Published in final edited form as:

Methods. 2009 September ; 49(1): 11–17. doi:10.1016/j.ymeth.2009.05.002.

Cost-effective sequence-based nonhuman primate MHC class I genotyping from RNA

Julie A. Karl^a, Roger W. Wiseman^a, and David H. O'Connor^{a,b,*}

^aWisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

^bDepartment of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA

Abstract

The major histocompatibility complex (MHC) class I region of nonhuman primates (NHPs) is highly polymorphic and has undergone complex segmental duplications such that the number of expressed genes differs between individuals. In addition the relative abundance of transcripts varies more than 100-fold between NHP class I genes. This unparalleled complexity makes rapid, efficient class I genotyping difficult for NHPs. The 'gold standard' of cDNA library construction, screening and sequencing is both costly and labor-intensive. Several rapid genotyping methods have been utilized, but all require some degree of prior sequence knowledge. Here, we describe a method for sequence-based MHC class I genotyping which reduces cost by (1) pooling molecularly barcoded class I cDNA-PCR amplicons for cloning and (2) targeting sequencing of a region of concentrated polymorphism spanning the two exons encoding the peptide binding domain. This method can efficiently genotype both known and novel MHC class I alleles. In addition, full-length cDNA amplicons with novel sequences can be resequenced in their entirety to expand the repertoire of characterized MHC class I sequences for NHPs.

Keywords

nonhuman primates; major histocompatibility complex; class I; sequence-based; genotyping

1. Introduction

The major histocompatibility complex (MHC) class I genes play a vital role in the adaptive immune response mounted against viral and bacterial infections [1]. In macaque monkeys, common models for infectious disease and transplantation research, the MHC class I genes are highly polymorphic and have undergone segmental duplication, leading to high allelic diversity and variable numbers of MHC transcripts in each animal [2]. Variation between MHC class I alleles is concentrated within the peptide binding domain. Thus, the repertoire of MHC class I alleles expressed by a macaque determines the specificity of cytotoxic T lymphocyte responses as different alleles have the ability to bind and present different peptides [3-5].

*Corresponding author: David H. O'Connor, University of Wisconsin-Madison, 555 Science Drive, Madison, Wisconsin 53711, Phone: 608-890-0845, Fax: 608-265-8084, Email: doconnor@primate.wisc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Therefore, it is important to determine the MHC class I genotypes of macaques utilized in biomedical research protocols.

Various methods have been applied to macaque MHC genotyping including denaturing gradient gel electrophoresis (DGGE), reference strand conformational analysis (RSCA), sequence-specific PCR (PCR-SSP), and microsatellite analysis. DGGE and RSCA can both resolve multiple MHC alleles in a single run and can potentially identify both known and novel alleles present in an animal. However, analysis with these techniques is complicated by co-migration of heteroduplexes and both methods require clones of known sequence as exemplars for mobility for positive identification [6,7]. PCR-SSP assays are difficult to multiplex and only available for a very small fraction of nearly 1000 nonhuman primate (NHP) class I alleles identified to date (primarily alleles common among Indian-origin rhesus macaques) [5,8,9]. Development of PCR-SSP assays for additional alleles is dependent on identifying one or more single nucleotide polymorphisms that differentiate an allele of interest from all other alleles found in a particular species. Microsatellite analysis can easily and economically identify MHC alleles in a single run and can potentially identify both known and novel alleles present in an animal. However, analysis with these techniques is complicated by co-migration of heteroduplexes and both methods require clones of known sequence as exemplars for mobility for positive identification [6,7]. PCR-SSP assays are difficult to multiplex and only available for a very small fraction of nearly 1000 nonhuman primate (NHP) class I alleles identified to date (primarily alleles common among Indian-origin rhesus macaques) [5,8,9]. Development of PCR-SSP assays for additional alleles is dependent on identifying one or more single nucleotide polymorphisms that differentiate an allele of interest from all other alleles found in a particular species. Microsatellite analysis can easily and economically identify shared MHC class I haplotypes, particularly between related individuals, but provide no information on the alleles encoded on each haplotypes unless an independent method of genotyping is employed to associate specific class I alleles with microsatellite haplotypes [10,11]. Overall, all of these methods require prior knowledge of allele sequences for each macaque species.

The ‘gold standard’ for MHC class I genotyping is cloning and sequencing from a cDNA library of an animal since this has the potential to detect all expressed alleles [12]. The disadvantage of this approach is the prohibitive cost associated with library construction, screening and sequencing enough clones per animal for effective genotyping. We have developed a rapid method for sequence-based genotyping that amplifies full-length NHP class I cDNAs that are molecularly barcoded with multiplex identifier (MID) tags. This approach allows for ligation and transformation of pools of PCR amplicons from multiple animals, reducing cloning costs. Additionally, our method focuses sequencing on a concentrated region of high polymorphism in exons 2 and 3, which encodes the peptide binding domain. By selectively sequencing only this region, we can unambiguously resolve the majority of known macaque alleles while significantly reducing the cost of sequencing. Finally, this method also detects previously unidentified alleles that can easily be sequenced using additional primers to fully characterize these novel MHC class I sequences.

2. Method overview

The workflow outlined in Fig. 1 describes a method for rapid, cost-effective sequence-based genotyping of MHC class I alleles expressed in NHPs. Total cellular RNA is converted to first-strand cDNA and PCR amplicons containing the complete open reading frame of class I A and I B transcripts are generated using primers in the 5' and 3' untranslated regions. Purified amplicons are cloned in *E. coli* and the resulting plasmid DNAs (typically 192/animal) serve as templates for single-pass Sanger sequencing with a primer that flanks the highly polymorphic peptide binding domains encoded by exons 2 and 3 of the class I genes. BLASTn analysis of the resulting sequences against a database of known NHP MHC sequences yields

a summary of the class I transcripts expressed by an animal. By altering the depth of sequence coverage the balance between cost and sensitivity of this genotyping approach may be controlled. Depending on experimental goals, barcoded cDNA amplicons from four or more animals may be pooled and genotyped simultaneously. Since plasmids containing novel cDNA sequences may be prioritized for further characterization by full-length sequencing, this approach is particularly useful for allele discovery efforts in less well-characterized NHP populations such as pig-tailed macaques.

3. Materials and reagents

1. RNA isolation: MagNA Pure LC RNA Isolation Kit – High Performance (Roche Applied Science 03542394001).
2. cDNA synthesis: Superscript® III First-Strand Synthesis System for RT-PCR (Invitrogen 18080-051).
3. PCR amplification DNA oligos:

5' PCR primer	5'MHC-UTR	5' [MID]AGAGTCTCCTCAGACGCCGAG 3'
3' IA PCR primer	3'MHC-UTR-A	5' CAGGAACAYAGACACATTCAGG 3'
3' IB PCR primer	3'MHC-UTR-B	5' GGCTGTCTCTCCACCTCCTCAC 3'
MID tag set	MID1	5' ACGAGTGCGT 3'
	MID2	5' ACGCTCGACA 3'
	MID3	5' AGACGCACTC 3'
	MID4	5' AGCACTGTAG 3'
	MID5	5' ATCAGACACG 3'
	MID6	5' ATATCGCGAG 3'
	MID7	5' CGTGTCTCTA 3'
	MID8	5' CTCGCGTGTC 3'
	MID9	5' TAGTATCAGC 3'
	MID10	5' TCTCTATGCG 3'
	MID11	5' TGATACGTCT 3'
	MID12	5' TACTGAGCTA 3'
	MID13	5' CATAGTAGTG 3'
	MID14	5' CGAGAGATAC 3'

4. PCR amplification: Phusion™ High-Fidelity PCR Master Mix with HF Buffer (New England BioLabs F-531), nuclease-free water, FlashGel® DNA Cassette 1.2% (Lonza 57023), FlashGel® 5× loading dye (Lonza 50462), FlashGel® Quant Ladder 100bp-1.5kb (Lonza 50475).
5. Gel purification of PCR products: 1× TAE, genetic analysis grade agarose, 6× loading dye, SYBR® Safe DNA gel stain 10,000× concentrate in DMSO (Invitrogen S33102), BenchTop 1kb DNA ladder (Promega G7541), MinElute® Gel Extraction Kit (Qiagen 28606).
6. Purified DNA quantitation: Quant-iT™ dsDNA HS Assay Kit (Invitrogen Q32854), Qubit™ assay tubes (Invitrogen Q32856).
7. Ligation and transformation: Zero Blunt® TOPO® PCR Cloning Kit with One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen K2800).

8. Bacterial plate cultures: 50µg/ml kanamycin stock, 50µg/ml ampicillin stock, Ready2AUTOCLAVE LB (Miller's) broth agar powder (ISC BioExpress G-3320-700).
9. Liquid bacterial growth medium: 50µg/ml kanamycin stock, CIRCLEGROW® broth powder (Qbiogene 3000-152).
10. Plasmid purification: Perfectprep™ Plasmid 96 Vac Bind Kit (5 PRIME 2300220).
11. Restriction digestion analysis: EcoRI and 10× NEBuffer EcoRI (New England BioLabs R0101), nuclease-free water, 1× TAE, genetic analysis grade agarose, SYBR® Safe DNA gel stain 10,000× concentrate in DMSO (Invitrogen S33102), 6× loading dye, BenchTop 1kb DNA ladder (Promega G7541).
12. Sequencing oligos: 5' Refstrand (5' GCTACGTGGACGACACGC 3'), SBT190-R (5' TCGCTCTGGTTGTAGTAGC 3').
13. Sequencing reactions: DYEnamic™ ET Terminator Cycle Sequencing Kit (GE Healthcare US81070), magnesium chloride (MgCl₂), Tris, nuclease-free water.
14. Sequencing reaction clean-up: CleanSEQ® Dye-Terminator Removal Kit (Agencourt A29154), 85% ethanol (EtOH), reagent grade water.

4. Step-by-step protocol

4.1 Total RNA isolation from whole blood or cells

To facilitate rapid total RNA isolation from fresh whole blood or fresh or frozen cell samples from up to 32 animals simultaneously, we use the Roche MagNA Pure LC Instrument. Other RNA isolation techniques can be substituted if this robot is not available. Regardless of isolation technique, all reagents and consumables used should be strictly RNase-free. Proper protective equipment should be worn when handling reagents from the MagNA Pure LC RNA Isolation Kit. Typical RNA concentrations from robotic isolation range between 10-35ng/µl with a 50µl elution volume.

- a. Ensure that the MagNA Pure LC Instrument has the proper RNA purification protocols installed as detailed by the MagNA Pure LC RNA Isolation Kit manual.
- b. Prepare the isolation kit reagents as detailed in the manual.
- c. Aliquot 200µl of whole blood or white blood cells, or 1×10^6 peripheral blood mononuclear or cultured cells as directed by the manual.
- d. Set up the instrument using the appropriate protocol for the starting material. Follow instructions supplied by the software, as it will calculate reagent volumes based upon number of samples to be isolated.
- e. Immediately upon completion of the isolation run, quantitate RNA samples with a NanoDrop™ (Thermo Scientific) prior to proceeding with cDNA synthesis. If necessary, RNA samples can be stored in RNase-free microcentrifuge tubes at -80° C.

4.2 Synthesis of first-strand cDNA

We use an MJ Research Tetrad Thermocycler (Bio-Rad Laboratories) for all incubation steps but other thermocyclers may be utilized.

- a. Prepare a reaction mixture of 1µl 10mM dNTP mix, 1µl 50µM Oligo(dT)₂₀, and 50ng total RNA (or up to 8µl if 50ng RNA would exceed 8µl). If necessary, add DEPC-treated water to obtain a 10µl final reaction volume.

- b. Incubate reactions at 65°C for 5 minutes to denature. During this incubation, prepare the cDNA synthesis master mix of 2µl 10× RT Buffer, 4µl 25mM MgCl₂, 2µl 0.1M DTT, 1µl RNaseOUT™, and 1µl SuperScript™ III RT.
- c. Incubate initial reaction mixture at 4°C or on ice for at least 1 minute. During this incubation, add 10µl cDNA synthesis master mix to initial reaction mixture.
- d. Incubate at 50°C for 50 minutes.
- e. Terminate cDNA synthesis by incubating at 85°C for 5 minutes.
- f. Add 1µl 2U/µl *E. coli* RNase H to reaction mixture and incubate at 37°C for 20 minutes to remove any remaining RNA.
- g. Upon completion, proceed directly to PCR or store cDNA at -20°C.

4.3 PCR amplification of MHC class I cDNA

The PCR primers are specific for the MHC class I untranslated regions and generate an ~1.25kb amplicon containing the entire open reading frame. A common 5' primer is paired with two locus-specific 3' primers, so two PCR reactions are performed for each cDNA sample. The 3' MHC-UTR-A oligo has a higher degree of specificity for and preferentially amplifies MHC class I A loci alleles; likewise 3'MHC-UTR-B amplifies primarily class I B loci alleles. Additionally, MID tag sequences can be engineered into the 5' ends of each oligo during synthesis to generate molecularly barcoded PCR products. By using a unique set of MID tagged primers for each animal, it is possible to pool PCR products from multiple animals, thus reducing the total number of ligation and transformation reactions required for cloning. Fourteen such MID tag sequences designed by Roche/454 Life Sciences are provided and additional tags can be designed as necessary. To ensure the generation of sufficient clones per animal for routine genotyping, it is not recommended to pool more than four animals together. However, up to 14 animals may be pooled if the primary goal is discovery of novel class I alleles in understudied populations or other such studies which only require a low depth of coverage.

To reduce the amount of PCR-induced error, the total number of cycles for each reaction should be kept to the absolute minimum required for sufficient amplification and will vary by sample. To rapidly check an aliquot of each reaction for amplification initially after 20 cycles and periodically thereafter, we use the FlashGel DNA cassette system (Lonza). Alternatively, several PCR reactions could be set up for each sample and run in parallel with varying numbers of cycles. Sufficient PCR amplification should be at least 0.4ng/µl (a 1.5kb band visible on a FlashGel in ambient light contains at least 1.6ng DNA). Cloning and sequencing will likely reveal an unacceptably high number of amplicons with PCR errors if more than 30 PCR cycles are used.

- a. Prepare 50µl primer-specific reaction mixtures of 25µl 2× Phusion™ High-Fidelity PCR Master Mix, 20µl nuclease-free water, 1.25µl 5µM 5' PCR primer, 1.25µl 5µM 3' PCR primer (A or B), and 2.5µl cDNA.
- b. PCR amplify using the following cycle conditions for the 3' MHC class I A primer: 98°C for 3 minutes, 20-30 cycles of 98°C for 5 seconds, 60°C for 1 second, 72°C for 20 seconds, and a final elongation of 72°C for 5 minutes. For the 3' MHC class I B primer, increase the annealing temperature to 63°C; all other temperatures and times are unchanged.
- c. After 20 cycles, check a 4µl aliquot of each reaction on a FlashGel DNA cassette (Lonza) following the manufacturer's protocol. Add additional cycles to reactions as needed to obtain sufficient amplification of each sample.

4.4 Agarose gel electrophoresis and amplicon purification

In addition to the desired ~1.25kb product, the PCR amplification process may yield primer dimers and truncated ~1.0kb products that result from aberrant internal priming by the 5' oligo. Therefore gel electrophoresis is required to isolate the full-length MHC class I amplicon. We use SYBR® Safe DNA gel stain (Invitrogen) to visualize products, but ethidium bromide could also be used. Additionally, we have found that the Qiagen MinElute® Gel Extraction Kit yields the best concentration of purified PCR amplicon, but other gel purification kits can be substituted. The MinElute® Gel Extraction Kit includes both spin and vacuum protocols.

- a. Prepare a 1% agarose gel by mixing 150ml 1× TAE with 1.5g agarose.
- b. Heat the mixture in the microwave to melt the agarose, swirling occasionally.
- c. Add 15µl SYBR® Safe DNA gel stain to the mixture, swirling to mix. Pour gel and allow to solidify.
- d. Add 10µl 6× loading dye to each PCR reaction and load entire volume into a single well of the agarose gel. Space samples appropriately leaving at least one empty well between samples to avoid cross-contamination. Load 6µl ladder into an empty well for sizing of the ~1.25kb amplicon.
- e. Run agarose gel in 1× TAE buffer at 120V for 1 hour. Ensure that the artifactual products have sufficiently separated from the ~1.25kb amplicon to allow for clean excision. Extend running time if necessary.
- f. Once sufficient separation is observed, visualize with a 'DNA safe' blue light transilluminator (Invitrogen) and excise the ~1.25kb amplicon band. Place the gel slice into a colorless 1.5ml microcentrifuge tube.
- g. Proceed with gel slice purification following the appropriate MinElute® protocol. Elute the DNA with 10µl nuclease-free water. Typical post-elution DNA concentrations range from 2-15ng/µl.

4.5 Purified DNA quantitation and pooling

Accurate quantitation of the purified DNA products is essential if pooling PCR products from multiple animals. If products are not pooled in equal concentrations, there may be an over representation of clones from the most concentrated sample and some animals in the pool may not be adequately genotyped. We use the Qubit™ fluorometer (Invitrogen) to accurately quantitate our samples, but other approaches may also be used.

- a. Prepare the Quant-iT™ working solution, dilute standards and samples, calibrate the fluorometer, read samples, and calculate concentrations of samples as directed by the manufacturer's protocol.
- b. If pooling products for ligation, normalize all products for each pool to the concentration of the least concentrated sample. Combine equal volumes of the normalized products to create each pool.
- c. Proceed directly to ligation and transformation or store DNA products at -20°C.

4.6 Ligation and transformation of blunt-end PCR amplicons

As stated previously, molecularly barcoding the PCR amplicons can greatly reduce the required number of ligation and transformations. However, since a routine transformation typically results in several hundred clones, this limits the number of animals that can be pooled together and still achieve sufficient coverage for genotyping. The MHC class I A region contains at most 6 alleles per animal and at least half of those are considered to be comparatively rare

transcripts ('minor'), so a depth of coverage of 48 class I A clones per animal is adequate to detect the abundant transcripts ('major'). The macaque class I B region is considerably more complex, containing up to 24 expressed loci per animal and therefore requiring more clones to satisfactorily detect the major B locus alleles. For convenience we use a depth of coverage of 48 class I A and 144 class I B clones per animal for routine genotyping (192 clones total in two 96-well deep-well culture plates). As seen in Fig. 2, similar coverage can also be achieved with strategic pooling of MID barcoded class I A and I B products from multiple animals, reducing the number of ligation/transformation reactions for both regions. This pooling strategy reduces the number of ligations/transformations required for both loci in four animals from eight to three for a significant decrease in overall cloning cost. Depth of coverage can be varied by picking more or less clones per animal, depending on genotyping needs.

- a. Prepare LB agar plates containing 50µg/ml kanamycin (2 or more per ligation to be performed) and LB agar plates containing 50µg/ml ampicillin (positive control).
- b. Prepare ligation mixture of 1µl salt solution (1.2M NaCl + 0.06M MgCl₂), 1µl pCR®-BluntII-TOPO®, and 4µl purified PCR product (single or pooled amplicons; typical concentration range 8-60ng per ligation reaction). Use nuclease-free water in place of PCR product for a negative control.
- c. Incubate ligation at room temperature for 30 minutes.
- d. During the last 10 minutes of the ligation incubation, thaw TOP10 chemically competent cells on ice.
- e. Add 4µl of incubated ligation mixture to TOP10 cells, stirring with pipette tip to mix (do not pipette up and down). Add 1µl 10pg/µl pUC19 control DNA to cells for a positive transformation control.
- f. Incubate ligation mixture with TOP10 cells on ice for 5 minutes.
- g. Heat shock cells at 42°C for 30 seconds.
- h. Recover cells on ice for 2 minutes.
- i. Add 250µl room temperature S.O.C. medium and incubate at 37°C shaking the tubes horizontally at 200 rpm for 1 hour.
- j. During the incubation, warm the LB + kanamycin plates (LB + ampicillin plates for the positive control) to 37°C.
- k. Spread two 150µl volumes of each transformation mixture onto two LB + kanamycin plates. Spread 150µl of the positive control onto an LB + ampicillin plate.
- l. Incubate LB selective plates at 37°C overnight.
- m. Prepare CIRCLEGROW® (or LB) broth plus 50µg/ml kanamycin.
- n. Pick 48 class I A clones per animal (192 class I A clones for a ligation pool of 4 animals), 144 class I B clones per animal (192-288 class I B clones for a ligation pool of 2 animals) into 1.3ml selective liquid media in 96-well deep-well culture plates. For a control of the kanamycin-selective liquid media, pick a clone from the positive transformation control plate into the broth + kanamycin in a separate bacterial culture tube (since the positive control colonies are resistant to ampicillin, not kanamycin, there should be no growth in the control).
- o. Incubate deep-well culture plates at 37°C overnight, shaking at 200 rpm.

4.7 Plasmid DNA purification and restriction enzyme digestion

We use a four station vacuum manifold (5 PRIME) to miniprep up to four 96-well plasmid plates simultaneously, but other plasmid purification protocols such as magnetic bead-based methods may also be used. The pCR®-BluntII-TOPO® vector contains EcoRI cleavage sites flanking the PCR product ligation site, so an EcoRI restriction digest can easily confirm cloning efficiency with a subset of samples before proceeding to sequencing.

- a. Centrifuge the deep-well culture plates at $1,900 \times g$ for 5 minutes to pellet bacteria. Pour off supernatant, invert and blot any remaining liquid medium to a paper towel.
- b. Proceed with plasmid DNA purification following manufacturer's protocol. Typical post-miniprep concentrations range from 50-150ng/μl.
- c. To perform the EcoRI restriction digest, make a reaction mix of 0.5μl EcoRI enzyme, 1μl 10× NEBuffer EcoRI, 5.5μl nuclease-free water, and 3μl purified plasmid DNA.
- d. Incubate digestion mixture at 37°C for 1 hour.
- e. During incubation, prepare a 1% agarose gel as described above in 4.4 steps (a)-(c).
- f. Add 2μl 6× loading dye to each digest and load into agarose gel. Load 6μl ladder into an empty well for sizing the ~3.5kb vector and the ~1.25kb insert.
- g. Run agarose gel in 1× TAE buffer at 120V for 15 minutes. Analyze the gel to calculate the frequency of clones with the desired insert (generally >85%).

4.8 Single-pass cycle sequencing of cDNA clones

MHC class I alleles exhibit polymorphism throughout the entire ~1.1kb open reading frame. However, there is a short highly conserved region of sequence just upstream of the peptide binding domains encoded by exons 2 and 3. We developed a sequencing primer, 5'Refstrand, in this conserved region that we can use to perform single-pass sequencing of these highly polymorphic exons. With the ~500bp of sequence generated from sequencing from this single oligo, we can unambiguously resolve the majority of the known MHC class I alleles of macaques. For the subset of alleles that only differ from each other outside of the region we are sequencing with this method, we can still unambiguously assign them to a unique allele lineage with the sequence we generate. For example, as seen in Fig. 3, *Mamu-A1*0040101* and *Mamu-A1*0040102* are identical through exons 2 and 3 so we cannot distinguish the specific allele but can unambiguously assign that sequence to the *Mamu-A1*00401* lineage. Since sufficient information for genotyping can be generated from a single sequencing reaction (rather than five reactions required to span the entire open reading frame), our method dramatically reduces the sequencing costs. Previously unidentified sequences noted in the single pass sequencing can be selectively sequenced with additional primers to characterize the entire open reading frame.

If using the pooled ligation approach, a second sequencing reaction using the oligo SBT190-R is required to generate sequence back to the MID tags incorporated in the 5'MHC-UTR oligo to link each sequence with the animal from which it was amplified. This reaction is necessary since the 5'Refstrand primer is located ~1.1kb away from the 3'MHC-UTR primers and a typical sequencing reaction produces only ~500bp of high quality sequence, so a single sequencing reaction would not be sufficient to generate enough sequence to reach the MID tags if they were incorporated into the 3' primers. SBT190-R is located ~350bp from the 5' MHC-UTR primer. Even with this additional reaction, sequencing costs are still reduced by more than half.

- a. Prepare a 5mM MgCl₂, 20mM Tris sequencing buffer by adding 0.102g MgCl₂ and 0.242g Tris to 80ml double-distilled water, stir to dissolve, adjust pH to 9.0, and bring volume to 100ml with double-distilled water.
- b. Prepare sequencing reaction mix containing 1μl DYEnamic™ ET Terminator, 1μl 3.2μM oligo (either 5'Refstrand or SBT190-R), 6μl sequencing buffer, 7μl nuclease-free water, and 5μl purified plasmid DNA.
- c. Cycle sequence using the following thermocycler profile: 30 cycles of 95°C for 20 seconds, 50°C for 15 seconds, and 60°C for 1 minute.
- d. Purify the sequencing reactions using the Agencourt® CleanSEQ® Dye-Terminator Removal kit following the manufacturer's protocol for ET Terminator clean-up.
- e. Resolve the cleaned sequencing products by capillary electrophoresis on an ABI 3730 Genetic Analyzer (Applied Biosystems), or a comparable Sanger sequencing platform.

4.9 Sequence analysis

We maintain an internal custom database of NHP MHC class I alleles which we can use to perform Basic Local Alignment Search Tool (BLAST) analyses. In the absence of a curated MHC-specific database, similar analyses can be performed through the nucleotide BLAST link on the National Center for Biotechnology Information (NCBI) website (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

- a. Edit single-pass sequence results for high quality data using CodonCode Aligner (CodonCode Corporation) or similar sequence analysis software. Clip sequence ends to maximize the region with an error rate below 0.01 and delete short sequences with less than 50 Phred 20 quality bases.
- b. Export high quality sequence for BLASTn analysis. Use the following parameters to most effectively analyze MHC class I alleles through the NCBI website: 'Others (nr etc.)' database, 'Nucleotide collection (nr/nt)', limit organism to the genus and species of NHP being analyzed (such as *Macaca mulatta*), and optimize for 'Highly similar sequences (megablast)'. Under 'Algorithm parameters', 'General Parameters', limiting max target sequences to '10' will also simplify the results.
- c. Determine which cDNA clones are identical to one or more known MHC class I alleles by looking for 100% identity to any alleles listed in GenBank.
- d. Determine the closest match for any sequences lacking 100% identity with any alleles deposited in GenBank.
- e. Perform any additional sequencing reactions necessary to fully characterize these novel alleles as described above in section 4.8, using other previously described MHC class I bidirectional sequencing oligos T7, M13, 3'Refstrand, and 5'Transmembrane [14].
- f. Deposit novel full-length sequences to GenBank and the Immuno Polymorphism Database (IPD) [15] to obtain official allele names.

5. Closing remarks

High-resolution sequence-based genotyping of human HLA alleles is widely available [16], but comparable capabilities are currently lacking for the analogous MHC alleles in NHPs that serve as preclinical models for human infectious disease and transplantation research. Our method for sequence-based MHC class I genotyping seeks to begin addressing this issue. It has been applied to over 120 samples from 3 different NHP species to date (*Macaca mulatta*,

M. fascicularis, and *M. nemestrina*), proving that the same primers can effectively amplify MHC class I alleles in several species of macaques and likely other related NHP species. An example of the typical results obtained for a single Indian-origin rhesus macaque by this method is shown in Fig. 3. Of the 192 cDNA plasmids originally cloned and sequenced for this animal, 170 (88.5%) resulted in high quality sequences that were subsequently compared against known rhesus alleles using BLASTn. The efficiency of performing transformation reactions with pooled PCR products and reduction in required number of sequencing reactions has proven this to be a relatively rapid and cost-effective genotyping method. Additionally, many of the steps involved in this method, such as colony picking and plasmid minipreps, are amenable to automation. This method also simplifies allele discovery efforts since it generates full-length cDNAs and rapidly identifies novel clones for selective additional sequencing.

The general principles of this method could be applied to any highly polymorphic gene of interest, provided there are diagnostic regions of polymorphism like exons 2 and 3 of MHC class I alleles. For example, we have adapted this genotyping method for DRB and other MHC class II loci using analogous polymorphic domains [17]. Finally, this method can also be adapted for use on next-generation sequencing platforms such as the Roche/454 Genome Sequencer FLX system (Wiseman et al., manuscript under review). Pooled, MID barcoded amplicons of appropriate length, currently approximately 200bp, can be amplified and pyrosequenced in massively parallel fashion for ultra high-throughput genotyping.

Acknowledgments

We thank members of the O'Connor laboratory for assistance developing these protocols. This work was supported by NIAID contract number HHSN266200400088C/N01-A1-40088 and by grant number 1 R24 RR021745-01A1. This publication was made possible in part by grant numbers P51 RR000167 and P40 RR019995 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) to the Wisconsin National Primate Research Center, University of Wisconsin-Madison. This research was conducted at a facility constructed with support from Research Facilities Improvement Program grant numbers RR15459-01 and RR020141-01.

References

1. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC Jr, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S. Nat Rev Genet 2004;5:889–899. [PubMed: 15573121]
2. Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE. Genome Res 2004;14:1501–1515. [PubMed: 15289473]
3. Rosenzweig M, Connole M, Glickman R, Yue SP, Noren B, DeMaria M, Johnson RP. Blood 2001;97:1951–1959. [PubMed: 11264158]
4. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Science 2002;296:1439–1443. [PubMed: 12029127]
5. Sauermann U, Siddiqui R, Suh YS, Platzer M, Leuchte N, Meyer H, Matz-Rensing K, Stoiber H, Nurnberg P, Hunsmann G, Stahl-Hennig C, Krawczak M. Genes Immun 2008;9:69–80. [PubMed: 18094710]
6. Knapp LA. Tissue Antigens 2005;65:211–219. [PubMed: 15730514]
7. Blasky AJ, Karl JA, Wiseman RW, Read DS, O'Connor DH. Immunogenetics 2008;60:575–584. [PubMed: 18629489]
8. Kaizu M, Borchardt GJ, Glidden CE, Fisk DL, Loffredo JT, Watkins DI, Rehrauer WM. Immunogenetics 2007;59:693–703. [PubMed: 17641886]
9. Karl JA, Wiseman RW, Campbell KJ, Blasky AJ, Hughes AL, Ferguson B, Read DS, O'Connor DH. Immunogenetics 2008;60:37–46. [PubMed: 18097659]
10. Penedo MC, Bontrop RE, Heijmans CM, Otting N, Noort R, Rouweler AJ, de Groot N, de Groot NG, Ward T, Doxiadis GG. Immunogenetics 2005;57:198–209. [PubMed: 15900491]

11. Wiseman RW, Wojcechowskyj JA, Greene JM, Blasky AJ, Goon T, Soma T, Friedrich TC, O'Connor SL, O'Connor DH. J Virol 2007;81:349–361. [PubMed: 17035320]
12. Loffredo JT, Friedrich TC, Leon EJ, Stephany JJ, Rodrigues DS, Spencer SP, Bean AT, Beal DR, Burwitz BJ, Rudersdorf RA, Wallace LT, Piaskowski SM, May GE, Sidney J, Gostick E, Wilson NA, Price DA, Kallas EG, Piontkivska H, Hughes AL, Sette A, Watkins DI. PLoS One 2007;2:e1152. [PubMed: 18000532]
13. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE. Proc Natl Acad Sci U S A 2005;102:1626–1631. [PubMed: 15665097]
14. Pendley CJ, Becker EA, Karl JA, Blasky AJ, Wiseman RW, Hughes AL, O'Connor SL, O'Connor DH. Immunogenetics 2008;60:339–351. [PubMed: 18504574]
15. Robinson J, Waller MJ, Stoeckl P, Marsh SG. Nucleic Acids Res 2005;33:D523–6. [PubMed: 15608253]
16. Maiers M, Gragert L, Klitz W. Hum Immunol 2007;68:779–788. [PubMed: 17869653]
17. O'Connor SL, Blasky AJ, Pendley CJ, Becker EA, Wiseman RW, Karl JA, Hughes AL, O'Connor DH. Immunogenetics 2007;59:449–462. [PubMed: 17384942]

Fig. 1.

A workflow overview of sequence-based MHC class I genotyping method. Numbers in front of each step refer to the detailed step-by-step protocol in section 4 of the text.

Fig. 2.

Pooling strategy for routine genotyping using MID barcoded PCR amplicons. (a) Cloning of non-barcoded class I A and I B products of four animals (NHP 1-NHP 4) can be performed with eight separate ligation/transformation reactions, picking 48 class I A clones and 144 class I B clones from each animal, for a total of 192 clones per animal. In this instance, eight ligation/transformation reactions are performed to obtain 768 total clones (8 96-well plates) from the four animals. (b) Alternatively, MID barcoded PCR products could be pooled into three total pools, one pool with the class I A products from all four animals (red) and two pools with the class I B products from two animals each (blue and green). Picking 192 clones from the class I A pool (2 96-well plates) and 288 clones from each of the two class I B pools (3 96-well plates from each pool) results in a comparable depth of coverage for four animals.

Allele	# Obs	%	
<i>Mamu-A</i>			
<i>A1*0040101/0102</i>	16	9.4	
<i>A1*00801</i>	42	24.7	
<i>Mamu-B</i>			
<i>B*0010101</i>	17	10.0	*
<i>B*00703</i>	19	11.2	*
<i>B*01201</i>	4	2.4	#
<i>B*03001</i>	10	5.9	#
<i>B*03002/0302/04/05</i>	6	3.6	*
<i>B*03801</i>	8	4.7	#
<i>B*05301/02</i>	1	0.6	#
<i>B*05701</i>	1	0.6	#
<i>B*05702</i>	2	1.2	*
<i>B*06001</i>	2	1.2	#
<i>B*07201/02</i>	7	4.1	*
<i>B*07401/02</i>	1	0.6	#
<i>B*07402-V1</i>	2	1.2	#
<i>B*08202</i>	10	5.9	#
<i>Mamu-I</i>			
<i>I*0115;I*03014</i>	3	1.8	
<i>B*00903</i>	1	0.6	
Unknown	18	10.6	
Total	170		
* Associated with haplotype InB26			
# Associated with haplotype InB11a			

Fig. 3.

MHC class I sequence-based genotyping results for a single Indian-origin rhesus macaque. Listed are all of the known alleles identified in this animal as well as the number of clones (# Obs) and the relative frequency expressed as a percentage of the total clones examined (%) for each allele. We identified two *Mamu-A*, 14 *Mamu-B*, and two *Mamu-I* alleles (the *I* locus is a *B*-like locus with unusually low variability). Based on this data we can infer that this animal has two of the most common previously reported *Mamu-B* haplotypes for Indian rhesus macaques, InB26 and InB11a [13]. Eighteen sequences did not match any previously identified *Mamu* alleles and are either putative novel alleles or single-read PCR sequencing artifacts.