Available online at www.blood-genomics.com

Blood and Genomics

2025, 9(1): 23−30

B & G

Original Article

The molecular mechanism analysis of 64 new HLA alleles discovered in Beijing HLA Typing Laboratory

Dan Zhang, Chengyan Fan, Yujie Wen, Na Liu, Dongmei Wang*

Beijing HLA Typing Laboratory, Beijing Red Cross Blood Center, Beijing 100088, China.

ABSTRACT

Human leukocyte antigen (HLA) genes rank among the most polymorphic genes in the human genome, with an extensive variation in HLA alleles. To elucidate the mechanisms underlying the generation of novel HLA alleles and to establish a reference for future HLA genotyping and allele discovery, we analyzed a total of 64 newly identified HLA alleles from the Beijing HLA Typing Laboratory of Beijing Red Cross Blood Center. The novel alleles were analyzed against their most closely related sequence homologs. This study indicates that point mutations constitute the primary molecular mechanism driving the emergence of new gene, with missense mutations being the predominant type. Among the 124 base mutations identified in this study, cytosine C exhibited the highest mutational propensity, while thymine T showed the lowest propensity to mutate. Additionally, we identified 10 mutagenic positions at the HLA-B locus, specifically at positions 379, 412, 419, 527, 538, 539, 603, 605, 610, and 618. This study provides critical insights into the origin and evolution of HLA genes and reveals the molecular basis of immune response diversity.

Keywords: HLA, new alleles, point mutation, synonymous mutation, missense mutation

INTRODUCTION

Human leukocyte antigen (HLA) is a complex encoded by a group of tightly linked loci on human chromosome 6p21.3, which is the main histo-compatibility antigen in humans. It plays a critical role in blood transfusion, organ transplantation, as well as the occurrence, progression, and prognosis of multiple diseases[1]. Since Dausset's groundbreaking identification of the first HLA antigen in 1958, advances in genotyping technologies have facilitated the continuous discovery of novel HLA alleles. As of December 2024, the IMGT/HLA database has released a total of 28 062 HLA-I alleles, comprising 8472 for HLA-A, 10 183 for HLA-B, and 8570 for HLA-C, as well as a total of 12 375 HLA-II alleles, including 3751 HLA-DRB1 and 12 771 HLA-DQB1[2].

The China Marrow Donor Program (CMDP) serves as a critical source of hematopoietic stem cell transplantation for patients with blood diseases and various other diseases. Annually, the Beijing HLA Typing Laboratory performs high-resolution genotyping for approximately 2000 patients and donors from the CMDP. A total of 64 new alleles have been discovered by the Beijing HLA Typing Laboratory and have been officially named by the WHO Nomenclature Committee for Factors of the HLA System. This study presents a comprehensive analysis of the molecular mechanism underlying these newly discovered alleles, aiming to summarize the mutagenic sites of HLA genes and provide a reference

*Corresponding to: Dongmei Wang, Beijing HLA Typing Laboratory, Beijing Red Cross Blood Center, Beijing 100088, China. E-mail: wang3dm@aliyun.com.

Received 6 March 2025, Revised 7 April 2025, Accepted 28 April 2025

https://doi.org/10.46701/BG.20250100771

for future discovery and detection of new alleles.

MATERIALS AND METHODS

Subjects

Between 2005 and 2024, the Beijing HLA Typing Laboratory identified 64 new HLA alleles among CMDP-registered patients and donors. The novel alleles were comprehensively characterized using a genotyping approach combining PCR-based sequence-based typing (PCR-SBT) with group-specific sequencing primer (GSSP) amplification, followed by bidirectional Sanger sequencing.

Detection and certification of new HLA alleles

DNA was extracted from the blood samples of study subjects. HLA genotyping of HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQ1 of these samples was performed using the TBG HLAssure SE SBT HLA genotyping kit (Texas BioGene Inc., Taiwan, China), run on the ABI 3730xl Genetic Analyzer (Applied Biosystems, Foster City, California, US), and analyzed with uTYPE6.0 analysis software (Invitrogen, Carlsbad, USA). For samples exhibiting sequence discrepancies with reference sequences or ambiguous typing results, group-specific primers from the TBG HLAssure kit were employed to amplify both DNA strands. Then the sequencing reaction was continued with the amplification products to obtain the single-strand sequencing results. Point mutations were identified through comparative sequence analysis against the most closely related known alleles using specialized software.

Nomenclature of new alleles

The coding nucleotide sequences of the new alleles were deposited in the GenBank database to get an Accession Number, followed by submission to the IPD-IMGT/HLA Database. Subsequently, the WHO Nomenclature Committee for Factors of the HLA System officially assigned new names of these sequences.

RESULTS

Characteristics of new HLA alleles

Among the 64 new alleles involved in this study, 19 alleles belonged to HLA-A, 29 alleles belonged to HLA-B, 7 alleles belonged to HLA-C, 7 alleles belonged to HLA-DRB1, and 2 alleles belonged to HLA-DQB1. Comparative analysis with their closest allele base sequences revealed that 45 alleles (70.3%, 45/64) emerged by one point mutation, while the remaining 19 alleles (29.7%, 19/64) resulted from two or more point mutations. These findings indicate that point mutations were the predominant molecular mechanism for the generation of new HLA alleles (***Table 1***).

Table 1 **Characteristics of mutations in 64 novel HLA alleles**

No.	Novel allele	Most similar allele	Variable sites	Exons	Position and base change	Codon and amino acid change	Mutation type	Property conversion	Location
1	A*01:28	01:01:01:01	2	2	299 C>T	76 Ala>Val	Missense	Non-polar = [a]	α1 domain
					301 A>G	77 Asn>Asp	Missense	Neutral-polar > acidic	α1 domain
2	A*02:99:01	02:06:01:01	1	3	527 T>A	152 Val>Glu	Missense	Non-polar > acidic	α2 domain
3	A*02:127	02:06:01:01	1	3	560 C>A	163 Thr>Lys	Missense	Neutral-polar > basic	α2 domain
4	A*02:129	02:87	5	2	25 T>A	9 Phe>Tyr	Missense	Non-polar > neutral-polar	α1 domain
					29 A>C	10 Thr=	Synonymous	= [b]	α1 domain
					167 T>G	56 Gly=	Synonymous	=	α1 domain
					184 G>A	62 Gly>Glu	Missense	Neutral-polar > acidic	α1 domain
					192 C>G	65 Arg>Gly	Missense	Basic > neutral-polar	α1 domain
5	A*02:145	02:01:01:01	1	3	530 C>T	153 Ala>Val	Missense	Non-polar =	α2 domain
6	A*02:227N	02:01:01:01	1	3	610 C>T	180 Gln>/	Nonsense	/ [c]	α2 domain
7	A*03:01:09	03:01:01:01	1	2	126 G>A	18 Gly=	Synonymous	=	α1 domain
8	A*03:23	03:01:01:01	1	2	221 C>A	50 Pro>Gln	Missense	Non-polar > neutral-polar	α1 domain
9	A*11:01:10	11:01:01:01	1	3	510 G>A	146 Lys>Lys	Synonymous	=	α2 domain
10	A*11:27	11:01:01:01	4	3	599 C>A	163 Arg>Thr	Missense	Basic > neutral-polar	α2 domain
					560 G>C	163 Arg>Thr	Missense	Basic > neutral-polar	α2 domain
					570 G>C	166 Glu>Asp	Missense	Acidic =	α2 domain
					571 T>G	167 Trp>Gly	Missense	Non-polar > neutral-polar	α2 domain

(continued)

Table 1 **Characteristics of mutations in 64 novel HLA alleles (continued)**

No.	Novel allele	Most similar allele	Variable sites	Exons	Position and base change	Codon and amino acid change	Mutation type	Property conversion	Location
11	A*23:20	23:01:01:01	1	4	806 C>T	245 Ala=	Synonymous	=	α3 domain
12	A*24:02:12	24:02:01:01	1	2	303 C>T	77 Asn=	Synonymous	=	α1 domain
13	A*24:02:17	24:02:01:01	1	2	264 A>G	64 Thr>Thr	Synonymous	=	α1 domain
14	A*24:52	24:02:01:01	1	3	362 T>G	97 Met>Arg	Missense	Non-polar > basic	α2 domain
15	A*24:96	24:02:01:01	3	3	411 C>T	113 Tyr>Tyr	Synonymous	=	α2 domain
					412 C>A	114 His>Asn	Missense	Basic > neutral-polar	α2 domain
					419 A>T	116 Tyr>Phe	Missense	Neutral-polar > non-polar	α2 domain
16	A*24:127	24:02:01:01	1	3	614 G>T	181 Arg>Leu	Missense	Basic > non-polar	α2 domain
17	A*24:627	24:02:01:01	1	2	172 G>A	34 Val>Met	Missense	Non-polar =	α1 domain
18	A*26:46	26:01:01:01	1	3	495 G>C	141 Gln>His	Missense	Neutral-polar > basic	α2 domain
19	A*66:07	66:01:01:01	1	3	566 T>C	165 Val>Ala	Missense	Non-polar =	α2 domain
20	B*07:55	07:02:01:01	1	3	538 C>T	156 Arg=	Missense	Basic > non-polar	α2 domain
21	B*07:60	07:02:01:01	6	3	559 G>C	163 Glu>Leu	Missense	Acidic > non-polar	α2 domain
					560 A>T	163 Glu>Leu	Missense	Acidic > non-polar	α2 domain
					603 C>G	177 Asp>Glu	Missense	Acidic =	α2 domain
					605 A>C	178 Lys>Thr	Missense	Basic > neutral-polar	α2 domain
					610 G>C	180 Glu>Gln	Missense	Acidic > neutral-polar	α2 domain
					618 T>G	182 Ala>Ala	Synonymous	=	α2 domain
22	B*13:16	13:02:01:01	1	3	527 T>A	152 Val>Glu	Missense	Non-polar > acidic	α2 domain
23	B*13:18	13:02:01:01	1	3	539 T>G	156 Leu>Arg	Missense	Non-polar > basic	α2 domain
24	B*13:19	13:02:01:01	1	2	268 A>G	66 Ile>Ala	Missense	Non-polar =	α1 domain
25	B*13:20	13:01:01:01	1	3	527 T>A	152 Val>Glu	Missense	Non-polar > acidic	α2 domain
26	B*13:23	13:01:01:01	1	2	97 T>C	9 Tyr>His	Missense	Neutral-polar > basic	α1 domain
27	B*15:133	15:18:01:01	1	3	419 C>T	116 Ser>Phe	Missense	Neutral-polar > non-polar	α2 domain
28	B*15:137	15:04:01:01	4	3	379 G>C	103 Val>Leu	Missense	Non-polar =	α2 domain
					409 C>T	113 His>Tyr	Missense	Basic > neutral-polar	α2 domain
					412 G>A	114 Asp>Asn	Missense	Acidic > neutral-polar	α2 domain
					419 C>T	116 Ser>Phe	Missense	Neutral-polar > non-polar	α2 domain
29	B*15:138	15:40:01	4	3	603 G>C	177 Glu>Asp	Missense	Acidic =	α2 domain
					605 C>A	178 Thr>Lys	Missense	Neutral-polar > basic	α2 domain
					610 C>G	180 Gln>Glu	Missense	Neutral-polar > acidic	α2 domain
					618 G>T	182 Ala=	Synonymous	=	α2 domain
30	B*15:139	15:02:01:01	1	2	277 A>G	69 Thr>Ala	Missense	Neutral-polar > non-polar	α1 domain
31	B*27:36	27:04:01	10	3	463 A>C	131 Ser>Arg	Missense	Neutral-polar > basic	α2 domain
					499 A>T	143 Thr>Ser	Missense	Neutral-polar =	α2 domain
					512 G>T	147 Trp>Leu	Missense	Non-polar =	α2 domain
					527 A>T	152 Glu>Val	Missense	Acidic > non-polar	α2 domain
					603 G>C	177 Glu>Asp	Missense	Acidic =	α2 domain
					605 C>A	178 Thr>Lys	Missense	Neutral-polar > basic	α2 domain
					610 C>G	180 Gln>Glu	Missense	Neutral-polar > acidic	α2 domain
					618 G>T	182 Ala=	Synonymous	=	α2 domain
				4	693 C>T	207 Gly=	Synonymous	=	α3 domain
					704 G>C	211 Gly>Ala	Missense	Neutral-polar > non-polar	α3 domain
32	B*27:44	27:08	1	2	292 G>T	74 Asp>Tyr	Missense	Acidic > neutral-polar	α1 domain
33	B*37:13	37:01:01:01	1	3	527 T>A	152 Val>Glu	Missense	Non-polar > acidic	α2 domain
34	B*39:42	39:01:01:01	2	3	353 C>T	94 Thr>Ile	Missense	Neutral-polar > non-polar	α2 domain
					355 C>A	95 Leu>Thr	Missense	Non-polar > neutral-polar	α2 domain
35	B*39:46	39:01:01:01	1	4	646 G>A	192 Gly>Arg	Missense	Neutral-polar > basic	α3 domain

(continued)

Table 1 **Characteristics of mutations in 64 novel HLA alleles (continued)**

No.	Novel allele	Most similar allele	Variable sites	Exons	Position and base change	Codon and amino acid change	Mutation type	Property conversion	Location
36	B*40:01:06	40:01:02:01	1	3	435 G>A	121 Lys=	Synonymous	=	α2 domain
37	B*40:59	40:30	6	3	361 A>G	97 Arg>Val	Missense	Basic > non-polar	α2 domain
					362 G>T	97 Arg>Val	Missense	Basic > non-polar	α2 domain
					378 C>G	103 Leu>Val	Missense	Non-polar =	α2 domain
					387 C>G	105 Pro=	Synonymous	=	α2 domain
					412 G>A	114 Asp>Asn	Missense	Acidic > neutral-polar	α2 domain
					419 C>A	116 Ser>Tyr	Missense	Neutral-polar =	α2 domain
38	B*40:60	40:01:02:01	7	3	368 A>C	99 Tyr>Ser	Missense	Neutral-polar =	α2 domain
					369 C>T	99 Tyr>Ser	Missense	Neutral-polar =	α2 domain
					379 G>C	103 Val>Leu	Missense	Non-polar =	α2 domain
					387 G>C	105 Pro=	Synonymous	=	α2 domain
					409 C>T	113 His>Tyr	Missense	Basic > neutral-polar	α2 domain
					412 A>G	114 Asn>Asp	Missense	Neutral-polar > acidic	α2 domain
					419 A>C	116 Tyr>Ser	Missense	Neutral-polar =	α2 domain
39	B*40:75	40:06:01:01	1	2	272 C>G	67 Ser>Cys	Missense	Neutral-polar =	α1 domain
40	B*40:78	40:02:01:01	1	3	479 C>T	136 Ala>Val	Missense	Non-polar =	α2 domain
41	B*40:585	40:01:02:01	1	3	460 C>A	130 Leu>Met	Missense	Non-polar =	α2 domain
42	B*48:16	48:01:01:01	1	2	260 A>T	63 Glu>Val	Missense	Acidic > non-polar	α1 domain
43	B*48:19	48:01:01:01	2	3	538 C>T	156 Leu>Trp	Missense	Non-polar =	α2 domain
					539 T>G	156 Leu>Trp	Missense	Non-polar =	α2 domain
44	B*54:16	54:01:01:01	1	3	412 A>G	114 Asn>Asp	Missense	Neutral-polar > acidic	α2 domain
45	B*55:21	55:01:01:01	2	3	538 C>T	156 Leu>Trp	Missense	Non-polar =	α2 domain
					539 T>G	156 Leu>Trp	Missense	Non-polar =	α2 domain
46	B*55:34	55:02:01:01	3	2	259 A>G	63 Asn>Asp	Missense	Neutral-polar > acidic	α1 domain
					261 C>G	63 Asn>Asp	Missense	Neutral-polar > acidic	α1 domain
					272 A>C	67 Tyr>Ser	Missense	Neutral-polar =	α1 domain
47	B*56:18	56:10	4	3	379 C>G	103 Leu>Val	Missense	Non-polar =	α2 domain
					412 A>G	114 Asn>Asp	Missense	Neutral-polar > acidic	α2 domain
					419 T>C	116 Leu>Ser	Missense	Non-polar > neutral-polar	α2 domain
					420 A>C	116 Leu>Ser	Missense	Non-polar > neutral-polar	α2 domain
48	B*58:12	58:01:01:01	1	2	142 G>T	24 Ala>Ser	Missense	Non-polar > neutral-polar	α1 domain
49	C*01:274	01:02:01:01	1	4	791 C>T	240 Thr>Ile	Missense	Neutral-polar > non-polar	α3 domain
50	C*07:43	07:01:01:01	3	2	270 C>G	66 Asn>Lys	Missense	Neutral-polar > basic	α1 domain
				3	353 C>T	94 Thr>Ile	Missense	Neutral-polar > non-polar	α2 domain
					355 C>A	95 Leu>Ile	Missense	Non-polar =	α2 domain
51	C*07:56:01	07:02:01:01	1	2	343 G>A	91 Gly>Arg	Missense	Neutral-polar > basic	α1 domain
52	C*15:02:04	15:02:01:01	1	2	216 G>C	48 Arg=	Synonymous	=	α1 domain
53	C*15:21	15:02:01:01	1	2	302 A>G	77 Asn>Ser	Missense	Neutral-polar =	α1 domain
54	C*15:26	15:02:01:01	1	3	379 C>G	103 Leu>Val	Missense	Non-polar =	α2 domain
55	C*15:279	15:02:01:01	5	4	727 C>T	219 Arg>Trp	Missense	Basic > non-polar	α3 domain
					735 C>G	221 Gly=	Synonymous	=	α3 domain
					756 C>T	228 Thr=	Synonymous	=	α3 domain
				5	900 A>G	276 Pro=	Synonymous	=	Transmem-brane region
					982 A>G	304 Met>Val	Missense	Non-polar =	Transmem-brane region
56	DRB1*04:67	04:10:01	1	2	299 G>A	71 Arg>Lys	Missense	Basic =	β1 domain
57	DRB1*04:388	04:01:01:01	1	2	217 G>A	44 Val>Met	Missense	Non-polar =	β1 domain
58	DRB1*08:33	08:03:02:01	1	2	229 C>T	48 Arg>Trp	Missense	Basic > non-polar	β1 domain
59	DRB1*11:64	11:45	1	2	286 A>C	67 Ile>Leu	Missense	Non-polar =	β1 domain

(continued)

Table 1 **Characteristics of mutations in 64 novel HLA alleles (continued)**

No.	Novel allele	Most similar allele	Variable sites	Exons	Position and base change	Codon and amino acid change	Mutation type	Property conversion	Location
60	DRB1*13:72	13:01:01:01	5	2	135 T>A	16 His>Gln	Missense	Basic > neutral-polar	β1 domain
					195 A>T	37 Asn>Phe	Missense	Neutral-polar > non-polar	β1 domain
					196 A>T	37 Asn>Phe	Missense	Neutral-polar > non-polar	β1 domain
					227 T>A	47 Phe>Tyr	Missense	Non-polar > neutral-polar	β1 domain
					261 C>T	58 Ala=	Synonymous	=	β1 domain
61	DRB1*14:61	14:01:01:01	2	2	124 T>G	13 Ser>Gly	Missense	Neutral-polar =	β1 domain
					125 C>G	13 Ser>Gly	Missense	Neutral-polar =	β1 domain
62	DRB1*14:62	14:01:01:01	1	2	256 G>A	57 Ala>Thr	Missense	Non-polar > neutral-polar	β1 domain
63	DQB1*04:01:02	04:01:01:01	1	2	321 G>T	75 Val>Ala	Missense	Non-polar =	β1 domain
64	DQB1*06:01:04	06:01:01:01	1	2	156 G>C	20 Gly=	Synonymous	Neutral-polar =	β1 domain

a: Although amino acid of the novel allele is changed, its property remains the same; b: Synonymous mutation. No amino acid changes; c: Nonsense mutation. After the mutation, the codon that encodes a polar neutral amino acid changes to a stop codon.

Analysis of the 64 novel alleles revealed 124 base changes, comprising 21 synonymous mutations (16.9%, 21/124), 102 missense mutations (82.3%, 102/124), and only one nonsense mutation (0.8%, 1/124) (***Table 2***). Among the missense mutation cases, 36.3% (37/102) of amino acid changes preserved the

Table 2 **Statistics of mutations in 64 novel HLA alleles**

Characteristic	Cases	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1
Base mutation	64	19	29	7	7	2
One point mutation	45	15	18	5	5	2
Two or more points mutation	19	4	11	2	2	0
Amino acid exchange	124	29	68	13	12	2
Synonymous mutation	21	8	7	4	1	1
Missense mutation	102	20	61	9	11	1
Nonsense mutation	1	1	0	0	0	0
Property changes of amino acid [a]	102	20	61	9	11	1
No change	37	5	22	4	5	1
Non-polar → Neutral-polar	9	3	4	0	2	0
Non-polar → Basic	2	1	1	0	0	0
Non-polar → Acidic	4	1	3	0	0	0
Neutral-polar → Non-polar	10	1	5	2	2	0
Neutral-polar → Basic	9	2	5	2	0	0
Neutral-polar → Acidic	9	2	7	0	0	0
Basic → Non-polar	6	1	3	1	1	0
Basic → Neutral-polar	8	4	3	0	1	0
Basic → Acidic	0	0	0	0	0	0
Acidic → Non-polar	4	0	4	0	0	0
Acidic → Neutral-polar	4	0	4	0	0	0
Acidic → Basic	0	0	0	0	0	0
Location in HLA-I	58	19	30[b]	9[c]		
α1 domain	19	7	8	4		
α2 domain	33	11	20	2		
α3 domain	5	1	2	2		
Transmembrane region	1	0	0	1		
Location in HLA-II	9				7	2
β1 domain	9				7	2
β2 domain	0				0	0

a: property changes of amino acid in nonsense mutations only; b: HLA-B*27:36 mutates at both exon 3 and exon 4; c: HLA-C*07:43 mutates at both exon 2 and exon 3, HLA-C*15:279 mutate at both exon 4 and exon 5.

properties of amino acids, while 63.7% (65/102) altered the properties of amino acids. It is worth noting that these mutations localized to the antigens presenting α1 and α2 domains of HLA class I molecules, as well as the equivalent β1 domain of HLA class II molecules. Such alterations in amino acid properties may affect the three-dimensional structure of the HLA protein, potentially modifying the binding properties and specificity of the HLA molecule.

According to the statistics of 124 base mutations, 27 new alleles emerged through the change of adenine A, of which 8 cases were A>C, 12 cases were A>G, and 7 cases were A>T; 46 new alleles emerged through the change of cytosine C, including 10 cases for C>A, 13 cases for C>G, and 23 cases for C>T; 33 new alleles emerged through the change of guanidine G, comprising 12 cases for G>A, 13 cases for G>C, and 8 cases for G>T; 18 new alleles emerged through the change of thymine T, including 7 cases for T>A, 3 cases for T>C, and 8 cases for T>G (***Table 3***).

Table 3 **Base mutations in novel alleles**

Base mutation	Cases	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1
A	27	4	17	3	3	0
A > C	8	1	6	0	1	0
A > G	12	2	7	3	0	0
A > T	7	1	4	0	2	0
C	46	11	24	8	3	0
C > A	10	4	5	1	0	0
C > G	13	1	8	3	1	0
C > T	23	6	11	4	2	0
G	33	8	18	2	3	2
G > A	12	4	4	1	3	0
G > C	13	3	8	1	0	1
G > T	8	1	6	0	0	1
T	18	6	9	0	3	0
T > A	7	2	3	0	2	0
T > C	3	1	2	0	0	0
T > G	8	3	4	0	1	0
Sum	124	29	68	13	12	2

Mutagenic site analysis

As shown in ***Table 4***, among the 19 new HLA-A alleles, HLA-A*02:127 and HLA-A*11:27 share a common missense mutation at position 560 of exon 3, resulting in changes in the properties of their encoded amino acids. In detail, the coding change at codon 163 of HLA-A*02:127 resulted in the neutral-polar amino acid threonine replaced by basic amino acids lysine, while the coding change at codon 163 of HLA-A*11:27 led to the basic amino acid arginine replaced by neutral-polar amino acid threonine.

Table 4 **Mutation positions found in two or more novel alleles**

Position	Cases	Novel alleles
HLA-A		
560	2	02:127, 11:27
HLA-B		
272	2	40:75, 55:34
379	3	15:137, 40:60, 56:18
387	2	40:59, 40:60
409	2	15:137, 40:60
412	5	15:137, 40:59, 40:60, 54:16, 56:18
419	5	15:133, 15:137, 40:59, 40:60, 56:18
527	4	13:16, 13:20, 27:36, 37:13
538	3	7:55, 48:19, 55:21
539	3	13:18, 48:19, 55:21
603	3	07:60, 15:138, 27:36
605	3	07:60, 15:138, 27:36
610	3	07:60, 15:138, 27:36
618	3	07:60, 15:138, 27:36

Mutation-prone sites are notably more prevalent in HLA-B. In this study, we identified 13 mutation positions where two or more novel alleles exhibited variations. Among these, 10 positions at exon 3 demonstrated mutations in more than three new alleles, specifically at positions 379, 412, 419, 527, 538, 539, 603, 605, 610, and 618. Notably, HLA-B*15:137, HLA-B*40:60, and HLA-B*56:18 were all mutated at positions 379, 412, and 419, while HLA-B*07:60, HLA-B*27:36, and HLA-B*15:138 all shared mutations at positions 603, 605, 610, and 618. It is suggested that the HLA-B gene may be prone to generate point mutation or crossing-over (bidirectional sequence exchange) and gene conversion (unidirectional sequence exchange). In addition, HLA-B*13:16, HLA-B*13:20, and HLA-B*37:13 were all new alleles identified by single-base T>A mutation at position 527, suggesting that base T at this location may be unstable and prone to convert to A. Similarly, HLA-B*07:55, HLA-B*48:19, and HLA-B*55:21 all exhibited C>T base mutations at position 538, further implying that base C at this position may be unstable and prone to T mutation.

DISCUSSION

HLA loci are one of the most polymorphic genes in the human genome, evidenced by a large number of alleles and distinct distribution patterns across populations. HLA molecules are closely related to human genetics and immunology, playing a critical role in immunorecognition, transplantation, and

disease progression[3–4]. HLA polymorphism provides diverse antigen presentation and profoundly affects the quantity and quality of immune protection, which is of crucial importance in evolutionary process[5–6]. With advancements in HLA typing methods and the continued progress of CMDP[7], new HLA alleles are being discovered at an accelerating rate[8–9]. Consequently, the mechanisms and characteristics of HLA new alleles have attracted increasing scientific interest[10–11]. Several molecular mechanisms are believed to contribute to the generation of new HLA alleles, including point mutations, gene conversion, and crossing-over exchange. Point mutations can be categorized into three types: base substitutions, deletions, and insertions, with base substitutions being the most common mutation. Among the 64 newly identified alleles considered in this study, 45 new alleles (70.3%) are identified by point mutations, indicating that point mutations or substitutions are the main mechanism of HLA allele generation[12]. Besides, 14 new alleles mutate at three or more positions which are relatively closed in nucleotide sequences, suggesting potential involvement of bidirectional sequence crossing-over or intralocus/interlocus gene conversion. Further investigation is required to determine whether gene conversion or crossing-over is involved.

In general, point mutations can generate novel gene sequences, thereby producing new alleles. However, whether a new allele encodes a new amino acid chain depends on codon changes. In other words, there is a missense mutation. Among the 124 base mutations identified in this study, 102 (82.3%) resulted in missense mutations. Given the important role of HLA molecules in immune recognition and response, changes in amino acids within their function domains may influence antigen binding and recognition. In addition, 63.7% of amino acid changes altered the properties of amino acids, which may impact HLA protein folding, assembly, and three-dimensional structure. These changes may further alter the association with peptide-editing molecules and recognition by T cell receptors[13], thereby modulating the functional specificity of HLA molecules. However, whether the missense mutations observed in these 64 new alleles result in generation of new serological antigens remains unknown and requires further testing with serological methods.

This study identified multiple mutagenesis positions in the HLA-B gene. Meanwhile, the mutagenesis of positions 272, 387, 412, 419, and 420 can be confirmed by existing literature[14]. It is worth noting that HLA-B*13:26 reported in the literature as well as HLA-B*40:59 and HLA-B*40:60 (identified in this study) all mutate at positions 387, 412, and 419[15–16]. It can be speculated that this sequence may be unstable and prone to gene exchange or recombination, which may be responsible for the largest number of HLA-B alleles[17]. Given these findings, more consideration should be given to this sequence in future genotyping.

The application of next-generation sequencing technologies and in-depth studies of new HLA alleles[18] have stimulated growing research interest in the mechanisms of HLA allele polymorphisms and the regulation of HLA molecule expression. This study provides valuable insights into the origin and evolution of HLA genes and reveals the molecular basis of the diversity of immune responses. The discovery of new HLA alleles is of great significance for precise matching of transplantation and the provision of new disease targets. Future investigation could employ molecular dynamics simulations to characterize three-dimensional structural variances within HLA-peptide-MHC complexes, thereby elucidating how allelic polymorphisms at critical peptide-anchoring residues mechanistically modulate antigen presentation efficiency.

Author contributions

Conceptualization, Dan Zhang and Dongmei Wang; Methodology, Dan Zhang; Validation, Chengyan Fan, Yujie Wen, and Na Liu; Formal Analysis, Dan Zhang; Investigation, Dan Zhang, Chengyan Fan, and Yujie Wen; Resources, Dongmei Wang; Data Curation, Na Liu; Writing – Original Draft Preparation, Dan Zhang; Writing – Review & Editing, Na Liu; Visualization, Dan Zhang; Supervision, Dongmei Wang; Project Administration, Dongmei Wang.

Ethics statement

Not applicable.

Informed consent statement

Not applicable.

Funding

This research received no external funding.

Declaration of competing interest

The authors declared no conflict of interests.

References

[1] The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex[J]. *Nature*, 1999, 401(6756): 921−923.

[2] IPD-IMGT/HLA. Statistics[EB/OL]. [2025-01]. https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/

[3] Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation[J]. *Blood*, 2007, 110(13): 4576−4583.

[4] Sypek M, Kausman J, Holt S, et al. HLA epitope matching in kidney transplantation: an overview for the general nephrologist[J]. *Am J Kidney Dis*, 2018, 71(5): 720−731.

[5] Lawlor DA, Zemmour J, Ennis PD, et al. Evolution of class-I MHC genes and proteins: from natural selection to thymic selection[J]. *Annu Rev Immunol*, 1990, 8: 23−63.

[6] Gustafsson K, Wiman K, Emmoth E, et al. Mutations and selection in the generation of class II histocompatibility antigen polymorphism[J]. *EMBO J*, 1984, 3(7): 1655−1661.

[7] Zhang T, Ouyang X, Zhu S, et al. HLA typing data from 39628 hematopoietic stem cell donors: a retrospective study[J]. *J Clin Hemato*, 2021, 34(04): 251−255.

[8] Barker DJ, Maccari G, Georgiou X, et al. The IPD-IMGT/HLA database[J]. *Nucleic Acids Res*, 2023, 51(D1): D1053−1060.

[9] Robinson J, Barker DJ, Marsh SGE. 25 years of the IPD-IMGT/HLA database[J]. *HLA*, 2024, 103(6): e15549.

[10] Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms[J]. *J Mol Biol*, 2003, 331(3): 623−641.

[11] Baxter-Lowe LA. Defining and characterizing HLA diversity[J]. *Hum Immunol*, 2021, 82(7): 455−456.

[12] Adamek M, Klages C, Bauer M, et al. Seven novel HLA alleles reflect different mechanisms involved in the evolution of HLA diversity: description of the new alleles and review of the literature[J]. *Hum Immunol*, 2015, 76(1): 30−35.

[13] Abualrous ET, Sticht J, Freund C. Major histocompatibility complex (MHC) class I and class II proteins: impact of polymorphism on antigen presentation[J]. *Curr Opin Immunol*, 2021, 70: 95−104.

[14] Zhu C, Nie X, Song Y, et al. Analysis of molecular mechanisms for 37 novel HLA alleles in Chinese hematopoietic stem cell donors[J]. *Int J Blood Transfus Hematol*, 2022, 45(1): 68−73.

[15] Shan X, Gao G, He X, et al. Identification of a novel HLA-B allele HLA-B*4059 in Chinese bone marrow donors[J]. *Tissue antigens*, 2006, 67(4): 339−340.

[16] An R, Hu B, Jiao S, et al. Genomic full-length sequence of the HLA-B*40: 60 allele was identified by full-length group-specific sequencing[J]. *HLA*, 2022, 99(1): 44−46.

[17] Zhong YP, Zou HY, Quan ZR, et al. Analysis of full-length sequence and 18 point mutations of HLA-B in a leukemia patient and her family[J]. *Chin J Tissue Eng Res*, 2020, 24(1): 77−82.

[18] Bravo-Egana V, Sanders H, Chitnis N. New challenges, new opportunities: next generation sequencing and its place in the advancement of HLA typing[J]. *Hum Immunol*, 2021, 82(7): 478−487.