Chapter 3

MHC Genotyping in Human and Nonhuman Species by PCR-based Next-Generation Sequencing

Takashi Shiina, Shingo Suzuki and Jerzy K. Kulski

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61842

Abstract

The major histocompatibility complex (MHC) is a highly polymorphic genomic region that encodes the transplantation and immune regulatory molecules. It receives special attention for genetic investigation because of its important role in the regulation of innate and adaptive immune responses and its strong association with numerous infectious and/or autoimmune diseases. Recently, genotyping of the polymorphisms of MHC genes using targeted next-generation sequencing (NGS) technologies was developed for humans and some nonhuman species. Most species have numerous highly homologous MHC loci so the NGS technologies are likely to replace traditional genotyping methods in the near future for the investigation of human and animal MHC genes in evolutionary biology, ecology, population genetics, and disease and transplantation studies. In this chapter, we provide a short review of the use of targeted NGS for MHC genotyping in humans and nonhuman species, particularly for the class I and class II regions of the Crab-eating Macaque MHC (Mafa).

Keywords: HLA, MHC, polymorphism, genotyping, NGS

1. Introduction

The major histocompatibility complex (MHC) genomic region consists of a large group of evolutionary-related genes involved functionally with the innate and adaptive immune systems in jawed vertebrates [1]. In humans, the MHC is located on the short arm of chromosome 6, band p21.3, and the MHC class I and class II genomic regions encode the highly polymorphic gene complex classified as the human leukocyte antigen (HLA) complex [2, 3]. The HLA class I and class II molecules expressed by the MHC play important roles in restricted

cellular interactions and tissue histocompatibility due to cellular discrimination of "self" and "nonself" that require an essential knowledge of the effects of HLA allele matched and mismatched donors in transplantation medicine [4] and transfusion therapy [5]. While the HLA class I molecules are expressed by all nucleated cells to present processed peptides of intracellular origin to CD8+ cytotoxic T cells and serve as ligands for natural killer cells, the class II molecules are expressed by antigen-presenting cells such as B cells, dendritic cells, or macrophages to present exogenous peptides to CD4+ helper T cells of the immune system [6]. In addition, the classical HLA class I genes, HLA-A, HLA-B, and HLA-C, and the classical HLA class II genes, HLA-DR, HLA-DQ, and HLA-DP are distinguished by their extraordinary polymorphisms, whereas the nonclassical HLA class I genes, HLA-E, HLA-F, and HLA-G, are distinguished by their tissue-specific expression and limited polymorphism [2, 3, 7].

The highly polymorphic HLA genomic region is critically involved in the rejection and graft-versus-host disease (GVHD) of hematopoietic stem cell transplants [8, 9], the pathogenesis of numerous autoimmune diseases [10–13], and infectious diseases [14]. Apart from regulating immunity, the MHC genes may have a role in reproduction and social behavior, such as pregnancy maintenance, mate selection, and kin recognition [15]. The MHC genomic region also appears to influence drug adverse reactions [16, 17], CNS development and plasticity [18–22], neurological cell interactions [23, 24], synaptic function and behavior [25, 26], cerebral hemispheric specialization [27], and neurological and psychiatric disorders [28–32]. Hence, the MHC is one of the most biomedically important genomic regions that warrant special attention for genetic investigation.

In general, the study of the diversity and polymorphic variation of the MHC genomic region has been focused more on humans than any other species and animal population [1] largely because of the high cost and limited throughput of the first generation Sanger sequencing method [33, 34]. However, this is now changing because the next-generation sequencing (NGS) technologies are becoming the method of choice for lower-cost, high-throughput genotyping of MHC genes that are composed of highly homologous multiple loci such as those found in the macaque primate species [35]. Thus, the NGS technologies are expected to perform precise MHC genotyping in human and model animals that already have a collection of MHC allele references, and to facilitate MHC genotyping of wild animals that as yet have no MHC allele references. In addition, the NGS technologies are likely to replace traditional genotyping methods such as subcloning, Sanger sequencing, and previously developed PCR-based MHC typing methods (PCR-RFLP, PCR-SSP, and so on) in the near future. Recently, many articles concerning the development of NGS technologies for precise MHC genotyping and genotyping data of MHC genes using the new NGS technologies have been published on the investigations of human and nonhuman MHC polymorphisms in various fields of study such as medical science, evolutionary biology, ecology, and population genetics.

In this chapter, we provide a short review of the current HLA polymorphism information and the use of PCR-based NGS for MHC genotyping in human and nonhuman species, particularly for the Filipino crab-eating macaque MHC (*Mafa*) class I (Mafa-A, -B, -E, -F, and -I) and class II loci (Mafa-DPA1, -DPB1, -DQA1, -DQB1, and -DRB1).

http://dx.doi.org/10.5772/61842

2. HLA allele number

A total of 13,840 HLA allele sequences, 10,297 in the class I and 3543 in the class II gene regions, were released by the IMGT/HLA database [7] release 3.22 in October 2015 (Table 1).

Category	Locus	Allele no.	Protein no.
Class I	HLA-A	3285	2313
	HLA-B	4077	3011
	HLA-C	2801	1985
	HLA-E	18	7
	HLA-F	22	4
	HLA-G	51	17
	Pseudogene	43	0
	Total	10,297	7337
Class II	HLA-DRA	7	2
	HLA-DRB1	1825	1335
	HLA-DRB3	60	48
	HLA-DRB4	17	10
	HLA-DRB5	22	19
	HLA-DQA1	54	32
	HLA-DQB1	876	595
	HLA-DPA1	42	21
	HLA-DPB1	587	480
	HLA-DMA	7	4
	HLA-DMB	13	7
	HLA-DOA	12	3
	HLA-DOB	13	5
	Pseudogene	8	0
	Total	3543	2561

Table 1. Number and genomic distribution (loci) of HLA alleles

The IMGT/HLA database is a specialist database for HLA sequences. Ten years ago, the allelenumbers were only 2182, but since then, the numbers have increased by 1000 allele sequenceseach year. Of 10,297 HLA class I alleles, 3285, 4077, 2801, 18, 22, and 51 alleles were countedin HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, and HLA-G genes, respectively (Table 1);

10,163 and91 alleles were counted in the classical and nonclassical HLA class I genes, respectively.Of the 3543 HLA class II alleles, 7, 1825, 99, 54, 876, 42, 587, 7, 13, 12, and 13 alleles were countedin HLA-DRA, HLA-DRB1, HLA-DRB3/4/5, HLA-DQA1, HLA-DQB1, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DMB, HLA-DOA, and HLA-DOB genes, respectively (Table 1), with3490 and 45 alleles in the classical and nonclassical HLA class II genes, respectively.

3. History of HLA genotyping methods

Many variations of the conventional HLA genotyping methods such as incorporating restriction fragment polymorphisms (RFLP) [36], single strand conformation polymorphism (SSCP) [37], sequence-specific oligonucleotides (SSOs) [38], sequence-specific primers (SSPs) [39], and sequence-based typing (SBT), like the Sanger method [33], have been used for the efficient and rapid HLA matching in transplantation therapy [40–43], research into HLA-related diseases [2, 3], population diversity studies [44–46], and in forensic and paternity testing [47]. The HLA genotyping methods mainly applied today are PCR-SSOP, such as the Luminex commercial methodology [48, 49], and SBT by the Sanger method employing capillary sequencing based on chain–termination reactions [33, 34]. However, both methods often detect more than one pair of unresolved HLA alleles because of chromosomal phase (*cis*/*trans*) ambiguity [50, 51]. To solve the phase ambiguity problem, new HLA genotyping technologies have been reported and commercialized that combine the PCR amplification of targeted HLA genomic regions with NGS platforms such as the ion PGM system (Life Technologies), GS Junior system (Roche), and the MiSeq system (Illumina) [52}. The PCR/NGS methods are expected to produce genotyping results that detect new and null alleles efficiently without phase ambiguity.

4. Summary of NGS-based human MHC genotyping methods

Table 2 shows list of publications on NGS-based human MHC genotyping that includes information for PCR range, targeted HLA locus, NGS platform, and allele assignment method. The MHC genotyping methods in human are basically composed of three steps, PCR, NGS, and allele assignment. We summarize the important points in each of the three steps below. The more detailed information is described in our previous publication [52].

4.1. PCR step

4.1.1. Long- and short-range PCR

PCR methods produce amplicons of different sequence lengths depending on the primer design and the type of DNA polymerase used for the PCR. The amplicon sizes are usually classified into two size ranges: the short-range system where the amplicon size is <1 kb and the long-range system where the amplicon size is >1 kb as shown in Figure 1.

http://dx.doi.org/10.5772/61842

No.	PCR range	Sorting from PCR range	Targeted HLA locus	NGS platform	Allele assignment method	Ref.
1	410–790 bp	Short-range system	A, B, C, DRB1/3/4/5, DQA1, DQB1, DPB1	454 GS FLX	Conexio Assign ATF	[66]
2	400–900 bp	Short-range system	A, B	454 GS FLX	GS-FLX amplicon variant analyzer	[67]
3	Unknown	Long-range system	A, B, C, DRB1, DQB1	454 GS FLX	Conexio Assign-NG	[51]
4	410–790 bp	Short-range system	A, B, C, DRB1/3/4/5, DQA1, DQB1, DPB1	454 GS FLX	Conexio Genomics ATF	[68]
5	381–537 bp	Short-range system	A, B, C	454 GS FLX	SSAHA2	[69]
6	**4.6–11.2 kb**	**Long-range system**	**A, B, C, DRB1, DQA1, DQB1, DPA1, DPB1**	**454 GS Junior, Ion PGM**	**SeaBass**	[70]
7	2.7–4.1 kb	Long-range system	A, B, C, DRB1	HiSeq2000, Miseq	Alignment with IMGT/HLA data	[71]
8	410–790 bp	Short-range system	A, B, C, DRB1/3/4/5, DQA1, DQB1, DPB1	454 GS FLX, (or GS Junior)	Conexio Assign ATF 454	[72]
9	3.4–13.6 kb	Long-range system	A, B, C, DRB1, DPB1, DQB1	MiSeq	BWA, Samtools, GATK, PerlScript	[73]
10	**5.1–5.6 kb**	**Long-range system**	**DRB3, DRB4, DRB5**	**454 GS Junior**	**SeaBass**	[74]
11	250–270 bp	Short-range system	A, B, C, DRB1, DPB1, DQB1	MiSeq	neXtype	[75]
12	250–270 bp	Short-range system	DRB1/3/4/5, DQA1, DQB1, DPA1, DPB1	MiSeq	Genetics Management System	[76]
13	Unknown	Long-range system	A, B, DRB1	PacBio	Bayes' theorem, NGSengine	[77]
14	**4.0–7.2 kb**	**Long-range system**	**A, B, C, DRB1/3/4/5, DQB1, DPB1**	**Ion PGM**	**SeaBass**	[54]

Table 2. Publication list of NGS-based MHC genotyping in human. Bold letter shows publications from the author's group

The short-range PCR system is a method based on PCR amplification of each exon that includes polymorphic exons 2 and 3 in HLA-A, HLA-B, and HLA-C and exon 2 in HLA-DR, HLA-DQ, and HLA-DP. One of the advantages of the short-range system is that it is the most suitable for application of physically fragmented DNA samples as templates such as those extracted from swabs because the PCR length is relatively short, ranging from 250 bp to 900 bp, per

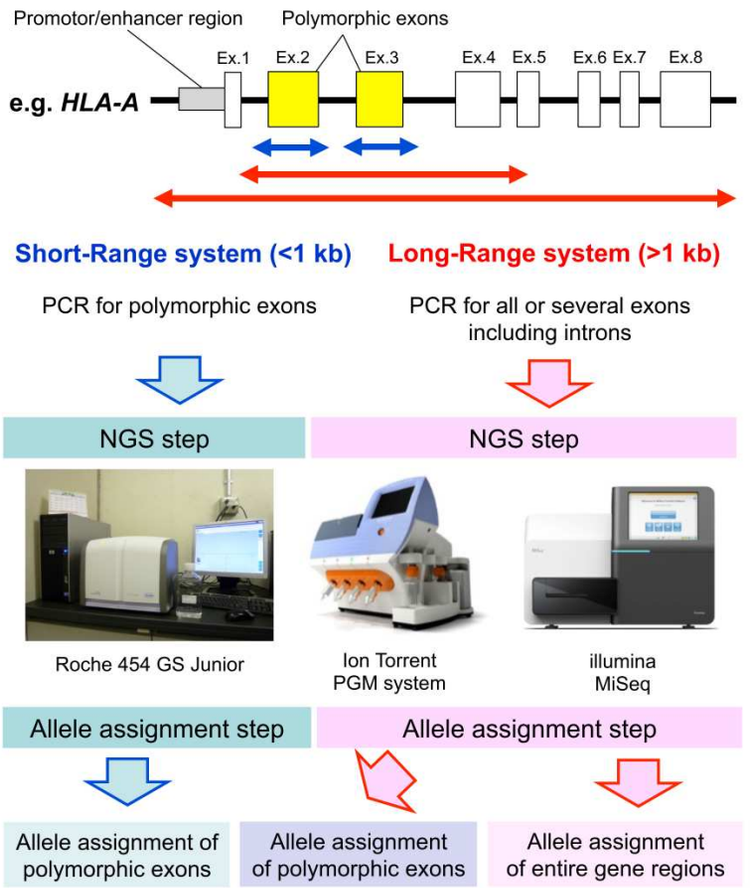

Figure 1. Outline of NGS-based human MHC typing.

amplicon. On the other hand, the short-range system is less effective for genotyping recombinant alleles that have been generated by recombination events of the HLA genes because it is difficult to avoid the phase ambiguities generated by recombinations. For example, in Figure 2, B*15:20 has an identical nucleotide sequence with B*15:01 in exon 2 and B*35:01 in exon 3, but B*35:43 has an identical nucleotide sequence with B*35:01 in exon 2 and B*15:01 in exon 3. When we genotype a DNA sample that has B*15:01 or B*15:20 and B*35:01 or B*35:43, ambiguous genotyping can result in assignments such as B*15:01/20 and B*35:01/43 that are difficult to assign correctly and definitively.

The long-range PCR system is a method based on PCR amplification of the entire HLA gene region including the promotor-enhancer region, 5′ untranslated region (UTR), all exons, all introns, and the 3′ UTR or partial gene regions that include polymorphic exons and adjacent introns (Figure 1). Primer sets for long-range systems have already been developed and published for HLA-A, HLA-B, HLA-C, HLA-DRB1/3/4/5, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1 (Table 2). The advantage of long-range PCR is that this system can easily solve phase ambiguity even if recombinant alleles such as those shown in Figure 2 are present in DNA samples. Also, the long-range PCR method is expected to detect new polymorphisms and variations throughout the entire HLA gene region. Therefore, the long-range

http://dx.doi.org/10.5772/61842

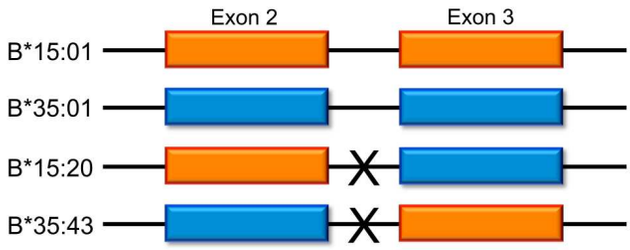

Figure 2. Example of recombinant HLA alleles. B*15:01 and B*15:20 and B*15:01 and B*35:43 have identical nucleotide sequences in exon 2 and in exon 3, respectively (red boxes), and B*35:01 and B*35:43 and B*35:01 and B*15:20 have identical nucleotide sequences in exon 2 and in exon 3, respectively (blue boxes). "X" indicates the recombination site.

system is an important and useful alternative to the short-range system for donor-recipient matching in bone marrow transplantation and HLA-related disease studies. In fact, one of the main themes of the upcoming 17th International HLA and Immunogenetics Workshop (IHIWS) in 2017 [53] is "NGS of full length HLA genes," with the following objectives: (1) to complete the sequence of all HLA alleles of the reference cell lines from the 13th IHIWS and (2) to perform HLA genotyping of 10,000 quartet families of varied ancestry, utilizing at least one NGS method.

4.1.2. Development of multiplex PCR methods

Recently, we developed four kinds of multiplex PCR methods based on the long-range system for genotyping nine HLA loci (HLA-A, -B, -C, -DRB1/3/4/5, -DQB1, and -DPB1) [54] (Figure 3).

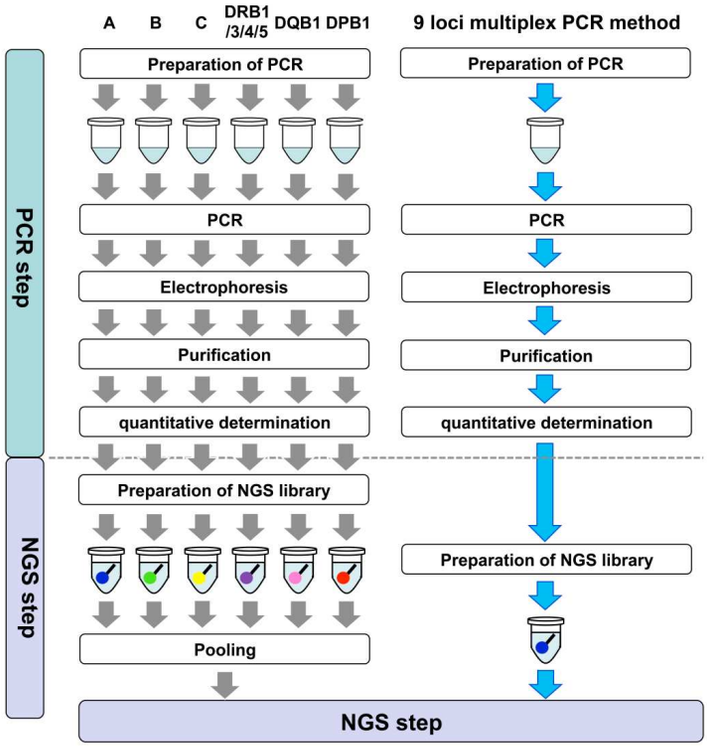

Figure 3. Two different nine loci HLA genotyping procedures at the PCR step.

The multiplex PCR methods contributed greatly to simplifying, accelerating, and reducing costs and the number of reagents for the PCR step that is used to prepare samples and libraries for NGS in the NGS-based HLA genotyping method. The multiplex methods also conserved on the amounts of DNA samples needed to genotype a multiple number of HLA loci. Overall, the multiplex PCR method is a powerful tool for providing precise genotyping data without phase ambiguity, with a strong potential to replace the current routine genotyping methods to find polymorphisms. Commercialized PCR amplification reagents such as NEType (One-Lambda) that are based on multiplex PCR methods will be made available in the near future, whereas those based on the one-locus, one-tube PCR methods (left side of Figure 3) such as the TruSight HLA panel (Illumina) and NGSgo (GenDX) are already available in the market place.

4.2. NGS step

Although the 454 GS FLX was used often in the early stages of development of NGS-based HLA genotyping, the benchtop next-generation sequencers such as the GS Junior system, Ion Torrent PGM system, and the MiSeq system have been used more recently for the development and application of the HLA genotyping methods (Table 2). At the moment, complicated operations such as the preparation of NGS libraries are necessary for each of the different second generation sequencing platforms. However, the NGS companies are attempting to overcome these procedural bottlenecks by simplifying, automating, and speeding-up of the preparatory steps for NGS. For example, a new protocol using Ion Isothermal Amplification Chemistry that enables sequence reads of up to and beyond 500 bp, and Ion Hi-Q™ Sequencing Chemistry that reduces consensus insertion and deletion (indel) errors, including homopolymer errors, might lead to further simplification and cost reduction with higher data quality.

4.3. Allele assignment step

A variety of different allele assignment methods have been developed with some allele assignment software packages such as Assign (CONEXIO), OMIXON Target (OMIXON), and NGSengine (GenDX) commercially available, and others such as TypeStream (Life Technologies) still to be made commercially available in the near future. From our knowledge, Assign and NGSengine only support NGS data obtained from the one-locus, one-tube PCR method, whereas OMIXON Target and TypeStream also support NGS data obtained by the multiplex PCR methods. However, accuracy rates of the assignment methods are not 100% with genotyping errors caused by (1) missing HLA allele sequences, (2) generation of excessive allelic imbalance (ratio of sequence read numbers of allele 1 and allele 2), and (3) interference of HLA-DRB1 genotyping by participation of sequence reads originating from highly homologous HLA-DRB3/4/5 and other HLA-DRB pseudogenes. To avoid the errors raised in point 1, it is necessary to have a full and proper collection of all the HLA allele sequences to achieve precise HLA genotyping. In this regard, a much greater collection of high-quality full-length HLA allele sequences are expected to be obtained by way of international collaborations at the 17th IHIWS meeting in 2017 [53].

http://dx.doi.org/10.5772/61842

4.3.1. In-house Sequence Alignment-Based Assigning Software (SeaBass)

Recently, we developed a new next sequence allele assignment program (Sequence Alignment-Based Assigning Software; SeaBass) to solve the problems previously outlined in points 2 and 3 above. The program includes (1) output of sequence reads, (2) homology search using the Blat program [55] with the "match" variable set to 100% to detect identical exons within the known HLA alleles released from the IMGT-HLA database [7], (3) selection of allele candidates, (4) mapping of the sequence reads to the selected allele candidates as references with the "match" set at 100% using Reference Mapper (Roche), (5) calculation of coverages, and (6) confirmation of the mapping data and allele assignment (Figure 4).

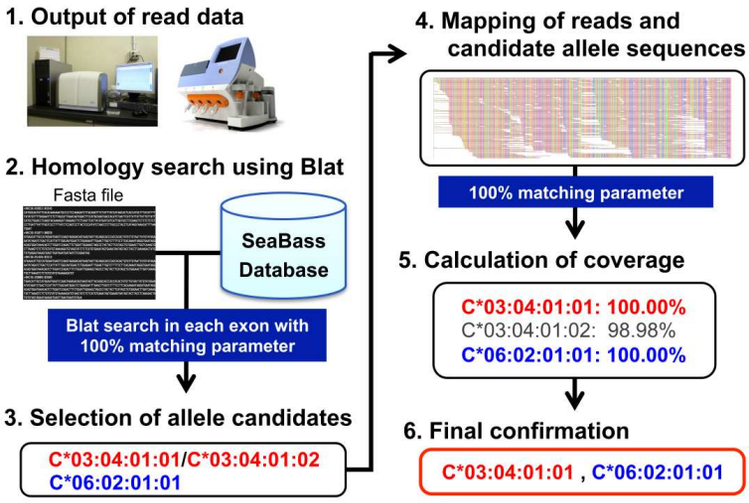

Figure 4. Allele assignment method using the newly developed Sequence Alignment-Based Assigning Software, SeaBass.

The operations from Eqs. (2) to (5) are automatically processed. If a new polymorphism is included in the exon, we can detect its presence at the Blat search stage as shown in Figure 5, and if a new polymorphism is included in the intron, we can detect its presence during the calculation of the coverage and the final confirmation stages (Figure 6).

After the detection of the new polymorphisms, we further confirm them by traditional methods such as Sanger sequencing and subcloning. In addition, we validated the use of the SeaBass assignment methods for three next-generation sequencers, the GS Junior system, the Ion Torrent PGM system, and the MiSeq system. To evaluate the SeaBass program, we used a total of 2414 HLA sequences from all the classical HLA loci that have frequent HLA alleles in Caucasians, African-Europeans, and Japanese, and we obtained an overall accuracy rate of >99.8% and 100% for the Japanese subjects (Table 3).

The accuracy rate was not 100% for HLA-DRB1/3/4/5 and HLA-DPB1 of the non-Japanese subjects because the complete coding sequences have not been determined as yet for some of their HLA-DRB and HLA-DPB1 alleles. Nevertheless, the allele assignment method that we developed for SeaBass appears to be the most accurate and efficient way to detect new and null alleles by NGS.

(A)

Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7
C*01:02:01	C*03:04:01:01	C*03:03:01	C*01:04	C*01:02:11	C*01:02:01	C*01:02:01
C*03:04:01:01	C*03:04:01:02	C*03:04:01:01	C*03:04:01:01	C*03:02:02:01	C*02:02:02:01	C*03:02:02:01
C*03:04:01:02	C*03:04:02	C*03:04:01:02	C*03:04:01:02	C*03:02:02:02	C*03:04:01:01	C*03:04:01:01
C*03:13	C*06:02:01:01	C*03:97	C*06:02:01:01	C*03:02:03	C*03:04:01:02	C*03:04:01:02
C*06:02:01:01	C*06:02:03	C*06:02:01:01	C*16:15:02	C*03:03:01	C*05:01:01:01	C*06:02:01:01
C*06:02:03	C*18:01	C*12:33		C*03:03:02	C*05:01:01:02	C*08:01:01
C*06:02:11	C*18:02			C*03:04:01:01	C*05:01:04	C*08:01:03
C*16:04:01	C*18:03			C*03:04:01:02	C*05:03	C*08:01:05
				C*04:06	C*05:88	C*08:02:01:01
				C*06:02:01:01	C*06:02:01:01	C*08:02:01:02
				C*06:02:03	C*17:01:01:01	C*08:03:01
				C*12:02:01	C*18:01	C*08:04:01
				C*12:03:01:01		C*18:02
				C*14:03		
				C*18:02		

C*03:04:01:01/C*03:04:01:02
C*06:02:01:01

(B-1)

Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7
B*15:18:01	B*15:18:01	B*15:18:01	B*15:18:01	B*15:18:01	B*15:18:01	B*15:18:01
B*44:03:01	B*44:03:01	Not detected	B*44:03:01	B*44:03:01	B*44:03:01	B*44:03:01

(B-2)

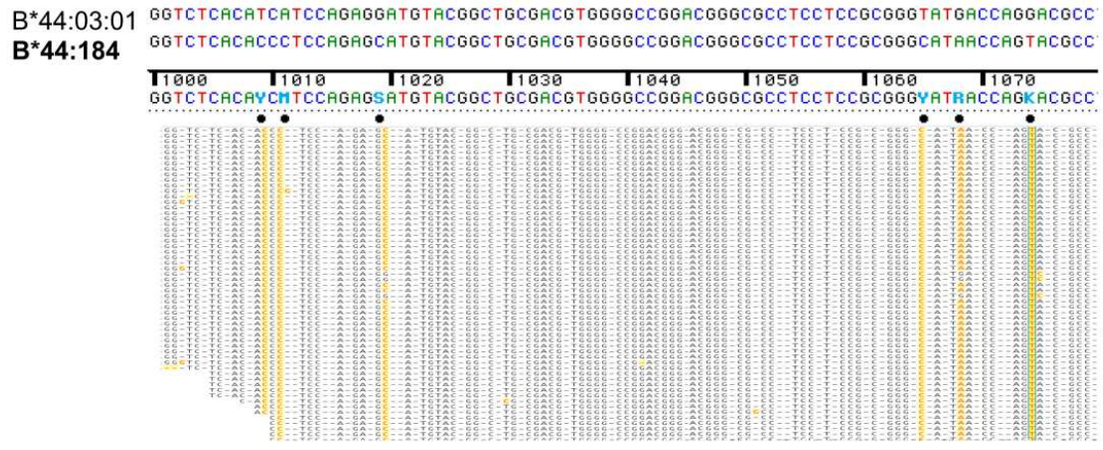

Figure 5. Detailed information concerning selection of allele candidates using the SeaBass computer program. (A) "Extraction of allele candidates" by Blat search. We select allele candidates that are extracted in all of the exons. (B-1) New allele detection. In this example, one allele was called B*15:18:01, but the other allele was called B*44:03:01 excluding the exon 3. (B-2) Confirmation of the new allele by NGS. Mapping of the sequence reads with B*44:03:01 as a reference suggested six nucleotide differences with B*44:03:01 were detected in exon 3. We confirmed the polymorphisms by Sanger sequencing and deposited the sequence to DDBJ and IMGT-HLA database. Now the formal allele name is B*44:184 [94].

http://dx.doi.org/10.5772/61842

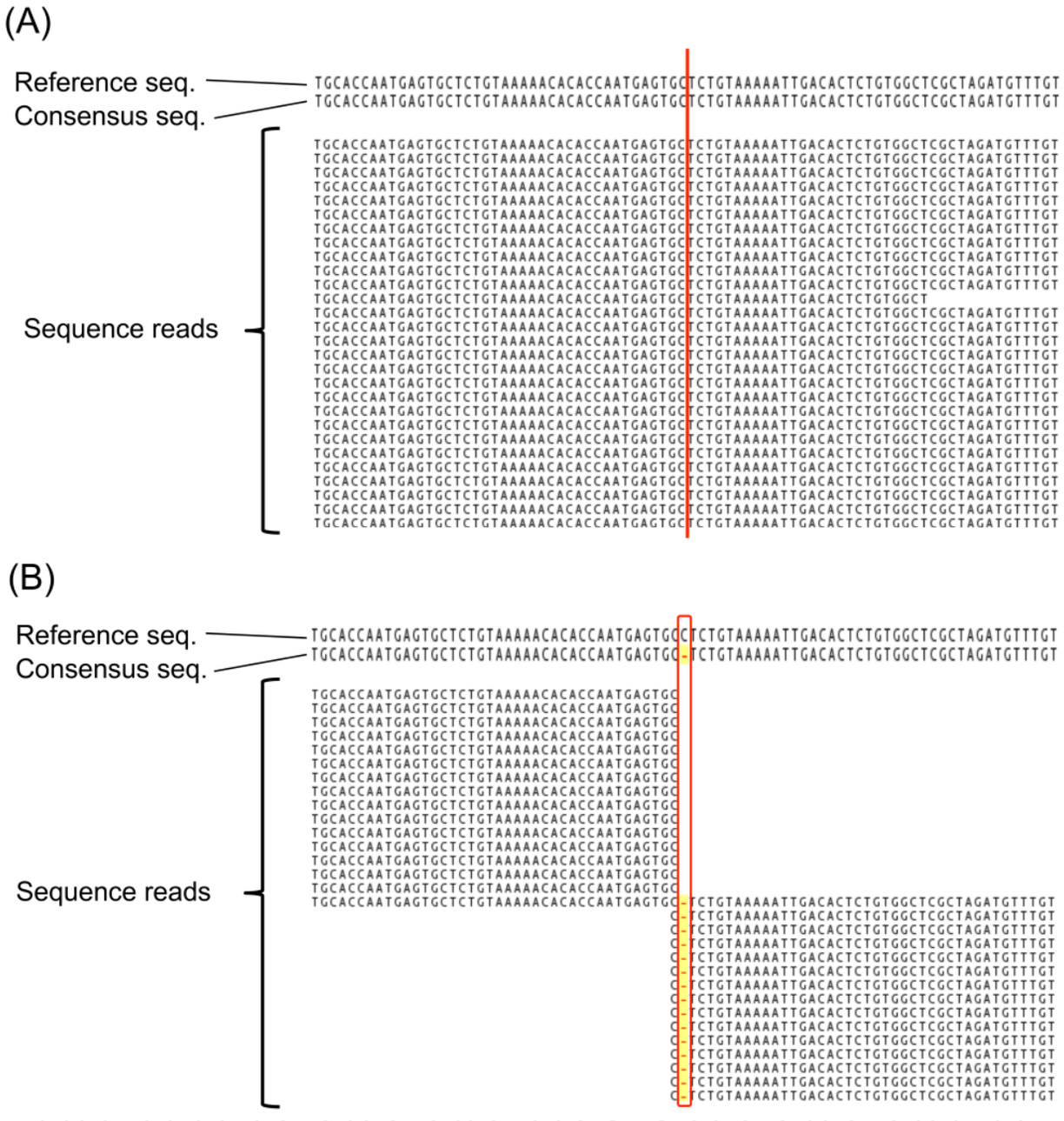

Figure 6. Detection of a new allele during the calculation of the coverage and final confirmation stages in SeaBass. Mapping results of the sequence reads using GS Reference Mapper are shown. (A) In this case, there is no mismatch between the reference and consensus sequence. (B) In this case, there is a mismatch between the reference and consensus sequence (reference: C; consensus: -) indicated by yellow background.

5. NGS-based MHC genotyping methods in nonhuman species

NGS technology provides the opportunity to genotype MHC sequences either by PCR targeted DNA sequencing or by PCR targeted RNA sequencing, that is, by DNA sequencing after converting the RNA samples to cDNA by reverse transcriptase. Usually, one or other of the

Worldwide subject (1916 loci)										
	Total	**A**	**C**	**B**	**DRB345**	**DRB1**	**DQA1**	**DQB1**	**DPA1**	**DPB1**
Locus number	1916	250	250	242	186	239	140	234	140	235
Allele number	3832	500	500	484	372	478	280	468	280	470
Accuracy rate (%)	99.8	**100**	**100**	**100**	99.2	99.6	**100**	**100**	**100**	99.6
Japanese subject (498 loci)										
	Total	**A**	**C**	**B**	**DRB345**	**DRB1**	**DQA1**	**DQB1**	**DPA1**	**DPB1**
Locus number	498	86	80	77	50	68	4	65	4	64
Allele number	996	172	160	154	100	136	8	130	8	128
Accuracy rate (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 3. Evaluation of the SeaBass program

sequencing methods is chosen rather than using both methods on the same samples. In the following sections, we compare the use and limitations of targeted NGS sequencing using DNA or RNA samples for MHC genotyping of MHC class I and class II genes in nonhuman species such as the Filipino cynomolgus macaques.

5.1. Advantage and disadvantage of using DNA and RNA samples for NGS

Table 4 shows a summary of the advantages and disadvantages of using DNA and RNA samples for NGS-based MHC genotyping.

	DNA	**RNA**
Difficulty of sampling	Easy	Difficult
Extraction cost of nucleic acid	Cheap	Expensive
Preparation before PCR	No	RT reaction
Primer location	Both of exons and introns	Exons only
Required sequence read number	Few	Many
Exclusion of pseudogene	Difficult	Easy
Estimation of expression level	Impossible	Possible

Table 4. Advantages and disadvantages of DNA and RNA samples for NGS-based MHC genotyping

The advantages of using DNA samples instead of RNA samples are that (1) the sampling and extraction of the DNA nucleic acids are easier and cheaper than RNA samples, (2) PCR amplification can be perform directly without an additional reaction such as the reverse transcriptase (RT) reaction, (3) design of primers in the exon and intron regions, and (4) fewer read sequences are required for DNA than RNA samples if all alleles are amplified without allelic imbalance. Although many more read sequences are necessary for RNA samples than DNA samples to genotype all the MHC alleles that have different transcription levels, the advantages of using RNA samples for genotyping are that (1) they provide an opportunity to examine MHC gene expression, (2) transcription levels are possible to be estimated for each of MHC alleles from the read sequence depth [56], and (3) only transcribed MHC genes are detected without contamination of PCR products originating from pseudogenes if the primer locations cross over to at least two homologous exons. Thus, the use of RNA samples is thought to be more effective for precise MHC genotyping on duplicated MHC genes that have high similarities among the genes. However, DNA and RNA samples have their own unique advantages and disadvantages for informative NGS-based MHC genotyping and widen the choices for experimentation and data collection.

5.2. Methodology

Table 5 shows a publication list of the MHC genotyping by PCR-based NGS methods in different animal species, and it includes the MHC species name, target gene, PCR method, degree of allele data accumulation, and the allele assignment method.

	Species	**MHC name**	**Animal model or nonmodel type**	**Template**	**Target gene**	**NGS platform**	**Degree of allele data accumulation**	**Allele assignment method**	**Ref.**
Mammal	Rhesus macaque	*Mamu*	Model	RNA	Class I and II	454, Illumina	Relatively rich	Mapping *de novo* assembly	[78, 79]
	Cynomolgus macaque	*Mafa*	Model	RNA	Class I and II	454, Illumina, PacBio	Relatively rich	Mapping *de novo* assembly	[35, 78, 80, 81]
	Pig-tailed macaque	*Mane*	Model	RNA	Class I and II	454, Illumina	Relatively rich	Mapping *de novo* assembly	[78, 82]
	Swine	*SLA*	Model	RNA	Class I	454	Relatively rich	Mapping *de novo* assembly	[56]
	Grey mouse lemur	*Mimu*	Nonmodel	DNA	DRB and DQB	454	Poor	*De novo* assembly	[83]

	Species	MHC name	Animal model or nonmodel type	Template	Target gene	NGS platform	Degree of allele data accumulation	Allele assignment method	Ref.
	Alpine marmots	*Mama*	Nonmodel	DNA	Class I and DRB	454	Poor	*De novo* assembly	[84]
	New Zealand sea lion	*Phho*	Nonmodel	DNA	DRB and DQB	454	Poor	*De novo* assembly	[85]
Avian	Collared flycatcher	*Fial*	Nonmodel	DNA	Class II	454	Poor	*De novo* assembly	[86]
	Great tit	*Pama*	Nonmodel	DNA	Class I	454	Poor	*De novo* assembly	[87]
	House Sparrows	*Pado*	Nonmodel	DNA	Class I	454	Poor	*De novo* assembly	[88]
	Berthelot's pipittawny pipit	*AnbeAnca*	Nonmodel	DNA	Class II	454	Poor	*De novo* assembly	[89]
	New Zealand passerine	*Peph*	Nonmodel	DNA	Class II	PGM	Poor	*De novo* assembly	[90]
	Eurasian Coot	*Fuat*	Nonmodel	DNA	Class II	454	Poor	*De novo* assembly	[91]
Reptile	Ornate dragon lizard	*Ctor*	Nonmodel	DNA	Class I	454	Poor	*De novo* assembly	[92]
Fish	Stickleback fish	*Gaac*	Nonmodel	DNA	Class II	454	Poor	*De novo* assembly	[93]

Table 5. Publication list of MHC genotyping by PCR-based NGS methods in nonhuman species

As discussed previously, for humans, the HLA alleles obtained by next-generation sequencers are mainly assigned by mapping to known allele sequences that are used as the read references because a large number of HLA allele sequences already have been collected in the IMGT-HLA database [7] (Table 2). On the other hand, *de novo* assembly of read sequences and subcloning of PCR products identifies novel allele sequences. Of the nonhuman species, RNA samples tend to be used for MHC genotyping in experimental animals (model animals) such as macaque species and swine, whereas DNA samples are mainly used for MHC genotyping wild (nonmodel) animals because collecting RNA samples from them in their natural environment is more difficult than sampling captured or domesticated experimental animals (Table 5).

5.2.1. MHC genotyping RNA samples collected from Filipino cynomolgus macaques

MHC alleles in humans and experimental animals such as the macaque species and swine are mainly assigned by mapping methods because of the large amount of MHC allele information

http://dx.doi.org/10.5772/61842

already available for them than for most other species. This allele information is collected and released by the IPD-MHC database [57]. When novel alleles are detected, *de novo* assembly of the read sequences and subcloning of PCR products identifies the sequences.

We identified homozygous and heterozygous cynomolgus macaques (Mafa) that have specific Mafa MHC haplotypes by genotyping the MHC of more than 5000 Filipino animals, and we found that they have a smaller number of different Mafa-class I and Mafa-class II alleles than the Indonesian and Vietnamese populations. In this section, we outline the MHC genotyping method using RNA samples and provide some results as an example of the method. Figure 7 shows a comparative genomic map of MHC regions between human and Filipino cynomolgus macaque.

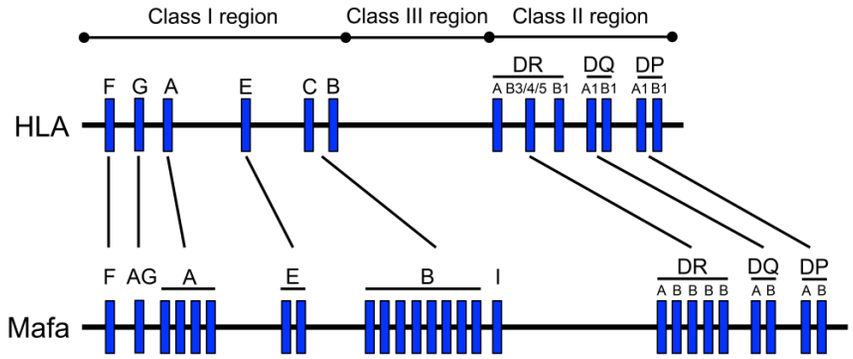

Figure 7. Comparative genomic map of the human (HLA) and the Filipino cynomolgus macaque (Mafa) Class I and Class II transcribed genes.

The MHC class I genomic region has many more Mafa-class I genes than HLA-class I genes generated by gene duplication events, whereas the organization of Mafa-class II genes are well conserved between the two species. Also, there are many Mafa-class I pseudogenes located in the Mafa-class I region. Therefore, we performed MHC genotyping by amplicon sequencing with the Roche GS Junior system using RNA samples from the Filipino cynomolgus macaques to prevent contamination of PCR products originating from the pseudogenes (Figure 8).

The workflow that we used is composed mainly of five steps: (1) RNA extraction and cDNA synthesis, (2) multiplex PCR amplification, (3) pooling of the PCR products, (4) amplicon NGS sequencing, and (5) allele assignment. In step 1, we usually extracted total RNA from the peripheral white blood cell samples using the TRIzol reagent (Invitrogen/Life Technologies/Thermo Fisher Scientific, Carlsbad, CA) and synthesized cDNA by oligo d(T) primer using the ReverTraAce for the reverse transcriptase reaction (TOYOBO, Osaka, Japan) after treatment of the isolated RNA with DNase I (Invitrogen/Life Technologies/Thermo Fisher Scientific, Carlsbad, CA). In step 2, we designed a single Mafa-class I-specific primer set in exon 2 and exon 4 (PCR product size: 514 bp or 517 bp) that could amplify all known Mafa-class I alleles, whereas the Mafa-class II locus-specific primer sets included the polymorphic exon 2 in Mafa-DRB (420 bp), Mafa-DQA1 (435 bp), Mafa-DQB1 (396 bp), Mafa-DPA1 (407 bp), and Mafa-DPB1 (333, 336 or 339 bp) for massively parallel pyrosequencing (Figure 9).

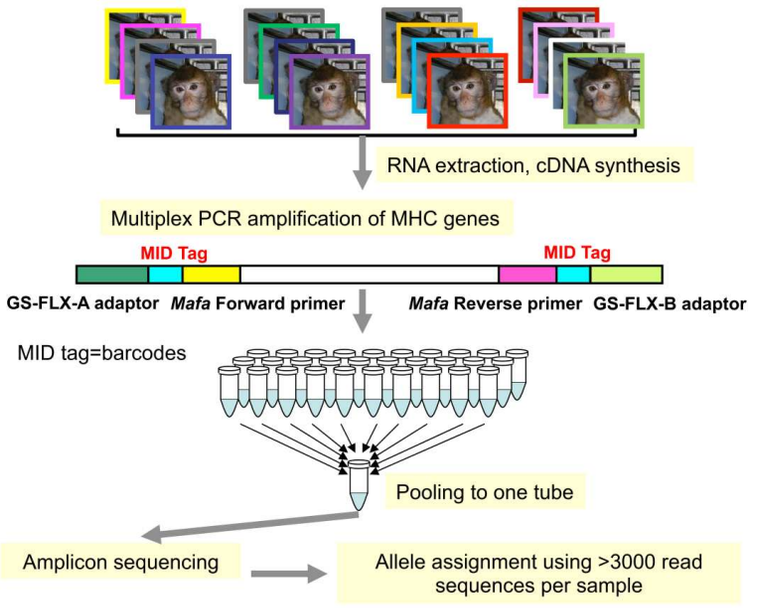

Figure 8. A schematic workflow of the successive steps of the MHC genotyping method by NGS amplicon sequencing for the Filipino cynomolgus macaques.

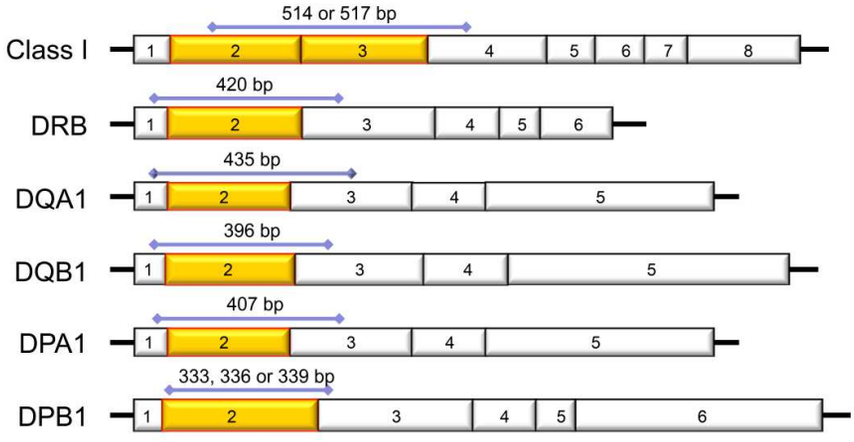

Figure 9. Location of primer sites to amplify Filipino cynomolgus macaque MHC genes. Yellow boxes and blue arrows indicate polymorphic exons and PCR regions, respectively. Numbers indicate exon numbers.

In addition to these primer sets, we also designed 50 different types of fusion primers that contained the 454 titanium adaptor (A in forward and B in reverse primer), 10 bp MID (multiple identifier), and MHC-specific primers (Figure 8). Moreover, we constructed a multiplex PCR method using the primer sets by carefully optimizing primer composition and PCR conditions and by comparing the sequence read data obtained by NGS (Figure 10).

As a result of these primer designs, 51.5%, 13.6%, and 8.6–8.9% of all read sequence numbers were detected in Mafa-class I, Mafa-DRB, and the other Mafa-class II genes, respectively, and we confirmed that the genotypes obtained by the multiplex PCR method were consistent with

http://dx.doi.org/10.5772/61842

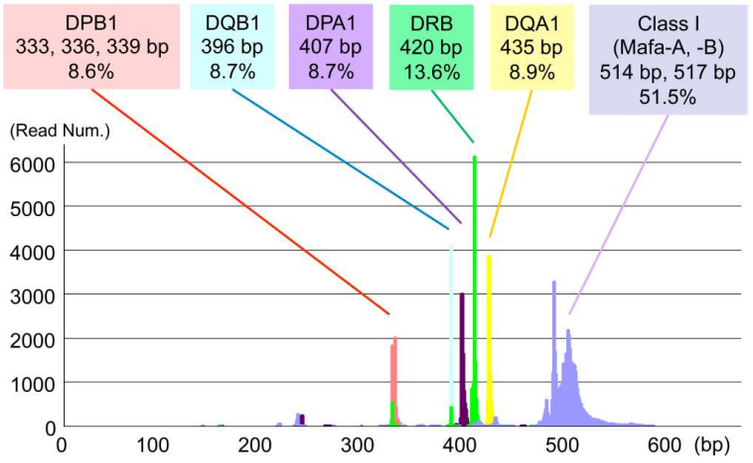

Figure 10. Ratio of read sequence numbers obtained by amplicon sequencing of multiplex PCR products.

our previous uniplex PCR methods. Therefore, the multiplex PCR method greatly simplified the procedures required in preparing the DNA samples for NGS by reducing the time of preparation and the amount and cost of reagents. In the pooling step of the PCR products, we quantified the purified PCR products by the Picogreen assay (Invitrogen) with a Fluoroskan Ascent micro-plate fluorometer (Thermo Fisher Scientific, Waltham, MA), mixed each of the PCR products at equimolar concentrations and then diluted them according to the manufacture's recommendation. In the NGS amplicon sequencing step, we perform emulsion PCR (emPCR) and emulsion-breaking according to the manufacturer's protocol (Roche, Basel, Switzerland). After the emulsion-breaking step, we enriched and counted the beads carrying the single-stranded DNA templates, and deposited them into a PicoTiterPlate to obtain the sequence reads.

A schematic workflow of the allele assignment process as a follow on from Figure 8 is shown in Figure 11.

After the sequencing run, image processing, signal correction, and base calling are performed by the GS Run Processor Ver. 3.0 (Roche) with full processing for shotgun or paired-end filter analysis. Quality-filter sequence reads that are passed by the assembler software (single sff file) are binned according to the MID labels into each separate sequence sff file using the sff file software (Roche). These files are further quality trimmed to remove poor sequence at the end of the reads with quality values (QVs) of less than 20. After separation of the trimmed and MID-labeled sequence reads in each of forward and reverse side read sequences, we independently detect the Mafa-class I and Mafa-class II allele candidates from both sides of the forward and reverse reads by using the BLAT program to match the trimmed and MID labeled sequence reads at 99% and 100% identity while setting the minimum overlap length at 200 and the alignment identity score parameter at 10 against all the known Mafa-class I and Mafa-class II allele sequences released in the IMGT/MHC-NHP database [58]. After the extraction of common allele candidates from both sequencing sides, we finally assign the "real alleles" by confirming nucleotide sequences of the allele candidates using the GS Reference Mapper Ver.

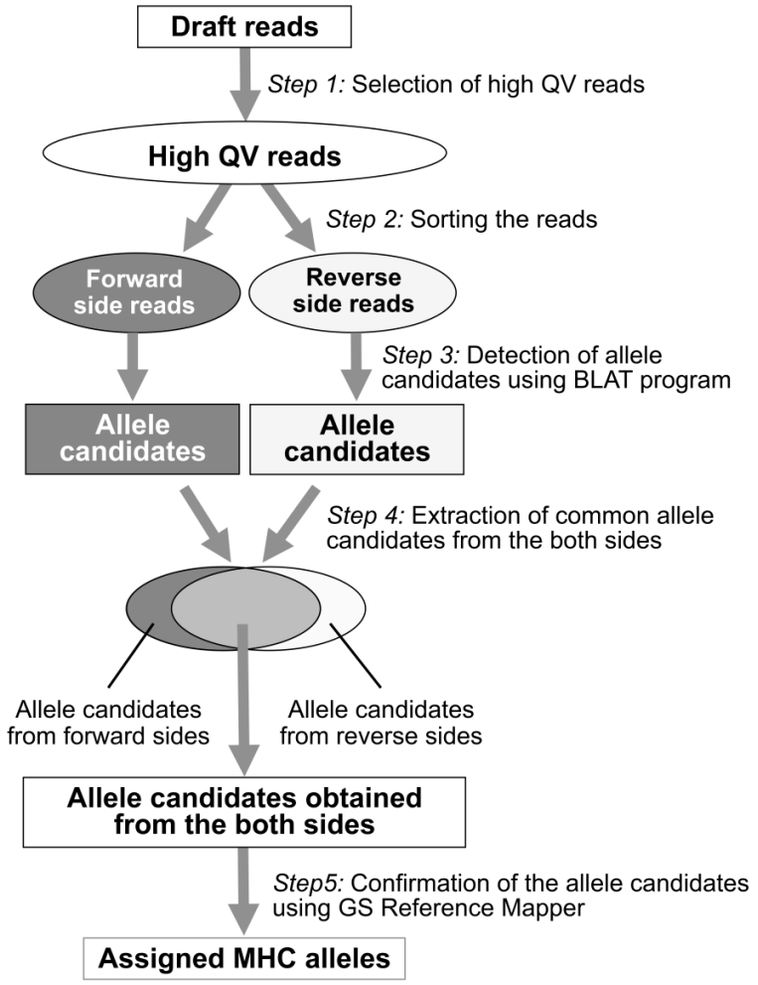

Figure 11. A schematic workflow of the allele assignment process using the SeaBass software.

3.0. To discover novel Mafa-class I sequences, we perform the de novo assembly set to detect >85% matches using the trimmed and MID-binned sequences after converting the outputs to ace files for the Sequencher Ver. 5.01 DNA sequence assembly software (Gene Code Co., Ann Arbor, MI). We then use the defined consensus sequence obtained from the de novo assembly as a reference sequence to identify and map the correct allele sequences. Using this process, we genotyped a set of 400 unrelated animals by the Sanger sequencing method and high resolution pyrosequencing and identified 190 different alleles, 28 at Mafa-A, 54 at Mafa-B, 12 at Mafa-I, 11 at Mafa-E, 7 at Mafa-F, 34 at Mafa-DRB, 13 at Mafa-DQA1, 13 at Mafa-DQB1, 9 at Mafa-DPA1, and 9 at Mafa-DPB1 alleles [35, 59].

On the basis of our large-scale project to genotype the MHC of 5000 Filipino cynomolgus macaques by NGS, we so far have detected 15 different types of Mafa haplotypes (HT1~HT15) in 45 homozygous animals. These Mafa homozygous animals provided the basis to efficiently estimate other Mafa haplotypes. For example, we estimated a variety of Mafa-A, Mafa-B/I, Mafa-E, and Mafa-class II (Mafa-DRB, Mafa-DQA1, Mafa-DQB1, Mafa-DPA1, and Mafa-DPB1) haplotypes by comparing the homozygous animals with heterozygous animals that

carry the identical Mafa-class I and Mafa-class II alleles in the homozygous animals. In addition, we estimated the Mafa haplotypes and haplotype frequencies by the PHASE 2.1.1 program [60] using the allele data obtained by amplicon sequencing. From these procedures, we estimated a total of 84 Mafa-class I and 18 Mafa-class II haplotypes. Of the 15 different Mafa HT haplotypes, the haplotype frequencies of HT1, HT2, HT4, and HT8 were the highest. Of them, HT1 and HT8 have entirely different Mafa alleles, whereas HT2 and HT4 are thought to be recombinants of HT1 and HT8 (Figure 12).

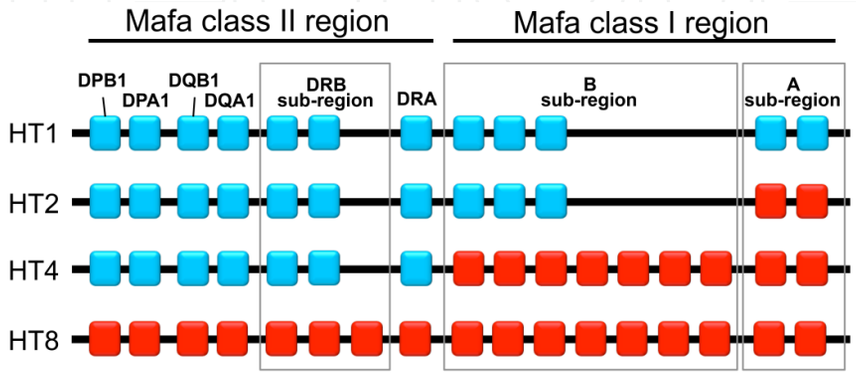

Figure 12. Gene composition of representative Mafa MHC haplotypes HT1 and HT8 and their recombinants HT2 and HT4.

Namely, the Mafa-A allele in HT2 is identical to that in HT8, whereas HT2 also has alleles at other loci that are identical with those in HT1. Similarly, HT4 has alleles in Mafa-class I loci that are identical with those in HT8, and alleles in the Mafa-class II loci that are identical with those in HT1. Therefore, Mafa homozygous animals with known haplotypes such as H1 and H2 are important for biomedical research, such as the transplantation outcomes of induced pluripotent stem (iPS) cells (Figure 13) because such studies are undertaken on animals with a defined genetic background and relatively well-characterized MHC haplotypes that might regulate the adaptive immune system in different ways and efficiencies.

5.2.2. MHC genotyping using DNA samples of wild animals

At this time in the development of MHC genotyping by NGS, it is difficult to apply the RNA-sequencing mapping method to accurately genotype the MHC of wild animals using known allele sequences as references. This is because the present allele information is relatively poor for most of them (Table 5). Therefore, MHC genotyping of wild animals or poorly studied species by NGS is based on *de novo* assembly of DNA sequences. In this case, the definition of "real alleles" and "artifact alleles" is important because NGS errors such as monostretch sequences are frequently observed in the assembled consensus sequences. Some of the allele assignment approaches based on *de novo* assembly that have been published include the allele validation threshold (AVT) method [61], clustering method [62–64], and the relative sequencing depth modeling methods [65]. These methods suppose that the contigs that have a sequence depth greater than the threshold level are the "real alleles," and they are determined by

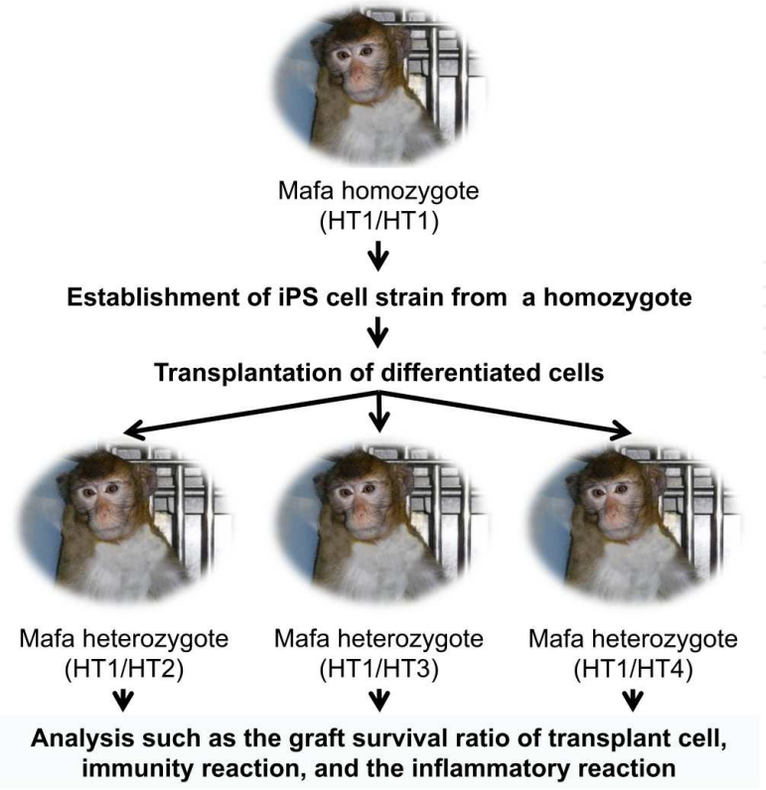

Figure 13. Application of Mafa homozygous and heterozygous animals for nonclinical trials of induced pluripotent stem (iPS) cells.

statistical calculation of the threshold using the sequence depth values of all contigs obtained in *de novo* assembly. Therefore, the detection of exact or "real" alleles depends largely on the setting of the threshold level and the quality of the sequence reads [65]. To enable the correct setting of the threshold level, it is important to use primers that can amplify all alleles of the target locus or loci without allelic imbalance. Furthermore, additional considerations such as repeating independent NGS experiments at least three times and detecting identical allele sequences in at least two animals are necessary to distinguish between real and artifactual alleles.

6. Conclusion

Genotyping the polymorphisms of MHC genes using targeted NGS technologies has been developed for humans and some nonhuman species to replace the use of other more cumbersome and less accurate procedures. We found that targeted NGS of DNA or RNA samples is feasible, productive, and generates high-quality MHC allele information from a large number of samples not easily achievable by other genotyping methods. We used second-generation sequencing protocols to target the DNA region and RNA subsets of interest in our NGS studies. It is likely that the longer sequence reads produced by third-generation platforms such as the

Pacific Biosciences single-molecule real-time sequencing or the Oxford nanopore sequencing platform will enable and improve the task of MHC sequence phasing and haplotyping, although this has yet to be demonstrated and proved to be advantageous and more economical. Continued allele data collection for different species, improvements to the reagents, protocols, and data analysis tools also are likely to simplify procedures and lower the costs of generating sequencing data in future. Most species have numerous highly polymorphic MHC loci; hence, the many benefits of using NGS technologies are likely, in the near future, to replace many of the traditional genotyping methods for the investigation of human and animal MHC genes and their role in evolutionary biology, ecology, population genetics, disease, and transplantation.

Author details

Takashi Shiina[1*], Shingo Suzuki[1] and Jerzy K. Kulski[1,2]

*Address all correspondence to: tshiina@is.icc.u-tokai.ac.jp

1 Division of Basic Medical Science and Molecular Medicine, Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa, Japan

2 Centre for Forensic Science, The University of Western Australia, Nedlands, WA, Australia

References

[1] Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunological Reviews. 2002;190:95–122. DOI: 10.1034/j.1600-065X.2002.19008.x.

[2] Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens. 2004;64(6):631–49. DOI: 10.1111/j.1399-0039.2004.00327.x.

[3] Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. Journal of Human Genetics. 2009;54(1):15–39. DOI: 10.1038/jhg.2008.5.

[4] Claas FH, Duquesnoy RJ. The polymorphic alloimmune response in clinical transplantation. Current Opinion in Immunology. 2008;20(5):566–7. DOI: 10.1016/j.coi.2008.08.001.

[5] Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. Yonsei Medical Journal. 2007;48(1):11–23. DOI: 10.3349/ymj.2007.48.1.11.

[6] Kulski JK, Inoko H. Major histocompatibility complex (MHC) genes. In: Cooper DN, editor. Nature Encyclopedia of the Human Genome. London: Nature Publishing Group; 2003. pp. 778–85.

[7] IMGT/HLA Database. Available from: http://www.ebi.ac.uk/imgt/hla/.

[8] Zinkernagel RM, Doherty PC. The discovery of MHC restriction. Immunology Today. 1997;18(1):14–7. DOI: http://dx.doi.org/10.1016/S0167-5699(97)80008-4.

[9] Sasazuki T, Juji T, Morishima Y, Kinukawa N, Kashiwabara H, Inoko H, et al. Effect of matching of class I HLA alleles on clinical outcome after transplantation of hematopoietic stem cells from an unrelated donor. Japan Marrow Donor Program. New England Journal of Medicine. 1998;339(17):1177–85. DOI: 10.1056/NEJM199810223391701.

[10] International MHC, Autoimmunity Genetics N, Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(44):18680–5. DOI: 10.1073/pnas.0909307106.

[11] Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genetics. 2011;7(8):e1002254. DOI: 10.1371/journal.pgen.1002254.

[12] International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214–9. DOI: 10.1038/nature10251.

[13] Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature Genetics. 2012;44(3):291–6.DOI: 10.1038/ng.1076.

[14] International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551–7. DOI: 10.1126/science.1195271.

[15] Ziegler A, Kentenich H, Uchanska-Ziegler B. Female choice and the MHC. Trends in Immunology. 2005;26(9):496–502. DOI: 10.1016/j.it.2005.07.003.

[16] McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, et al. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. New England Journal of Medicine. 2011;364(12):1134–43. DOI: 10.1056/NEJMoa1013297.

[17] Illing PT, Vivian JP, Dudek NL, Kostenko L, Chen Z, Bharadwaj M, et al. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature. 2012;486(7404):554–8. DOI: 10.1038/nature11147.

[18] Xiao BG, Link H. Immune regulation within the central nervous system. Journal of the Neurological Sciences. 1998;157(1):1–12. DOI: http://dx.doi.org/10.1016/S0022-510X(98)00049-5.

[19] Huh GS, Boulanger LM, Du H, Riquelme PA, Brotz TM, Shatz CJ. Functional requirement for class I MHC in CNS development and plasticity. Science. 2000;290(5499):2155–9. DOI:10.1126/science.290.5499.2155.

[20] Boulanger LM, Shatz CJ. Immune signalling in neural development, synaptic plasticity and disease. Nature Reviews Neuroscience. 2004;5(7):521–31. DOI: 10.1038/nrn1428.

[21] Cullheim S, Thams S. The microglial networks of the brain and their role in neuronal network plasticity after lesion. Brain Research Reviews. 2007;55(1):89–96. DOI: 10.1016/j.brainresrev.2007.03.012.

[22] Ohtsuka M, Inoko H, Kulski JK, Yoshimura S. Major histocompatibility complex (Mhc) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. BMC Genomics. 2008;9:178. DOI: 10.1186/1471-2164-9-178.

[23] Matsuo R, Asada A, Fujitani K, Inokuchi K. LIRF, a gene induced during hippocampal long-term potentiation as an immediate-early gene, encodes a novel RING finger protein. Biochemical and Biophysical Research Communications. 2001;289(2):479–84. DOI: 10.1006/bbrc.2001.5975.

[24] Patino-Lopez G, Hevezi P, Lee J, Willhite D, Verge GM, Lechner SM, et al. Human class-I restricted T cell associated molecule is highly expressed in the cerebellum and is a marker for activated NKT and CD8+ T lymphocytes. Journal of Neuroimmunology. 2006;171(1–2):145–55. DOI: 10.1016/j.jneuroim.2005.09.017.

[25] Goddard CA, Butts DA, Shatz CJ. Regulation of CNS synapses by neuronal MHC class I. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(16):6828–33. DOI: 10.1073/pnas.0702023104.

[26] Tonelli LH, Postolache TT, Sternberg EM. Inflammatory genes and neural activity: involvement of immune genes in synaptic function and behavior. Frontiers in Bioscience. 2005;10:675–80. DOI: http://dx.doi.org/10.2741/1562.

[27] Lengen C, Regard M, Joller H, Landis T, Lalive P. Anomalous brain dominance and the immune system: do left-handers have specific immunological patterns? Brain and Cognition. 2009;69(1):188–93. DOI: 10.1016/j.bandc.2008.07.008.

[28] O'Keefe GM, Nguyen VT, Benveniste EN. Regulation and function of class II major histocompatibility complex, CD40, and B7 expression in macrophages and microglia:

implications in neurological diseases. Journal of Neurovirology. 2002;8(6):496–512.DOI: 10.1080/13550280290100941.

[29] Raha-Chowdhury R, Andrews SR, Gruen JR. CAT 53: a protein phosphatase 1 nuclear targeting subunit encoded in the MHC Class I region strongly expressed in regions of the brain involved in memory, learning, and Alzheimer's disease. Brain Research. Molecular Brain Research. 2005;138(1):70–83. DOI: 10.1016/j.molbrainres.2005.04.001.

[30] Cohly HH, Panja A. Immunological findings in autism. International Review of Neurobiology. 2005;71:317–41. DOI: 10.1016/S0074-7742(05)71013-8.

[31] Bailey SL, Carpentier PA, McMahon EJ, Begolka WS, Miller SD. Innate and adaptive immune responses of the central nervous system. Critical Reviews in Immunology. 2006;26(2):149–88. DOI: 10.1615/CritRevImmunol.v26.i2.40.

[32] McElroy JP, Oksenberg JR. Multiple sclerosis genetics. Current Topics in Microbiology and Immunology. 2008;318:45–72.

[33] Santamaria P, Lindstrom AL, Boyce-Jacino MT, Myster SH, Barbosa JJ, Faras AJ, et al. HLA class I sequence-based typing. Human Immunology. 1993;37(1):39–50. DOI: 10.1016/0198-8859(93)90141-M.

[34] Hutchison CA, 3rd. DNA sequencing: bench to bedside and beyond. Nucleic acids Research. 2007;35(18):6227–37. DOI: 10.1093/nar/gkm688.

[35] Shiina T, Yamada Y, Aarnink A, Suzuki S, Masuya A, Ito S, et al. Discovery of novel MHC-class I alleles and haplotypes in Filipino cynomolgus macaques (*Macaca fascicularis*) by pyrosequencing and Sanger sequencing : Mafa-class I polymorphism. Immunogenetics. 2015;67(10):563–78. DOI: 10.1007/s00251-015-0867-9.

[36] Ota M, Fukushima H, Kulski JK, Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. Nature Protocols. 2007;2(11):2857–64. DOI: 10.1038/nprot.2007.407.

[37] Arguello JR, Madrigal JA. HLA typing by Reference Strand Mediated Conformation Analysis (RSCA). Reviews in Immunogenetics. 1999;1(2):209–19.

[38] Saiki RK, Walsh PS, Levenson CH, Erlich HA. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(16):6230–4.

[39] Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue Antigens. 1992;39(5):225–35. DOI: 10.1111/j.1399-0039.1992.tb01940.x.

[40] Sheldon S, Poulton K. HLA typing and its influence on organ transplantation. Methods in Molecular Biology. 2006;333:157–74. DOI: 10.1385/1-59745-049-9:157.

[41] Mahdi BM. A glow of HLA typing in organ transplantation. Clinical and Translational Medicine. 2013;2(1):6. DOI: 10.1385/1-59745-049-9:157.

[42] Petersdorf EW. Optimal HLA matching in hematopoietic cell transplantation. Current Opinion in Immunology. 2008;20(5):588–93. DOI: 10.1016/j.coi.2008.06.014.

[43] Erlich HA, Opelz G, Hansen J. HLA DNA typing and transplantation. Immunity. 2001;14(4):347–56. DOI: http://dx.doi.org/10.1016/S1074-7613(01)00115-7.

[44] Fernandez Vina MA, Hollenbach JA, Lyke KE, Sztein MB, Maiers M, Klitz W, et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. Philosophical Transactions of the Royal Society of London Series: B. Biological Sciences. 2012;367(1590):820–9. DOI: 10.1098/rstb.2011.0320.

[45] Gourraud PA, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M, et al. HLA diversity in the 1000 genomes dataset. PLoS One. 2014;9(7):e97282. DOI: 10.1371/journal.pone.0097282.

[46] Nakaoka H, Mitsunaga S, Hosomichi K, Shyh-Yuh L, Sawamoto T, Fujiwara T, et al. Detection of ancestry informative HLA alleles confirms the admixed origins of Japanese population. PLoS One. 2013;8(4):e60793. DOI: 10.1371/journal.pone.0060793.

[47] Grubic Z, Stingl K, Martinez N, Palfi B, Brkljacic-Kerhin V, Kastelan A. STR and HLA analysis in paternity testing. International Congress Series. 2004;1261:535–7. DOI: 10.1016/S0531-5131(03)01654-6.

[48] Itoh Y, Mizuki N, Shimada T, Azuma F, Itakura M, Kashiwase K, et al. High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. Immunogenetics. 2005;57(10):717–29. DOI: 10.1007/s00251-005-0048-3.

[49] Itoh Y, Inoko H, Kulski JK, Sasaki S, Meguro A, Takiyama N, et al. Four-digit allele genotyping of the HLA-A and HLA-B genes in Japanese patients with Behcet's disease by a PCR-SSOP-Luminex method. Tissue Antigens. 2006;67(5):390–4. DOI: 10.1111/j.1399-0039.2006.00586.x.

[50] Adams SD, Barracchini KC, Chen D, Robbins F, Wang L, Larsen P, et al. Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification. Journal of Translational Medicine. 2004;2(1):30. DOI: 10.1186/1479-5876-2-30.

[51] Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Human Immunology. 2010;71(10):1033–42. DOI: 10.1016/j.humimm.2010.06.016.

[52] Kulski JK, Suzuki S, Ozaki Y, Mitsunaga S, Inoko H, Shiina T. In phase HLA genotyping by next generation sequencing—a comparison between two massively paral-

lel sequencing bench-top Systems, the Roche GS Junior and Ion Torrent PGM. In: Xi Y, editor. HLA and Associated Important Diseases. Croatia: Intech; 2014. p. 141–81.

[53] 17th International HLA and Immunogenetics Workshop (IHIWS) in 2017. Available from: http://ihiws.org/ngs-of-full-length-hla-genes/.

[54] Ozaki Y, Suzuki S, Kashiwase K, Shigenari A, Okudaira Y, Ito S, et al. Cost-efficient multiplex PCR for routine genotyping of up to nine classical HLA loci in a single analytical run of multiple samples by next generation sequencing. BMC Genomics. 2015;16:318. DOI: 10.1186/s12864-015-1514-4.

[55] Blat program. Available from: http://genome.ucsc.edu/

[56] Kita YF, Ando A, Tanaka K, Suzuki S, Ozaki Y, Uenishi H, et al. Application of high-resolution, massively parallel pyrosequencing for estimation of haplotypes and gene expression levels of swine leukocyte antigen (SLA) class I genes. Immunogenetics. 2012;64(3):187–99. DOI: 10.1007/s00251-011-0572-2.

[57] IPD-MHC database. Available from: http://www.ebi.ac.uk/ipd/mhc/.

[58] Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG. IPD—the Immuno Polymorphism Database. Nucleic Acids Research. 2013;41(Database issue):D1234–40. DOI: 10.1093/nar/gks1140.

[59] Blancher A, Aarnink A, Yamada Y, Tanaka K, Yamanaka H, Shiina T. Study of MHC class II region polymorphism in the Filipino cynomolgus macaque population. Immunogenetics. 2014;66(4):219–30. DOI: 10.1007/s00251-014-0764-7.

[60] Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics. 2001;68(4): 978–89. DOI: 10.1086/319501.

[61] Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichon M, Radwan J. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. BMC Evolutionary Biology. 2010;10:395.DOI: 10.1186/1471-2148-10-395.

[62] Sommer S, Courtiol A, Mazzoni CJ. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics. 2013;14:542. DOI: 10.1186/1471-2164-14-542.

[63] Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire PA, et al. Nonparallelism in MHCIIbeta diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. Molecular Ecology. 2013;22(14):3833–49. DOI: 10.1111/mec.12358.

[64] Lamaze FC, Pavey SA, Normandeau E, Roy G, Garant D, Bernatchez L. Neutral and selective processes shape MHC gene diversity and expression in stocked brook charr

populations (*Salvelinus fontinalis*). Molecular Ecology. 2014;23(7):1730–48. DOI: 10.1111/mec.12684.

[65] Lighten J, van Oosterhout C, Bentzen P. Critical review of NGS analyses for de novo genotyping multigene families. Molecular Ecology. 2014;23(16):3957–72. DOI: 10.1111/mec.12843.

[66] Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens. 2009;74(5):393–403. DOI: 10.1111/j.1399-0039.2009.01345.x.

[67] Gabriel C, Danzer M, Hackl C, Kopal G, Hufnagl P, Hofer K, et al. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. Human Immunology. 2009;70(11):960–4. DOI: 10.1016/j.humimm.2009.08.009.

[68] Holcomb CL, Hoglund B, Anderson MW, Blake LA, Bohme I, Egholm M, et al. A multi-site study using high-resolution HLA genotyping by next generation sequencing. Tissue Antigens. 2011;77(3):206–17. DOI: 10.1111/j.1399-0039.2010.01606.x.

[69] Erlich RL, Jia X, Anderson S, Banks E, Gao X, Carrington M, et al. Next-generation sequencing for HLA typing of class I loci. BMC Genomics. 2011;12:42. DOI: 10.1111/j.1399-0039.2010.01606.x.

[70] Shiina T, Suzuki S, Ozaki Y, Taira H, Kikkawa E, Shigenari A, et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. Tissue Antigens. 2012;80(4):305–16. DOI: 10.1111/j.1399-0039.2012.01941.x.

[71] Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(22):8676–81. DOI: 10.1073/pnas.1206614109.

[72] Moonsamy PV, Williams T, Bonella P, Holcomb CL, Hoglund BN, Hillman G, et al. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array System for simplified amplicon library preparation. Tissue Antigens. 2013;81(3):141–9. DOI: 10.1111/tan.12071.

[73] Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics. 2013;14:355. DOI: 10.1186/1471-2164-14-355.

[74] Ozaki Y, Suzuki S, Shigenari A, Okudaira Y, Kikkawa E, Oka A, et al. HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput sequencing. Tissue Antigens. 2014;83(1):10–6. DOI: 10.1111/tan.12258.

[75] Lange V, Bohme I, Hofmann J, Lang K, Sauter J, Schone B, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics. 2014;15:63.DOI: 10.1186/1471-2164-15-63.

[76] Smith AG, Pyo CW, Nelson W, Gow E, Wang R, Shen S, et al. Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors. Human Immunology. 2014;75(10):1040–6. DOI: 10.1016/j.humimm.2014.08.206.

[77] Chang CJ, Chen PL, Yang WS, Chao KM. A fault-tolerant method for HLA typing with PacBio data. BMC Bioinformatics. 2014;15:296. DOI: 10.1186/1471-2105-15-296.

[78] Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, et al. Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nature Medicine. 2009;15(11):1322–6. DOI: 10.1038/nm.2038.

[79] Dudley DM, Karl JA, Creager HM, Bohn PS, Wiseman RW, O'Connor DH. Full-length novel MHC class I allele discovery by next-generation sequencing: two platforms are better than one. Immunogenetics. 2014;66(1):15–24. DOI: 10.1007/s00251-013-0744-3.

[80] Budde ML, Wiseman RW, Karl JA, Hanczaruk B, Simen BB, O'Connor DH. Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing. Immunogenetics. 2010;62(11–12):773–80. DOI: 10.1007/s00251-010-0481-9.

[81] Westbrook CJ, Karl JA, Wiseman RW, Mate S, Koroleva G, Garcia K, et al. No assembly required: full-length MHC class I allele discovery by PacBio circular consensus sequencing. Human Immunology. 2015. DOI: 10.1016/j.humimm.2015.03.022.

[82] O'Leary CE, Wiseman RW, Karl JA, Bimber BN, Lank SM, Tuscher JJ, et al. Identification of novel MHC class I sequences in pig-tailed macaques by amplicon pyrosequencing and full-length cDNA cloning and sequencing. Immunogenetics. 2009;61(10):689–701. DOI: 10.1007/s00251-009-0397-4.

[83] Huchard E, Albrecht C, Schliehe-Diecks S, Baniel A, Roos C, Kappeler PM, et al. Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing. Immunogenetics. 2012;64(12):895–913. DOI: 10.1007/s00251-012-0649-6.

[84] Ferrandiz-Rovira M, Bigot T, Allaine D, Callait-Cardinal MP, Cohas A. Large-scale genotyping of highly polymorphic loci by next-generation sequencing: how to overcome the challenges to reliably genotype individuals? Heredity. 2015;114(5):485–93. DOI: 10.1038/hdy.2015.13.

[85] Osborne AJ, Zavodna M, Chilvers BL, Robertson BC, Negro SS, Kennedy MA, et al. Extensive variation at MHC DRB in the New Zealand sea lion (Phocarctos hookeri) provides evidence for balancing selection. Heredity. 2013;111(1):44–56. DOI: 10.1038/hdy.2013.18.

[86] Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichon M, Radwan J. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. BMC Evolutionary Biology. 2010;10:395. DOI: 10.1186/1471-2148-10-395.

[87] Sepil I, Moghadam HK, Huchard E, Sheldon BC. Characterization and 454 pyrosequencing of major histocompatibility complex class I genes in the great tit reveal complexity in a passerine system. BMC Evolutionary Biology. 2012;12:68. DOI: 10.1186/1471-2148-12-68.

[88] Karlsson M, Westerdahl H. Characteristics of MHC class I genes in house sparrows *Passer domesticus* as revealed by long cDNA transcripts and amplicon sequencing. Journal of Molecular Evolution. 2013;77(1–2):8–21. DOI: 10.1007/s00239-013-9575-y.

[89] Gonzalez-Quevedo C, Phillips KP, Spurgin LG, Richardson DS. 454 screening of individual MHC variation in an endemic island passerine. Immunogenetics. 2015;67(3): 149–62. DOI: 10.1007/s00251-014-0822-1.

[90] Sutton JT, Robertson BC, Jamieson IG. MHC variation reflects the bottleneck histories of New Zealand passerines. Molecular Ecology. 2015;24(2):362–73. DOI: 10.1111/mec.13039.

[91] Alcaide M, Munoz J, Martinez-de la Puente J, Soriguer R, Figuerola J. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian coot *Fulica atra* (Aves: Rallidae). Ecology and Evolution. 2014;4(6):688–98. DOI: 10.1002/ece3.974.

[92] Radwan J, Kuduk K, Levy E, LeBas N, Babik W. Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. Molecular Ecology. 2014;23(24):5966–78. DOI: 10.1111/mec.12984.

[93] Stutz WE, Bolnick DI. Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. PLoS One. 2014;9(7):e100587. DOI: 10.1371/journal.pone.0100587.

[94] Wada A, Shiina T, Michino J, Yasumura S, Sugiyama T. A novel HLA-B allele, HLA-B*44:184, identified by super high-resolution single-molecule sequence-based typing in a Japanese individual. Tissue Antigens. 2014;83(3):198–9. DOI: 10.1111/tan.12284.

IntechOpen