Human Immunology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

HAPCAD: An open-source tool to detect PCR crossovers in next-generation sequencing generated HLA data

Shana L. McDevitt [a,c,*], Jessen V. Bredeson [b], Scott W. Roy [c], Julie A. Lane [a], Janelle A. Noble [a]

[a] Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA
[b] Department of Molecular & Cell Biology, University of California, 142 LSA #3200, Berkeley, CA 94720, USA
[c] San Francisco State University, Department of Biology, 1600 Holloway Avenue, San Francisco, CA 94132, USA

ARTICLE INFO

Article history:
Received 30 May 2015
Revised 4 December 2015
Accepted 18 January 2016
Available online xxxx

Keywords:
Human Leukocyte Antigen
IMGT/HLA Database
Next-generation sequencing
PCR crossovers
Open-source tools

ABSTRACT

Next-generation sequencing (NGS) based HLA genotyping can generate PCR artifacts corresponding to IMGT/HLA Database alleles, for which multiple examples have been observed, including sequence corresponding to the *HLA-DRB1*03:42* allele. Repeat genotyping of 131 samples, previously genotyped as *DRB1*03:01* homozygotes using probe-based methods, resulted in the heterozygous call *DRB1*03:01 + DRB1*03:42*. The apparent rare *DRB1*03:42* allele is hypothesized to be a "hybrid amplicon" generated by PCR crossover, a process in which a partial PCR product denatures from its template, anneals to a different allele template, and extends to completion. Unlike most PCR crossover products, "hybrid amplicons" always corresponds to an IMGT/HLA Database allele, necessitating a case-by-case analysis of whether its occurrence reflects the actual allele or is simply the result of PCR crossover. The Hybrid Amplicon/PCR Crossover Artifact Detector (HAPCAD) program mimics jumping PCR *in silico* and flags allele sequences that may also be generated as hybrid amplicon.

1. Introduction

Amplicon-based Human Leukocyte Antigen (HLA) genotyping is most commonly performed using 454™ pyrosequencing chemistry, which offers the longest next generation sequencing (NGS) single-read lengths available at over 800 base pairs [1]. 454™-based NGS platforms are among the few NGS platforms capable of sequencing through most HLA amplicons, ranging from 300 to 750 base pairs, in one read. Amplicons are designed to contain the most polymorphic and functionally relevant regions of HLA, namely, the exons encoding the peptide-binding groove [2]. However, allelic polymorphism and sequence similarity among gene sequences limit PCR primer design. Improperly developed primers may amplify other loci in addition to the target exon, or they may fail to amplify all alleles of a particular locus. While difficult to design, primers specific to single HLA exons have been successfully developed for all classical loci except the *HLA-DRB* loci.

HLA-DR molecules comprise an alpha chain, which is essentially monomorphic, and a highly polymorphic beta chain [3]. Unlike other class II molecules, the beta chain of the DR molecule can be encoded by one of four *DRB* loci [3]. Every human chromosome 6 contains a *DRB1* gene. Depending upon the *DRB1* allele present, the same chromosome may or may not also contain a *DRB3*, *DRB4*, or *DRB5* gene. Thus, the surface of antigen presenting cells may include up to four different HLA-DR molecules, with each of the different DR-beta chains paired with the same DR-alpha chain [3,4]. Like many HLA loci, these DRB loci share extensive sequence homology [4], which, as noted previously, creates challenges in primer sensitivity (amplifying the peptide-binding region of all divergent alleles of the DRB loci) and specificity (amplifying only a single locus, e.g., *HLA-DRB1*, *DRB3*, *DRB4*, or *DRB5*). Current primers co-amplify all alleles of all *DRB* loci present in the same PCR [5]. One consequence of this co-amplification of these closely-related loci is that amplicons may be generated that do not reflect the genome. Specifically, spurious PCR products are generated that represent chimeras of different loci, through a phenomenon termed "jumping" PCR (or PCR crossover), in which a partial PCR product

Abbreviations: HAPCAD, Hybrid Amplicon/PCR Crossover Artifact Detector; HLA, Human Leukocyte Antigen; NGS, next-generation sequencing; SSO, sequence specific oligonucleotide; CWD, Common and Well-Documented; SBT, Sequence-Based Typing; P1, parent allele 1; P2, parent allele 2.

* Corresponding author at: Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA.

E-mail addresses: shana.mcdevitt@gmail.com (S.L. McDevitt), jessenbredeson@berkeley.edu (J.V. Bredeson), scottwroy@gmail.com (S.W. Roy), jlane@chori.org (J.A. Lane), jnoble@chori.org (J.A. Noble).

http://dx.doi.org/10.1016/j.humimm.2016.01.013
0198-8859/

dissociates from its template and anneals to a template generated from a different allele (Fig. 1) [5,6]. This produces an amplicon that combines sequences from two different HLA alleles into a hybrid sequence or PCR crossover product. For the purposes of clarification throughout this manuscript, a PCR crossover product is defined as a "hybrid amplicon" only if the PCR crossover product corresponds to an IMGT/HLA Database allele. This phenomenon can happen with amplification of two alleles of a single locus; however, amplification of multiple DRB loci with a single set of primers appears to increase the likelihood. PCR crossover products defined by Holcomb et al. [6], as a majority do not correspond to IMGT/HLA Database alleles and are easily detected due to mismatches between the crossover sequence and allele sequence [6]. The hybrid amplicon subgroup of PCR crossovers can however pass undetected because the sequences are exact matches to IMGT/HLA Database alleles.

Multiple examples of putative hybrid amplicons corresponding to reported HLA alleles have been observed (Table 1). A striking example was the apparent appearance of the *HLA-DRB1*03:42* allele after introducing NGS technology into the lab. One hundred and thirty-one samples, previously genotyped as DRB1*03:01/DRB1*03:01 homozygotes using sequence specific oligonucleotide (SSO) methods, were re-genotyped using the original SSO method primer sequences with Roche 454™ NGS-based technology. Fifty-nine of these samples (45%) resulted in the heterozygous genotype call *DRB1*03:01 + DRB1*03:42*. In our data sets, the *DRB1*03:42*

Table 1
Motif-specific primer pair combinations and expected amplifications. It displays twelve primer pair combinations used to amplify HAR, cell line DNA. Motif-specific targets and the allele sequences that may be amplified (DRB1*03:01, DRB1*03:42, DRB3*01:01, and DRB3*01:14) with specific combinations are noted.

Combination	Forward primer	Reverse primer	Expected allele amplification
1	YSTS	V	DRB1*03:01
2	YSTS	G	DRB3*01:14
3	YSTS	DRB general	DRB1*03:01, DRB3*01:14
4	LRKS	V	DRB1*03:42
5	LRKS	G	DRB3*01:01
6	LRKS	DRB general	DRB1*03:42, DRB3*01:01
7	LLKS	V	None
8	LLKS	G	None
9	LLKS	DRB general	None
10	DRB general	V	DRB1*03:42, DRB1*03:01
11	DRB general	G	DRB3*01:01, DRB3*02:02, DRB3*01:14
12	DRB general	DRB general	All

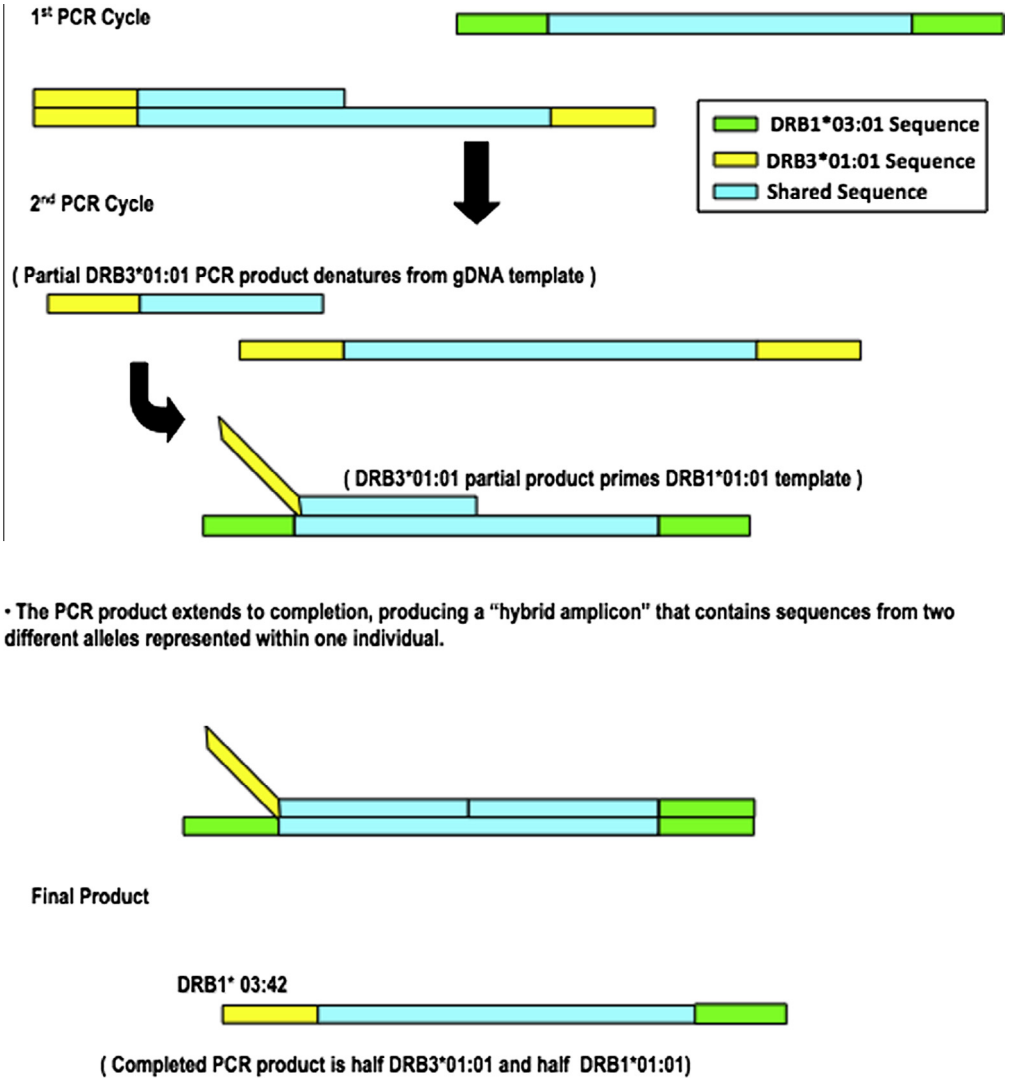

Fig. 1. "Jumping" PCR/PCR crossover schematic. It displays the "Jumping" PCR mechanism used to explain the *in vitro* production of DRB1*03:42 from partial DRB3*01:01:02:01 and DRB1*03:01:01:01 alleles often found on a DR52 haplotype [4].

Please cite this article in press as: S.L. McDevitt et al., HAPCAD: An open-source tool to detect PCR crossovers in next-generation sequencing generated HLA data, Hum. Immunol. (2016), http://dx.doi.org/10.1016/j.humimm.2016.01.013