读长测序能够覆盖跨外显子、内含子及 UTR 区域的连续序列,提供更完整的变异信息并减少相位解析歧义[6]。目前针对中国北方汉族人群 *HLA-A*、*-B*、*-C* 基因进化分析数据较少,本文基于 PacBio 长读长测序技术,采用 Ewens-Watterson 检验、核苷酸多样性(Nucleotide diversity, π)及 Tajima's D 滑动窗口分析,以及遗传距离、主坐标分析(Principal Coordinates Analysis, PCoA)及系统发育分析等方法揭示了北方汉族人群 *HLA-A*、*-B*、*-C* 基因序列的多态性模式及其潜在选择信号,现报道如下。

1　对象与方法

1.1　研究对象

纳入 408 名无血缘关系的健康献血者,标本采集时间为 2024 年 7 月至 2025 年 3 月。受试者年龄范围 21~60 岁,平均年龄 45± 10 岁,其中男性 317 例(77.7%),女性 91 例(22.3%)。

1.2　试剂与仪器

血液基因组 DNA 提取试剂盒(苏州海狸生物科技有限公司,批号:20240802)、PCR 扩增试剂 KOD-FX-NEO(Toyobo 公司,批号:346300)、*HLA-A*、*-B* 和*-C* 基因特异性引物(西安浩瑞基因科技有限公司)、Qubit dsDNA BR assay kit(Thermo Fisher Scientific,批号:Q32854)、ADAPTER INDEX 接头及相关建库酶试剂(西安浩瑞基因科技有限公司,批号:20240811)、VAHTS DNA Clean Beads 磁珠纯化试剂(南京诺唯赞生物科技有限公司,批号:N411-01)及 PacBio Binding Kit 3.2 测序引物及酶试剂(PacBio 公司,Binding Kit 3.2,批号:102-333-300)。PCR 扩增仪(Applied Biosystems,美国)和 PacBio Sequel II 高通量单分子实时测序系统(Pacific Biosciences,美国)。

1.3　方法

1.3.1　基因组 DNA 提取

采用磁珠法从 408 名献血者的外周血标本中提取基因组 DNA,最终 DNA 浓度调整至 15~30 ng/μL,纯度 A_{260}/A_{280} 为 1.6~1.8。

1.3.2　PCR 扩增

采用基因组 DNA 为模板,利用 *HLA-A*、*-B* 和*-C* 位点特异性引物对目标区域进行扩增。*HLA-B* 和 *HLA-C* 扩增片段覆盖 5′UTR、3′UTR、全部外显子及内含子;*HLA-A* 扩增片段覆盖 5′UTR、全部外显子和内含子及大部分 3′UTR,但 3′UTR 末端约 0.3 kb 区域未完全覆盖。所应用的特异性引物由西安浩瑞基因技术有限公司提供,其序列、设计方法、扩增条件见已授权发明专利[7]。PCR 产物经 Qubit dsDNA BR assay kit 定量及琼脂糖凝胶电泳质检,合格后用于后续文库构建。

1.3.3　文库构建

文库构建采用商用 ADAPTER INDEX 接头及配套建库酶试剂,主要操作流程参照厂家说明书。取 4 μL PCR 产物,加入 2.5 μL ADAPTER INDEX 接头、5.4 μL DNase/RNase-Free 去离子水及 3.1 μL Enzyme Mix 酶混合物,总体积 15 μL。反应条件:37℃孵育 40 min,25℃孵育 40 min,65℃孵育 10 min,完成接头连接,使每个标本获得唯一标识的环形 DNA 分子。随后加入 2 μL Exonuclease Mix 外切酶混合物,37℃消化 60 min,去除连接失败产物,并用 0.45× VAHTS DNA Clean Beads 磁珠纯化,获得测序文库。在文库中加入测序引物(Binding Kit 3.2),20℃孵育 30 min,使引物结合至文库接头;再加入测序酶(Binding Kit 3.2),30℃孵育 30 min,使

酶固定至引物上;磁珠纯化后即得可上机文库。

1.3.4 三代测序和 HLA 基因分型

将上机文库加入测序芯片,采用 PacBio Sequel II 高通量基因测序仪进行测序,测序完成后碱基信息以 BAM 格式输出。使用 lima(v2.12.0)软件拆分 HiFi reads 以区分标本;使用 pbMM2(v1.16.0)软件将 reads 与人类参考基因组 GRCh38 比对,配合 samtools(v1.13)、bedtools(v2.26.0)等软件提取目标染色体上的基因序列,生成 fasta 文件;将序列比对至 *HLA-A*,*-B*, *-C* 基因参考序列,过滤非目标基因序列;小变异(SNV 和 indel)通过 deepvariant(v. 1.6.1)软件识别,质量控制阈值设定为 QUAL >30;大变异辅以 IGV 和 SnapGene 软件进行人工复核;涉及复杂区域时辅以 SnapGene 进行序列比对。

HLA 等位基因分型基于编码区序列进行确认。根据参考序列注释提取各位点 CDS 序列,采用 blastn(v2.11.0)与标准 HLA 参考序列进行比对,并结合双算法矫正和人工复核输出 HLA 分型结果。

1.4 统计学分析

1.4.1 序列比对与分析区域确定

所有序列经多序列比对后,按照各位点目标扩增区域中所有标本均可可靠覆盖的共同区间进行统一修剪,以保证同一位点内所有序列具有相同起止坐标,后续分析均基于该统一坐标范围内的序列进行。

1.4.2 位点内序列类型识别与多样性分析

采用 DnaSP 软件(v6.12.03)[8]识别位点内序列类型,并计算位点内序列类型数及其多样性(按 DnaSP 中 haplotype diversity, *Hd* 计算)。*Hd* 按照 Nei 方法[9]计算,用于评价同一 HLA 位点内不同长读长序列类型的总体多样性水平。

本研究中的“位点内序列类型”是指在 *HLA-A*、*-B* 或*-C* 单个位点内,由一条同源染色体上的连续长读长序列定义的去重序列;不同于 *HLA-A*~B~C 多位点组合单体型,也不同于仅依据 IPD-IMGT/HLA 命名体系确定的标准 HLA 等位基因。每个个体在每个位点贡献两条单倍体序列,纯合个体则计为两条相同序列。

1.4.3 中性检验与局部选择信号分析

采用 R(v4.5.2)语言分析 Ewens-Watterson 中性检验[10-11],以评价各位点内序列类型频率分布是否偏离中性模型。为进一步分析多态性沿序列的空间分布,采用 R 语言 seqinr 包进行滑动窗口分析,分别计算核苷酸多样性(π)[12]和 Tajima's *D* 统计量[13]。窗口大小与步长分别固定为 100 bp 与 25 bp,计算时将插入缺失(indel)作为第五种碱基状态纳入分析。π 用于反映不同序列区段的核苷酸变异水平,Tajima's *D* 用于评估不同窗口内等位变异频率分布是否偏离中性预期。

1.4.4 遗传距离、主坐标分析与系统发育分析

基于 MEGA 软件(v12.1.2)计算位点内序列类型两两遗传距离矩阵,模型为 Kimura 双参数模型[14]。并进一步用热图和 PCoA 展示。采用最大似然法(maximum likelihood, ML)构建系统发育树[15],模型选择 Tamura-Nei 模型[16]。

2 结果

2.1 位点内序列类型总体特征

408 名健康献血者获得 *HLA-A*、*-B* 与*-C* 位点的长读长序列共 816 条。各位点的序列长度、序列种类、*Hd* 及 Ewens-Watterson 中性检验结果见表 1。3 个位点的位点内序列类型多样性均较高,其中 *HLA-B* 最高。Ewens-Watterson 检验显示,各位点的观察纯合性(F_{obs})与中性模型下的期望纯合性(F_{exp})差异无统计学意义($P>0.05$)。

表 1 ***HLA-A*、*-B*、*-C* 位点内长读长序列类型特征及 Ewens-Watterson 中性检验结果($n=408$)**

Table 1 Locus-specific sequence type characteristics of long-read sequences and Ewens-Watterson neutrality test results at the *HLA-A*, *-B*, and *-C* loci ($n=408$)

	序列长度(bp)	染色体位置区间	位点内序列类型数量	*Hd*	Ewens-Watterson 中性检验				
					F_{obs}	F_{exp}	P_{lower}	P_{upper}	P
HLA-A	3 257	347 711-350 943	44	0.905 2	0.095 8	0.101 2	0.460 5	0.539 4	0.921 1
HLA-B	3 794	1 758 891-1 762 657	82	0.964 9	0.036 3	0.036 8	0.506 1	0.493 8	0.987 7
HLA-C	3 831	1 673 770-1 677 570	56	0.939 8	0.061 3	0.066 9	0.285 5	0.714 4	0.571 1

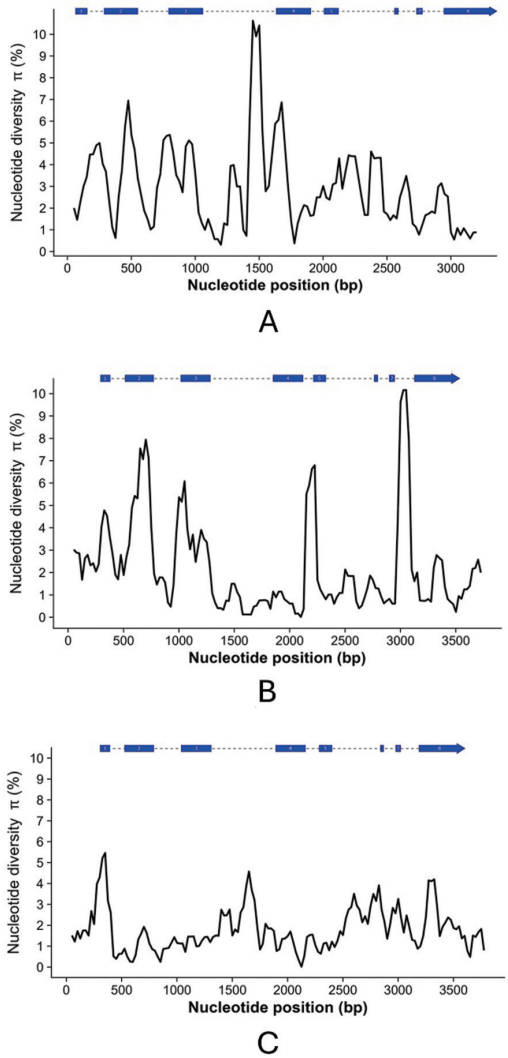
注:表中统计的是单个位点内部的序列多态性。*Hd* 为位点内序列类型多样性;F_{obs} 为样本观察纯合性(位点内序列类型频率平方和);F_{exp} 为中性模型下给定样本量和位点内序列类型数的期望纯合性;P_{lower} 和 P_{upper} 分别为单侧下尾和上尾检验概率;P 为双侧检验概率。显著性阈值 $\alpha=0.05$。

2.2 核苷酸多样性与 Tajima's *D* 滑动窗口分析

图 1 展示了 *HLA-A*、*-B* 与*-C* 三个位点沿全长序列的 π 滑动窗口分布,三个位点均呈现明显的区段性分布特征。*HLA-A* 在外显子 2 和 3 区域出现显著峰值,在内含子 3 及外显子 4 区域亦存在较高多样性水平。*HLA-B* 整体 π 值最高,除外显子 2 和

3 外，在内含子 4 及内含子 7 区域亦观察到明显升高。*HLA-C* 的 π 值整体低于 *HLA-A* 和 *HLA-B*，多样性主要集中于外显子 1 及内含子 3 区域，在内含子 5 至外显子 8 区域存在次级峰值。进一步观察 5′-UTR 和 3′-UTR 区域，π 值存在一定波动，但未见高于其他高变区域的突出峰值。

Tajima's *D* 分析结果与 π 结果类似，三个位点多数窗口的 *D* 值为正，部分窗口呈现较高的正值，少数窗口接近零或负值。*HLA-B* 的 *D* 值波动范围最大，说明其序列层面存在更强的区域异质性；*HLA-A* 与 *HLA-C* 的 *D* 值分布相对稳定。总体来看，π 值较高区域与 Tajima's *D* 正值区段在空间上基本一致，表明这些区域可能累积较多中等频率变异。



注：A、B、C 分别对应 *HLA-A*、*HLA-B*、*HLA-C* 的 π 值分布。

图 1　*HLA-A*、*HLA-B*、*HLA-C* 位点核苷酸多样性（π）滑动窗口图

Figure 1　Sliding-window analysis of nucleotide diversity (π) at the *HLA-A*, *HLA-B*, and *HLA-C* loci

2.3　遗传距离、主坐标分析与系统发育分析

位点内序列类型间遗传距离的热图显示，*HLA-A* 各等位基因家族内部形成低距离聚类区，家族间距离较高，呈现清晰的块状结构。*HLA-B* 遗传距离分布范围更广，多数位点内序列类型间呈中高距离，家族间存在交错重叠，谱系界限模糊。*HLA-C* 整体距离范围较小，聚类结构紧凑，家族间差异相对均匀。



注：A、B、C 分别对应 *HLA-A*、*HLA-B*、*HLA-C* 位点内序列类型的 PCoA。

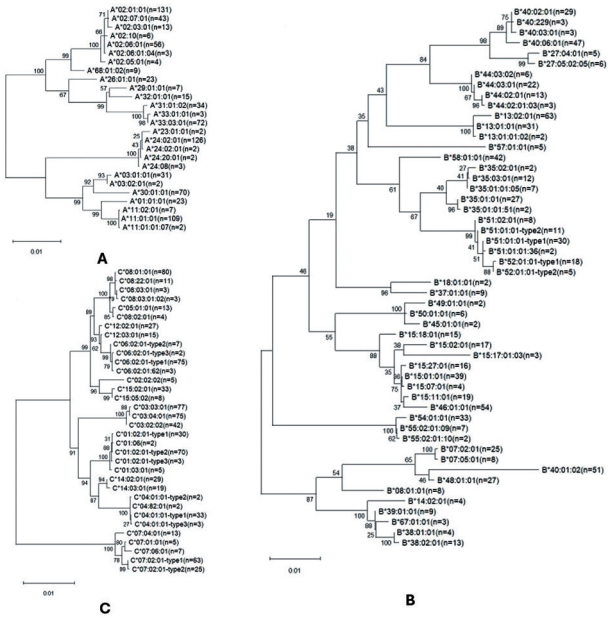
图 2　基于遗传距离矩阵的 *HLA-A*、*HLA-B* 与 *HLA-C* 位点内序列类型的主坐标分析（PCoA）

Figure 2　Principal coordinates analysis (PCoA) of locus-specific sequence types based on genetic distance matrices at the *HLA-A*, *HLA-B* and *HLA-C* loci

PCoA 进一步揭示了各位点的谱系分化结构（图 2）。*HLA-A*（图 2A）前两主坐标解释方差分别为 56.0% 和 23.2%，不同等位基因家族在二维空间中分离相对清晰，家族内部聚集紧密。*HLA-B*（图

2B)前两主坐标解释方差分别为 29.7%和 21.1%,不同家族之间存在明显重叠,整体结构呈连续分布。*HLA-C*(图 2C)主要由少数家族驱动分离,其中 C*07 家族在 PCoA1 方向上明显独立(主坐标解释方差为 67.87%),其余家族聚集于另一侧区域。

ML 系统发育树进一步揭示了 *HLA-A*、*HLA-B* 与 *HLA-C* 位点内序列类型之间的谱系关系(图 3)。*HLA-A* 位点的 ML 系统发育树(图 3A)显示,不同等位基因家族总体呈现较为清晰的谱系聚类结构。*HLA-B* 位点的 ML 系统发育树(图 3B)较 *HLA-A* 和-*C* 显示出更复杂的拓扑结构。*HLA-C* 位点的 ML 系统发育树(图 3C)显示出较为清晰的家族性谱系结构。



注:A、B、C 分别对应 *HLA-A*、*HLA-B*、*HLA-C* 位点内序列类型的系统发育树。

图 3 *HLA-A*、*HLA-B* 与 *HLA-C* 位点内序列类型最大似然(ML)系统发育树

Figure 3 Maximum likelihood (ML) phylogenetic trees of locus-specific sequence types at the *HLA-A*, *HLA-B*, and *HLA-C* loci

3 讨论

本研究中 *HLA-B* 表现出最复杂的谱系结构,与既往关于 HLA-I 类基因进化特征的认识一致。Hughes 和 Nei[4]对 MHC-I 抗原识别区的经典研究表明,抗原识别相关位点的非同义替换显著高于同义替换,这是平衡选择作用于 HLA-I 类基因的重要分子证据。但 HLA 基因高多态性的长期维持并不能完全归因于单一机制,而更可能是杂合优势、负频率依赖选择以及时空波动选择等多种进化过程共同作用的结果[17-18]。因此,本研究中 *HLA-B* 较高的多样性水平及更复杂的树形结构,更适宜理解为多重选择压力长期叠加的结果,而不宜简单归因于单一类型的平衡选择。

本研究 Ewens-Watterson 检验未发现 *HLA-A*、-*B* 和-*C* 三个位点在全局范围内显著偏离中性模型,但这并不意味着这些位点不存在平衡选择作用。Meyer 等[17]指出 HLA 基因保留了平衡选择的演化痕迹,但不同中性检验对选择信号的敏感性并不一致,尤其当选择主要集中于某些功能区段时,全局统计量很容易受到相对保守区段的“稀释”。Radwan 等[18]指出 MHC 多态性的遗传图景本身十分复杂,单一统计量通常只能反映是否偏离中性模型,而难以进一步判定其背后的具体选择机制。本研究中 Ewens-Watterson 检验未见显著偏离,而局部区段的 π 值和 Tajima's *D* 却出现明显高值,这两类结果并不矛盾;它们共同表明,HLA-I 类基因所受选择更可能表现为“局部增强、整体平缓”的模式。滑动窗口分析进一步支持了上述判断。Buhler 等[19]指出,HLA-I 类分子的主要功能多态性集中于外显子 2 和 3 所编码的肽结合区,且这些区域的非同义变异与分子功能差异高度相关。Goeury 等[20]基于全基因 HLA 序列的研究发现,不同位点、不同外显子及不同非编码区的核苷酸多样性分布并不一致,尤其 HLA-I 类不同位点之间存在明显的区段特异性差异。本研究中,π 值与 Tajima's *D* 在 *HLA-A* 和-*B* 的外显子 2、3 及部分内含子区域(如 *HLA-A* 内含子 3,*HLA-B* 内含子 4 和 7)出现明显峰值,与前人关于抗原识别区为主要高变异区的结论一致。Voorter 等[21]在总结 HLA-I 剪接相关研究时指出,部分同义替换、外显子-内含子边界附近变异以及 branch point 相关变异,均可能通过影响剪接模式而改变 HLA 表达水平,甚至形成缺失表达、低表达或表达可疑等表型。因此,本研究中部分内含子区域所观察到的高变异积累,可能不仅反映中性变异背景,也说明这些区段可能具有潜在的调控或剪接功能意义。

在三个位点中,*HLA-B* 整体 π 值较高,而且 Tajima's *D* 的波动范围也最大,反映其在序列层面具有更强的区段异质性。这一结果可能与 *HLA-B* 所承担的功能复杂性密切相关。Raghavan 和 Geng[22]指出,不同 *HLA-B* allotype 之间不仅在肽结合谱方面存在显著差异,而且在胞内装配模式、无

肽状态稳定性及对 tapasin 的依赖程度方面也表现出系统性不同，显示序列变异很可能对应着更高程度的功能分化。

HLA-C 整体 π 值较低，Tajima's *D* 波动相对较小，说明其进化模式可能较 *HLA-A* 和 *HLA-B* 更为保守，PCoA 与系统发育树也显示 *HLA-C* 具有更紧凑的家族聚类结构。Buhler 等[19]提出“联合分歧不对称选择”模型，认为 *HLA-C* 的肽结合多样性显著低于 *HLA-A* 和-*B*，其演化更大程度上受 KIR 配体功能驱动，而并非主要依赖持续扩大肽结合谱。Goeury 等[20]的全基因序列研究也表明，*HLA-C* 某些功能区段的多样性低于 *HLA-A* 和 *HLA-B*，支持其在 HLA-I 系统内部具有一定程度的功能特化。本研究中，C*07 家族在 PCoA1 方向上显著独立，说明 *HLA-C* 内部可能存在功能分化较强的谱系。此外，Voorter 等[21]发现 C*03 等位基因存在内含子 5 保留 18 nt 的选择性剪接亚型，而本研究在 *HLA-C* 内含子 3 和 5 区域观察到 π 峰值，表明这些内含子区段可能包含与剪接调控或表达调控相关的变异。

除内含子外，HLA-I 类基因 5′端启动子和 3′-UTR 等非编码调控区也可能具有进化意义。既往研究显示，中国汉族人群 *HLA-A*、*-B* 及 *HLA-C*w 全长序列中，SNP 和插入/缺失变异不仅分布于外显子，也存在于 5′-启动子、3′-UTR 及内含子等区域，且 *HLA-B* 调控区与外显子之间可能存在更频繁的重组事件，*HLA-C*w 部分内含子和外显子区段也可能受到平衡选择作用[23]。相比之下，本研究在 5′-UTR 和 3′-UTR 区域虽观察到 π 值波动，但未见突出高峰，表明主要多样性仍集中于抗原识别相关外显子及部分内含子区段。考虑到 *HLA-A* 的 3′-UTR 末端约 0.3 kb 未完全覆盖，本研究对 *HLA-A* 3′-UTR 多态性及其调控意义的解释具有一定限制。

遗传距离热图、PCoA 和最大似然系统发育树的结果总体一致：*HLA-A* 与 *HLA-C* 具有相对清晰的家族聚类结构，而 *HLA-B* 表现出更高的谱系异质性和更复杂的分支格局。*HLA-B* 的 ML 系统发育树中，多个家族（如 B*40、B*44、B*15）均由若干亚簇组成，且分支较深，指向其演化历史中可能经历了多次基因转换、重组及长期平衡选择作用，因此整体谱系关系呈现出更连续、更复杂的特征。Goeury 等[20]在西非 Mandenka 人群中的研究也发现，*HLA-B* 外显子 3 抗原识别区的多样性最高，而 *HLA-C* 的肽结合多样性显著偏低，与本研究的趋势一致。在本研究中 *HLA-C* 的 C*07 家族所表现出的显著独立性，则可能反映了较早发生并长期维持的谱系分化，但其形成机制仍需结合更广泛的人群数据和功能研究进一步阐明。

然而本研究也存在一定局限。首先，标本均来自单一区域健康献血者，虽然适合描述中国北方地区 HLA-I 类基因的基本多态性格局，但仍不足以代表更广泛的人群。其次，未纳入 HLA-Ⅱ类基因进行比较分析，限制了对 MHC 区域整体选择压力的理解。第三，本研究仅基于序列层面分析，未结合表达量数据或抗原结合功能实验验证非编码区变异的生物学意义。

综上所述，本研究基于长读长测序数据，系统揭示了中国北方汉族人群 *HLA-A*、*HLA-B* 和 *HLA-C* 基因的全长序列多态性、局部选择信号及谱系结构特征。研究结果表明，*HLA-B* 位点在多样性、选择信号及谱系复杂性方面均显著高于 *HLA-A* 和 *HLA-C*，而后两者则表现出更清晰的分化结构。本研究为理解 HLA-I 类基因在人群中的微进化模式提供了较传统外显子研究更完整的证据基础，也为后续疾病关联、移植匹配和功能研究提供了高分辨率的数据支持。

利益冲突说明/Conflict of Interests

所有作者均声明不存在利益冲突。

伦理批准及知情同意/Ethics Approval and Patient Consent

本研究通过大连市血液中心医学伦理委员会批准（伦理号：LL-05010）。所有研究对象均签署了知情同意书。

作者贡献/Authors' Contribution

宋文倩：研究设计、数据分析及论文撰写；余婉珍、白素宁、邵林楠：标本收集及实验实施；王丽颖：数据整理及统计分析；周世航：论文修订及学术性审阅；梁晓华：研究设计、论文指导及最终定稿。

[参考文献]

[1] NEEFJES J, JONGSMA M L M, PAUL P, et al. Towards a systems understanding of MHC class I and MHC class II antigen presentation[J]. Nat Rev Immunol, 2011, 11(12): 823-836.

[2] POLLOCK N R, HARRISON G F, NORMAN P J. Immunogenomics of killer cell immunoglobulin-like receptor and HLA class I: co-evolution and consequences for human health[J]. J Allergy Clin Immunol Pract, 2022, 10(7): 1763-1775.

[3] FÜRST D, MÜLLER C, VUCINIC V, et al. High-resolution HLA matching in hematopoietic stem cell trans-

plantation: a retrospective collaborative analysis [J]. Blood, 2013, 122(18):3221-3229.
[4] HUGHES A L, NEI M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection[J]. Nature, 1988, 335(6186): 167-170.
[5] KISHORE A, PETREK M. Next-generation sequencing based HLA typing: deciphering immunogenetic aspects of sarcoidosis[J]. Front Genet, 2018, 9:503.
[6] HOSOMICHI K, JINAM T A, MITSUNAGA S, et al. Phase-defined complete sequencing of the HLA genes by next-generation sequencing[J]. BMC Genomics, 2013, 14:355.
[7] 王博, 林奇翰, 杨龙, 等. HLA 基因扩增引物、试剂盒、测序文库构建方法及测序方法[P]. 中国, CN113817725B. 2024-05-14.
WANG B, LIN Q H, YANG L, et al. HLA gene amplification primers, kit, sequencing library construction method and sequencing method[P]. China, CN113817725B. 2024-05-14.
[8] ROZAS J, FERRER-MATA A, SÁNCHEZ-DELBARRIO J C, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets[J]. Mol Biol Evol, 2017, 34(12):3299-3302.
[9] NEI M. Molecular evolutionary genetics[M]. New York: Columbia University Press, 1987.
[10] EWENS W J. The sampling theory of selectively neutral alleles[J]. Theor Popul Biol, 1972, 3(1):87-112.
[11] WATTERSON G A. The homozygosity test of neutrality [J]. Genetics, 1978, 88(2):405-417.
[12] NEI M, LI W H. Mathematical model for studying genetic variation in terms of restriction endonucleases[J]. Proc Natl Acad Sci U S A, 1979, 76(10):5269-5273.
[13] TAJIMA F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism[J]. Genetics, 1989, 123(3):585-595.
[14] KIMURA M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences[J]. J Mol Evol, 1980, 16(2): 111-120.
[15] KUMAR S, STECHER G, LI M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. Mol Biol Evol, 2018, 35(6): 1547-1549.
[16] TAMURA K, NEI M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees[J]. Mol Biol Evol, 1993, 10(3):512-526.
[17] MEYER D, AGUIAR V R C, BITARELLO B D, et al. A genomic perspective on HLA evolution[J]. Immunogenetics, 2018, 70(1):5-27.
[18] RADWAN J, BABIK W, KAUFMAN J, et al. Advances in the evolutionary understanding of MHC polymorphism [J]. Trends Genet, 2020, 36(4):298-311.
[19] BUHLER S, NUNES J M, SANCHEZ-MAZAS A. HLA class I molecular variation and peptide-*B*inding properties suggest a model of joint divergent asymmetric selection [J]. Immunogenetics, 2016, 68(6):401-416.
[20] GOEURY T, CREARY L E, BRUNET L, et al. Deciphering the fine nucleotide diversity of full HLA class I and class II genes in a well-documented population from sub-Saharan Africa[J]. HLA, 2018, 91(1):36-51.
[21] VOORTER C E M, GERRITSEN K E H, GROENEWEG M, et al. The role of gene polymorphism in HLA class I splicing[J]. Int J Immunogenet, 2016, 43(2):65-78.
[22] RAGHAVAN M, GENG J. *HLA-B* polymorphisms and intracellular assembly modes[J]. Mol Immunol, 2015, 68(2):89-93.
[23] KLEIN J, SATO A, O'HUIGIN C. Evolution by gene duplication in the major histocompatibility complex[J]. Cytogenet Cell Genet, 1998, 80(1-4):123-127.

(2026-05-11 收稿,05-15 修回,05-28 接收)

本文编辑:李宜蔓