

Full-length characterization of novel HLA-DRB1 alleles for reference database submission

Kathrin Putke^{1*}, Viviane Albrecht^{1*}, Christin Paech¹, Madlen Pahlke¹, Bianca Schöne¹, Steffen Klasberg¹, Alexander H. Schmidt^{1,2}, Vinzenz Lange¹, Gerhard Schöfl¹, Anja Klussmeier¹

¹DKMS Life Science Lab, Dresden, Germany

²DKMS, Tübingen, Germany

*These authors contributed equally to this work

Correspondence:

Anja Klussmeier

klussmeier@dkms-lab.de

Keywords: genotyping, novel allele, PCR, primer, HLA, DRB1, sequencing

Running title: Characterization of Novel HLA-DRB1 alleles

Abstract

The prerequisite for successful HLA genotyping is the integrity of the large allele reference database IPD-IMGT/HLA. Consequently, it is in the laboratories' best interest that the data quality of submitted novel sequences is high. However, due to its long and variable length, the gene HLA-DRB1 presents the biggest challenge and as of today only 16% of the HLA-DRB1 alleles in the database are characterized in full length. To improve this situation, we developed a protocol for long-range PCR amplification of targeted HLA-DRB1 alleles. By subsequently combining both long-read and short-read sequencing technologies our protocol ensures phased and error-corrected sequences of reference grade quality. This dual redundant reference sequencing (DR2S) approach is of particular importance for correctly resolving the challenging repeat regions of DRB1 intron1. Until today, we used this protocol to characterize and submit 384 full-length HLA-DRB1 sequences to IPD-IMGT/HLA.

1 Introduction

The human leukocyte antigen (HLA) genes are the most diverse genes in the human genome. As of today, 3,389 HLA-DRB1 alleles are known and encode 2,203 distinct proteins (IPD-IMGT/HLA release 3.50, 10/2022) (*1*). Despite the high number of named alleles, the identification of novel HLA-DRB1 alleles is not a rare event during stem cell donor registry genotyping. In contrast to other HLA genes, most laboratories are continuing to submit only the sequences of exons 2 and 3 to the IPD-IMGT/HLA database resulting in poor representation of HLA-DRB1 diversity outside this region. One reason for this is the length of HLA-DRB1, and more specifically the varying length of the different HLA-DRB1 alleles. Due to large sequence variations in intron 1, HLA-DRB1 alleles can span between 11 and 16 kb. While this length in itself poses a substantial challenge for PCR success, preferential amplification will often result in the loss of the longer allele in the presence of a shorter allele. Due to these difficulties, most novel HLA-DRB1 alleles are submitted to the IPD-IMGT/HLA database as partial sequences only. Unfortunately, these partial sequences can then cause

38 inconsistencies in genotyping algorithms because missing bases need to be imputed. It would
 39 therefore be valuable for the IPD-IMGT/HLA database and the HLA genotyping laboratories if novel
 40 HLA-DRB1 alleles would be characterized and submitted in full length.

41 To overcome the challenge of HLA-DRB1 full-gene characterization, several strategies are available.
 42 One protocol splits the long gene in two or more PCR amplicons of defined length, which can then
 43 be sequenced separately (2). This will in all likelihood increase PCR success and might be the
 44 method of choice if the genomic DNA is of low molecular weight. However, it might not always be
 45 possible to phase the individual parts unambiguously. Therefore, one PCR covering the complete
 46 gene is preferable. To avoid the length-biased PCR amplification, we developed primer pairs that
 47 specifically amplify HLA-DRB1 alleles of similar length. This strategy requires the prior knowledge
 48 of the targeted allele in at least allele-group resolution (first field). This information is usually
 49 available if the workflow is used for the characterization of novel alleles.

50 For submission of novel alleles to IPD/IMGT-HLA, sequencing data from two independent PCR
 51 reactions per sample are obligatory. While this requirement is generally satisfied by using the same
 52 sequencing method twice, we prefer to exploit the strengths of two different technologies: long-read
 53 PacBio HiFi- or Oxford Nanopore (ONT) sequencing to obtain a fully phased sequence and short-
 54 read Illumina shotgun sequencing to correct for potential sequencing errors. We developed DR2S
 55 (Dual Redundant Reference Sequencing) to integrate the analysis of both long and short reads and
 56 generate reference quality sequence data (3, 4).

57 Within this chapter, we share the protocol for HLA-DRB1 full-length characterization starting at the
 58 laboratory and ending at sequence submission to the IPD-IMGT/HLA database. This protocol has
 59 been widely tested and used successfully for the characterization of over 300 novel HLA-DRB1
 60 alleles from buccal swab material.

61 2 Materials

62 2.1 PCR Primers

63 HLA-DRB1 PCR (see note 1):

64 Forward1 (F1): <Overlap Sequence>CAGATGCTGATTCGTTCTCCAACACT
 65 Forward2 (F2): <Overlap Sequence>TTCTCCAACACGAGATTACCCAACC
 66 Reverse1 (R1): <Overlap Sequence>AGCACAAAAGTTGGAAGATGAGGGG
 67 Reverse2 (R2): <Overlap Sequence>GAGCACAAAAGTTGAAGATGAGGCG
 68 Reverse3 (R3): <Overlap Sequence>GAGCACAAAAGTTGAAGATGAGGCAC
 69 Overlap Sequence PacBio forward primer: ACACTCTTCCCTACACGACGCTCTCCGATCT
 70 Overlap Sequence PacBio reverse primer: GTGACTGGAGTTCAGA
 71 Overlap Sequence ONT forward primer: ACTTCGTACGTACGGCGTCTTATAC
 72 Overlap Sequence ONT reverse primer: GAGACACGTCCGATTACGACTTGAC

73 Sample Barcode PCR:

74 PacBio barcode forward primer:
 75 GGTAG<BARCODE>ACACTCTTCCCTACACGACGCTCTCCGATCT
 76 PacBio barcode reverse primer:
 77 CCATC<BARCODE>GTGACTGGAGTTCAGA

78 For PacBio <BARCODE> sequences, use barcode sequences provided by PacBio (e.g. from
79 SMRTbell barcoded adapter plate 3.0)

80

81 ONT barcode forward primer:

82 <BARCODE>ACTTCGTACGTACGGCGTCTTATAC

83 ONT barcode reverse primer:

84 <BARCODE>GAGACACGTCCGATTACGACTTGAC

85 For ONT <BARCODE> sequences, see Supplementary Table 1

86 2.2 PCR

- 87 • High-quality genomic DNA (HLA-DRB1 pre-genotyped at allele-group resolution) (see note
- 88 2)
- 89 • Allele-group specific primers (see Table 1)
- 90 • Advantage Genomic LA Polymerase Mix (Takara Bio, Mountain View, California) or
- 91 another long amplification PCR kit
- 92 • dNTPs
- 93 • 96well plates or reaction tubes
- 94 • Thermocycler
- 95 • Gel-electrophoresis system

96 2.3 Purification

- 97 • SPRIselect beads
- 98 • Magnetic rack
- 99 • 80% ethanol

100 2.4 Library Preparation & Sequencing: Illumina Shotgun

- 101 • NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (NEB)
- 102 • NEBNext Ultra II Ligation Module (NEB)
- 103 • MgCl₂ (2.5 mM)
- 104 • NEBNext Multiplex Oligos for Illumina (Dual Index Set 1)
- 105 • NEBNext Multiplex Oligos for Illumina (Dual Index Set 2)
- 106 • Reagents and system for Illumina library quantification (e.g. qPCR)
- 107 • MiSeq
- 108 • MiSeq (500 cycle) Reagent Kit

109 2.5 Library Preparation & Sequencing: PacBio

- 110 • HulaMixer Sample Mixer (ThermoFisher Scientific)
- 111 • Qubit Fluorometer (ThermoFisher Scientific)
- 112 • Qubit 1X dsDNA HS Assay Kit (ThermoFisher Scientific)
- 113 • BluePippin (Sage Science)
- 114 • 0.75% Agarose Cassettes, dye-free, ext. markers, BluePippin (Sage Science)
- 115 • SMRTbell Express Template Prep Kit 2.0 (PacBio)

- 116 • AMPure PB Beads
- 117 • Sequel/Sequel II (PacBio)
- 118 • PreCR Repair Mix (NEB) (optional)

119 2.6 Library Preparation & Sequencing: ONT

- 120 • Qubit Fluorometer (ThermoFisher Scientific)
- 121 • Qubit 1X dsDNA HS Assay Kit (ThermoFisher Scientific)
- 122 • NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing
- 123 (NEB)
- 124 • ONT sequencing platform (MinION, GridION)
- 125 • MinION/GridION flowcell (10.4.1 chemistry)
- 126 • Ligation Sequencing Kit V14 (SQK-LSK114)

127 2.7 Genotyping & Sequence Submission to IPD-IMGT/HLA

- 128 • DR2S (download from <https://github.com/DKMS-LSL/dr2s>)
- 129 • TypeLoader2 (download from <https://github.com/DKMS-LSL/typeloader>)

130 3 Methods

131 If not stated otherwise in a kit manual, work at room temperature. For long-read sequencing, avoid
 132 fast pipetting through thin pipette tips to prevent shearing the DNA. In particular, do not mix by
 133 pipetting. Instead, mix by inversion or flicking the tube with a finger.

134 3.1 PCR

- 135 1. Choose the appropriate primers for the HLA-DRB1 allele you want to characterize in full
 136 length (Table 1).
- 137 2. Perform two independent long-range PCRs with 20-200 ng of genomic DNA in a 25 µl reaction
 138 using 0.2 µM primers, 0.4 mM dNTPs, 1x Advantage Genomic LA Buffer and 1.25 U
 139 Advantage Genomic LA Polymerase Mix. Cycling protocol: 94°C 1 min, 35 cycles: 98°C
 140 10 sec/68°C 12 min, 72°C 10 min. Verify amplification success and specificity by gel-
 141 electrophoresis (Figure 1, see notes 3 and 4).
- 142 3. Use one PCR reaction for Illumina shotgun sequencing and the other for long-read sequencing
 143 (PacBio or ONT).

144 3.2 Illumina Library Preparation & Sequencing

- 145 1. Fragment 10 µl of PCR product by adding 0.5 mM MgCl₂ (final concentration), 2.5 µl
 146 NEBNext Ultra II FS Reaction Buffer and 0.1 µl NEBNext Ultra II FS Enzyme Mix in a 25 µl
 147 reaction. Incubate at 37°C for 30 min and 65°C for 20 min.
- 148 2. Attach barcoded sample adapters by adding the following reagents to the 25 µl fragmentation
 149 reaction: 7.5 µl NEBNext Ultra II Ligation Master Mix, 0.25 µl NEBNext Ligation Enhancer,
 150 0.63 µl NEBNext Adaptor for Illumina. Incubate at 20°C for 15 min.
- 151 3. Add 0.75 µl USER enzyme and 0.75 µl water to the reaction. Incubate at 37°C for 15 min.

- 152 4. Purify the DNA by adding 15.5 μ l water and 40 μ l SPRIselect beads to the reaction. Make sure
 153 the beads are suspended completely by vortexing. Mix by inversion and spin down only for a
 154 short time such that the beads remain dispersed. Incubate for 5 min. Separate beads on a
 155 magnetic rack. Discard the supernatant. Perform a wash step with 100 μ l 80% ethanol leaving
 156 the tube on the magnetic rack. Discard the supernatant. Repeat the wash step. Spin down
 157 shortly to gather the ethanol at the bottom of the plate/tube. Separate the ethanol from beads
 158 using a magnetic rack and remove completely. Air-dry the pellet for 2 min. Elute with 17 μ l of
 159 water, mix by inversion and spin down shortly. Incubate for 5 min. Separate beads using the
 160 magnetic rack. Transfer the supernatant to a fresh plate/tube.
- 161 5. Now add the individual sample indexes to the fragmented, adapter-ligated and purified DNA.
 162 First, prepare the Index Primer Mix (NEBNext Multiplex Oligos for Illumina, see kit protocol
 163 for preparation of index combinations) by mixing 5 μ l of the i5 primer and 5 μ l of the i7 primer
 164 with 15 μ l water. In a new plate/tube, attach the sample index by mixing 3.75 μ l DNA with
 165 2.5 μ l of the prepared Index Primer Mix (NEBNext Multiplex Oligos for Illumina, fold adapter
 166 dilution, see kit protocol for preparation of index combinations) and 6.25 μ l NEBNext Ultra II
 167 Q5 Master Mix. Cycling protocol: 98°C 30 sec, 7 cycles: 98°C 10 sec/65°C 75 sec, 65°C
 168 5 min.
- 169 6. Prepare pools of up to 48 samples (5 μ l each).
- 170 7. Purify and remove small fragments by adding 0.7x SPRIselect beads (e.g. 168 μ l beads to a
 171 pool of 240 μ l). Make sure the beads are suspended completely by vortexing. Mix by inversion
 172 and spin down only for a short time such that the beads remain dispersed. Incubate for 5 min.
 173 Separate beads on a magnetic rack. Discard the supernatant. Perform a wash step with 500-
 174 1000 μ l 80% ethanol leaving the tube on the magnetic rack. Discard the supernatant. Repeat the
 175 wash step. Spin down shortly to gather the ethanol at the bottom of the tube, separate the
 176 ethanol from beads using a magnetic rack and remove completely. Air-dry the pellet for 2 min.
 177 Elute with 30 to 75 μ l of water (depending on the number of samples in the pool). Mix by
 178 inversion and spin down shortly. Incubate for 5 min. Separate beads using the magnetic rack.
 179 Transfer the supernatant to a fresh tube.
- 180 8. Quantify the sequencing libraries using an appropriate method for Illumina sequencing (e.g.
 181 qPCR).
- 182 9. Load samples on an appropriate Illumina MiSeq flow cell (500 cycle kit) to achieve a coverage
 183 of 50,000 reads per sample and run according to the manufacturer's instructions (see note 5).

184 3.3 PacBio Library Preparation & Sequencing

- 185 1. Perform an indexing PCR to add barcodes to each sample. Mix 1-3 μ l of the initial PCR
 186 product (estimation, use thickness of band in gel-electrophoresis), 0.2 μ M barcode primers,
 187 0.4 mM dNTPs, 1x Advantage Genomic LA Buffer and 1.25 U Advantage Genomic LA
 188 Polymerase Mix in a 25 μ l PCR reaction. Cycling protocol: 94°C 1 min, 15 cycles: 98°C
 189 10 sec/68°C 12 min, 72°C 10 min. Verify amplification success and specificity by gel-
 190 electrophoresis (Figure 1).
- 191 2. Pool up to 96 samples. Use 20 μ l for samples with weak amplification and 5 μ l for samples
 192 with strong bands. (Use 25 μ l if only few samples are pooled.) Do not include samples that did
 193 not yield PCR products of the expected size.
- 194 3. Reduce the volume by mixing the pool 1:1 with AMPure PB Beads (see note 6). Briefly
 195 centrifuge and rotate the tube on a HulaMixer for 10-15 min. Briefly centrifuge and put the
 196 tube on the magnetic rack. Incubate for 2-3 min and remove the supernatant. Perform a wash
 197 step with 1000 μ l 80% ethanol leaving the tube on the magnetic rack. Discard the supernatant.

- 198 Repeat the wash step. Spin down shortly to gather the ethanol at the bottom of the tube,
 199 separate the ethanol from beads using a magnetic rack and remove completely. Air-dry the
 200 pellet for 2 min. Elute with 30 to 200 μ l of water (depending on the number of samples in the
 201 pool). Mix by inversion and spin down carefully. Incubate for 5 min. Separate beads using the
 202 magnetic rack. Transfer the supernatant to a fresh tube.
 203 4. Quantify the pool using Qubit.
 204 5. Use 2500 ng of DNA and follow the “Procedure & Checklist – Preparing HiFi SMRTbell
 205 Libraries using the SMRTbell Express Template Prep Kit 2.0” protocol (PacBio), starting at
 206 step “Repair DNA Damage”. For HLA-DRB1 amplicons, select a range from 8,000 to
 207 50,000 bp using the BluePippin system. Elute the final purification step in 40 μ l water if you
 208 want to perform another step of DNA damage repair (continue at step 6) or in 11 μ l PB Elution
 209 buffer to proceed at step 8.
 210 6. Another step of DNA damage repair can improve the results. Add 5 μ l Thermopol-buffer 10x,
 211 0.1 mM dNTPs, 1x NAD⁺ and 2x PreCR Mix to a final volume of 50 μ l. Tap the tube and spin
 212 down briefly. Incubate at 37°C for 15 min and at 4°C for 10 min.
 213 7. Add 50 μ l water and 60 μ l AMPure PB Beads (0.6x) and repeat the final purification step
 214 according to PacBio’s protocol. Elute in 11 μ l PB Elution buffer.
 215 8. Check the library by Qubit and DNA fragment sizing system.
 216 9. Sequence the libraries on a Sequel or Sequel II system. Up to 192 samples can be multiplexed
 217 on a SMRT Cell 1M.
 218

219 3.4 ONT Library Preparation & Sequencing

- 220 1. When using DNA isolated from buccal swabs, gel-electrophoresis bands from the initial PCR
 221 (3.1 step 2) are usually weak to medium. If this is the case, proceed directly to step 2. However,
 222 in case of very strong bands, dilute the PCR products 1:10 with water before proceeding to
 223 step 2 (see note 7).
 224 2. Perform an indexing PCR to add barcodes to each sample. Mix 1-3 μ l of the initial PCR
 225 product (estimation, use thickness of band in gel-electrophoresis), 0.2 μ M barcode primers,
 226 0.4 mM dNTPs, 1x Advantage Genomic LA Buffer and 1.25 U Advantage Genomic LA
 227 Polymerase Mix in a 25 μ l PCR reaction. Cycling protocol: 94°C 1 min, 15 cycles: 98°C
 228 10 sec/68°C 12 min, 72°C 10 min. Verify amplification success and specificity by gel-
 229 electrophoresis (Figure 1).
 230 3. Pool up to 96 samples. Use 20 μ l for samples with weak amplification and 5 μ l for samples
 231 with strong bands. (Use 25 μ l if only few samples are pooled.) Do not include samples that did
 232 not yield PCR products of the expected size.
 233 4. Clean the pool by adding 0.6x SPRIselect beads (e.g. 180 μ l beads to a pool of 300 μ l). Mix by
 234 inversion and spin down only for a short time such that the beads remain dispersed. Incubate
 235 for 5 min. Separate beads on a magnetic rack. Discard the supernatant. Perform a wash step
 236 with 800 μ l 80% ethanol leaving the tube on the magnetic rack. Discard the supernatant. Repeat
 237 the wash step. Spin down shortly to gather the ethanol at the bottom of the tube. Separate the
 238 ethanol from beads using a magnetic rack and remove the ethanol completely. Air-dry the
 239 pellet for 2 min. Elute with 30 to 200 μ l of water (depending on the number of samples in the
 240 pool). Mix by inversion and spin down carefully. Incubate for 5 min. Separate beads using the
 241 magnetic rack. Transfer the supernatant to a fresh tube.
 242 5. Quantify the pool using Qubit.

- 243 6. It is recommended to use the latest chemistry (V14 SQK-LSK114) together with a R10.4.1
 244 flowcell. Consequently, prepare the sequencing library following the “Ligation Sequencing Kit
 245 V14 (SQK-LSK114) protocol” by Oxford Nanopore Technologies. When eluting with Elution
 246 Buffer (EB), use the given option for high molecular weight DNA (incubation at 37°C instead
 247 of room temperature).
- 248 7. Sequence the library on a MinION or GridION system. Up to 192 samples may be multiplexed
 249 on one MinION flowcell.

250 3.5 Data Analysis: DR2S

251 DR2S (Dual Redundant Reference Sequencing) is an R package designed to facilitate generating full-
 252 length phase-defined haplotype sequences in reference quality from a combination of long reads and
 253 short reads. It can be installed and run in a Linux environment and depends on local installations of
 254 samtools, bwa (version 0.7.11 or later), minimap2, and IGV. The latest version of DR2S can be
 255 installed in R from <https://github.com/DKMS-LSL/dr2s> using the commands:

```
256 > install.packages("devtools") # if not already installed
257 > devtools::install_github("DKMS-LSL/DR2S")
```

258 For ease of use we recommend, however, to use the DR2S docker image provided at docker hub. The
 259 image can be loaded and used with the following commands:

```
260 $ docker pull dkmsls1/dr2s:<version>
261 $ docker run --rm -p 8788:8787 -e PASSWORD=<pwd> -it dkmsls1/dr2s:<version>
```

262 Make sure to replace <pwd> with a password of your choice and <version> with the latest
 263 version of the software available (1.1.21 at the time of writing). An RStudio-session can now be
 264 accessed in your browser at localhost:8788. Log into RStudio using the username rstudio and
 265 the password <pwd>. A detailed tutorial on how to use DR2S is shipped as an R package vignette
 266 “DR2S”. Example data to follow the tutorial are also shipped as part of the docker image.

267 Briefly, the following steps constitute a DR2S workflow:

- 268 1. Place the demultiplexed FASTQs from a short-read and a long-read experiment into two
 269 separate directories, respectively, following the naming convention:
 270 <sampleid>_<locus>_*.fastq.gz
- 271 2. Prepare a yaml-based configuration file (or provide the relevant information via R in an
 272 interactive session) that provides information on the paths to input and output, the sample ids,
 273 the expected HLA locus, the type of long reads (pacbio or nanopore) as well as a couple of
 274 workflow parameters. A default configuration file and detailed documentation of the various
 275 options are provided as part of the docker image.
- 276 3. In an interactive R session, initialize a DR2S driver object:
 277
 278 > library(DR2S)
 279 > dr2sDriver <- InitDR2S(readDR2SConf(<path/to/configFile>))
 280
- 281 4. Create an initial sample-specific consensus sequence by mapping the short reads to a generic
 282 reference:

```

283
284 > dr2sDriver <- mapInit(dr2sDriver)
285
286 5. Partition long reads into haplotype-specific read clusters based on heterozygous positions in the
287 initial mapping:
288
289 > dr2sDriver <- partitionLongreads(dr2sDriver)
290
291 6. Iteratively, map long-read clusters to generate haplotype-specific consensus scaffolds:
292
293 > dr2sDriver <- mapIter(dr2sDriver)
294
295 7. Assign short reads to haplotype clusters:
296
297 > dr2sDriver <- partitionShortreads(dr2sDriver)
298
299 8. Polish the long-read-derived scaffold consensus sequences using the more accurate short reads:
300
301 > dr2sDriver <- mapFinal(dr2sDriver)
302
303 9. Report the polished consensus sequences as well as diverse quality diagnostics to a report/
304 subdirectory.
305 > dr2sDriver <- report(dr2sDriver)
306
307 10. To aid visual inspection and correction of possibly remaining problems, the DR2S package
308 provides the functions checkAlignmentFile(), remapAndReport(), and
309 reportCheckedConsensus(). For detailed instructions on their usage within a generic
310 quality-checking and postprocessing workflow consult the R package vignette “DR2S”.

```

311 For further details about the D2RS algorithm and its performance see Reference (3).

312 3.6 Submission to IPD/IMGT-HLA: TypeLoader2

313 Once the polished HLA-DRB1 consensus sequences are generated, novel allele sequences are ready
314 to be submitted to the ENA and IPD/IMGT-HLA databases. While you could also follow the
315 instructions for manual submission on the databases’ websites, we developed TypeLoader2 for better
316 performance, quality assurance and convenience (5). For a stepwise instruction on how to install and
317 use TypeLoader2 see the chapter of Schöne et al. in this book.

318 4 Notes

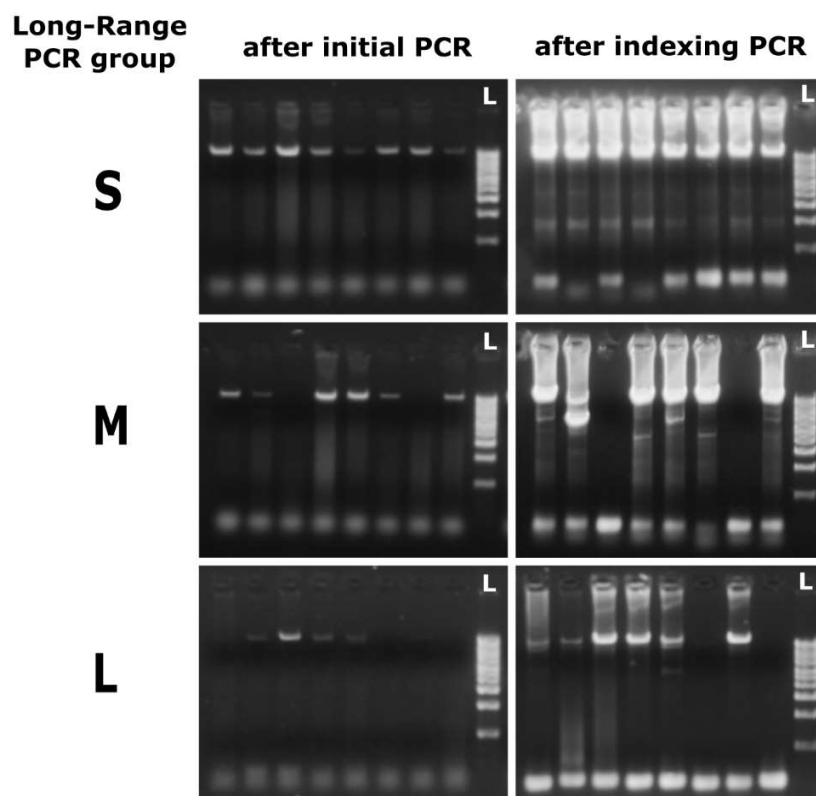
- 319 1. The overlap sequences are not required for short-read sequencing. However, since they do not
320 interfere, it is more convenient to use the same target-specific primers for both long- and short-
321 read sequencing.
- 322 2. Genomic DNA can be isolated from blood or buccal swabs. As DNA from buccal swabs is
323 usually more fragmented, higher PCR failure rates are to be expected. If in doubt, check DNA
324 quality and fragmentation with a suitable method before PCR.
- 325 3. PCR failure rates for 624 buccal swabs (DNA concentrations > 20 ng/μl, no prior
326 fragmentation analysis): 15%, 6%, 44% (long-range PCR groups S, M, L, respectively).
- 327 4. For long-read sequencing, weak bands are fine at this point as the initial PCR product is still
328 subjected to the barcode PCR. If insufficient PCR product is obtained for Illumina library

- 329 preparation, barcoding PCR (3.3 step 1) with any barcoding primers may be performed to
330 increase PCR product quantity before continuing the Illumina library preparation protocol.
331 5. Other Illumina instruments in combination with 300 cycle kits have not been tested but are
332 expected to provide similar results.
333 6. Washed SPRIbeads can be used instead of AMPure PB Beads.
334 7. Dilution ensures that every molecule is barcoded. Too much of the initial PCR in the barcoding
335 PCR will result in a high number of sequencing reads that cannot be demultiplexed. Try
336 diluting the PCR product even more if you experience this issue.

337 **5 References**

- 338 1. Robinson J, Barker DJ, Georgiou X, et al (2020) IPD-IMGT/HLA Database. *Nucleic Acids Res*
339 48:D948–D955
- 340 2. Gerritsen KEH, Groeneweg M, Meertens CMH, et al (2015) Full-length HLA-DRB1 coding
341 sequences generated by a hemizygous RNA-SBT approach. *Tissue Antigens* 86:333–342
- 342 3. Klasberg S, Schmidt AH, Lange V, et al (2021) DR2S: an integrated algorithm providing
343 reference-grade haplotype sequences from heterozygous samples. *BMC Bioinformatics* 22:236
- 344 4. Albrecht V, Zweiniger C, Surendranath V, et al (2017) Dual redundant sequencing strategy: Full-
345 length gene characterisation of 1056 novel and confirmatory HLA alleles. *HLA* 90:79–87
- 346 5. Schöne B, Fuhrmann M, Surendranath V, et al (2019) TypeLoader2: Automated submission of
347 novel HLA and killer-cell immunoglobulin-like receptor alleles in full length. *HLA* 93:195–202
348

349 6 Figures



350

351 Figure 1: Gel-electrophoresis control of eight exemplary samples of Long-Range PCR groups S, M
 352 or L after the initial PCR and after indexing PCR, respectively. DNA was isolated from buccal
 353 swabs. DNA ladder (lane L) ranges from 500 bp to 10 kb.

354

355 7 Tables

Long-Range PCR Group	Alleles	Length	Primer
S	DRB1*01/15/16	11 – 12 kB	F2 + R2 + R3
M	DRB1*03/08/10/11/12/13/14	13 – 14 kB	F1 + R2 + R3
L	DRB1*04/07/09	15 – 16 kB	F1 + R1

356 Table 1: HLA-DRB1 allele groups targeted by the respective primer combinations. Depending on the
357 targeted allele a particular set of primers is chosen for amplification resulting in an amplification
358 product of the given length.