PeerJ Preprints NOT PEER-REVIEWED

Preliminary characterization of little brown bats (*Myotis lucifugus*) immune MHC II DRB alleles using next-generation sequencing

Jon M. Palmer, Leah K. Berkman, Paula E Marquardt, Deahn M. Donner, Michelle A. Jusino, Daniel L. Lindner

Genes of the major histocompatibility complex (MHC) are of great interest to molecular ecologists due to their role in the immune response and the complex evolutionary mechanisms affecting their abundance and distribution. Next generation sequencing (NGS) offers the best potential for characterizing such genes by overcoming the low throughput constraints associated with traditional cloning and Sanger sequencing. However, available bioinformatics methods to effectively filter MHC data acquired through NGS are complex and have not been widely applied. Here we present a method to sequence MHC II DRB of the little brown bat (*Myotis lucifugus*) using semi-conductor sequencing on the Ion Torrent Personal Genome Machine (PGM) and subsequently analyze the data using a modified UPARSE clustering algorithm. As proof of concept, preliminary characterization of the MHC II DRB alleles of 15 little brown bats was accomplished through PGM sequencing. Analysis identified 24 MHC II DRB alleles overall and a high number of alleles per individual (average 6.8 ± 3.5, range 1 – 12), indicating little brown bats may have among the highest number of MHC DRB loci documented in mammals. Our method provides a framework for studying MHC II DRB diversity in little brown bats, which are experiencing significant population declines in North America due to the recent emergence of white-nose syndrome (WNS).

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016

1 **Preliminary characterization of little brown bats (*Myotis lucifugus*) immune MHC II DRB
2 alleles using next-generation sequencing.**

3 Jonathan M. Palmer[1], Leah K. Berkman[2], Paula E. Marquardt[3], Deahn M. Donner[3], Michelle A.
4 Jusino[1], Daniel L. Lindner[1]

5 [1] Center for Forest Mycology Research, US Forest Service, Madison, WI 53726, USA
6 [2] College of Environmental Science and Forestry, State University of New York, Syracuse, NY,
7 13210, USA
8 [3] Institute for Applied Ecosystem Services, US Forest Service, Rhinelander, WI 54501, USA
9
10 Corresponding author:
11 Paula Marquardt[3]
12 5985 Highway K, Rhinelander, WI, 54501, USA
13 Email address: pmarquardt@fs.fed.us
14

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016

15 **ABSTRACT**

16 Genes of the major histocompatibility complex (MHC) are of great interest to molecular
17 ecologists due to their role in the immune response and the complex evolutionary mechanisms
18 affecting their abundance and distribution. Next generation sequencing (NGS) offers the best
19 potential for characterizing such genes by overcoming the low throughput constraints associated
20 with traditional cloning and Sanger sequencing. However, available bioinformatics methods to
21 effectively filter MHC data acquired through NGS are complex and have not been widely
22 applied. Here we present a method to sequence MHC II DRB of the little brown bat (*Myotis*
23 *lucifugus*) using semi-conductor sequencing on the Ion Torrent Personal Genome Machine
24 (PGM) and subsequently analyze the data using a modified UPARSE clustering algorithm. As
25 proof of concept, preliminary characterization of the MHC II DRB alleles of 15 little brown bats
26 was accomplished through PGM sequencing. Analysis identified 24 MHC II DRB alleles overall
27 and a high number of alleles per individual (average 6.8 ± 3.5, range 1 – 12), indicating little
28 brown bats may have among the highest number of MHC DRB loci documented in mammals.
29 Our method provides a framework for studying MHC II DRB diversity in little brown bats,
30 which are experiencing significant population declines in North America due to the recent
31 emergence of white-nose syndrome (WNS).

32 **INTRODUCTION**

33 The multi-gene family of the major histocompatibility complex (MHC), called HLA in
34 humans, is of interest in molecular ecology due to its role in the activation of the adaptive
35 immune system (Hill 2001; Murphy et al. 2007) and the evolutionary mechanisms that generate
36 population level MHC variation (Spurgin & Richardson 2010; Sutton et al. 2011). MHC variants

37 can be associated with susceptibility or resistance to pathogens, which makes diversity of these
38 genes important for population viability (Dionne et al. 2009; Savage & Zamudio 2011).
39 Consequently, many studies of threatened, bottlenecked, or rapidly declining populations have
40 targeted the MHC to ascertain the loss of genetic diversity and adaptive potential of those
41 populations (e.g. Haig 1998; Hedrick et al. 2001; Miller et al. 2008; O'Brien et al. 1985).

42 Genotyping the MHC at the population level presents a challenge due to gene
43 duplications observed in many taxa (Kelley et al. 2005). Next generation sequencing (NGS)
44 provides the means to simultaneously analyze multiple MHC loci in many individuals, which
45 makes genotyping more tractable (Lighten et al. 2014). However, errors associated with
46 polymerase chain reaction (PCR), which is a requisite procedure for next generation sequencing
47 of amplicons (PCR products), can be problematic (Sommer et al. 2013). Additionally, primer
48 bias can cause preferential binding and an uneven degree of amplification among loci and alleles.
49 Many researchers have developed bioinformatics "pipelines", or a sequence of filters, to separate
50 PCR artifacts from real MHC alleles while attempting to avoid discarding low frequency alleles
51 (Babik et al. 2009; Galan et al. 2010; Pavey et al. 2013; Zagalska-Neubauer et al. 2010). These
52 pipelines may be effective for the taxon for which they were developed, but none have been
53 widely accepted or thoroughly tested (Lighten et al. 2014).

54 Studies of MHC diversity in bats are timely and of particular importance due to the recent
55 emergence of white-nose syndrome (WNS), a disease of bats caused by the fungus
56 *Pseudogymnoascus destructans* (Gargas et al. 2009; Lorch et al. 2011). This disease was first
57 observed in North American hibernating bats in 2006 (Blehert et al. 2009). The little brown bat
58 (*Myotis lucifugus*) is particularly susceptible and has suffered large-scale population decline as a
59 result of WNS (Blehert et al. 2009; Frank et al. 2014). However, very little is known about the

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016

60 immunological response of little brown bats to invasion by *P. destructans*. One way in which
61 bats and other mammals detect and respond to extracellular pathogens is through the variable
62 MHC II DRB locus, which encodes for an antigen-binding site responsible for sensing pathogens
63 and activating the adaptive immune system (Sommer 2005). Although the MHC Class II DRB
64 locus has been well-studied in other mammals, little is known about the genes involved in
65 disease resistance in bats, and the MHC Class II DRB is the only MHC gene that has been
66 examined to date (Del Real-Monroy et al. 2014; Mayer & Brunner 2007; Richman et al. 2010;
67 Schad et al. 2012). In this study we built a bioinformatics pipeline based on the Ion Torrent
68 Personal Genome Machine (PGM) sequencing platform and the UPARSE clustering algorithm
69 for studying the MHC II DRB locus in little brown bats (*Myotis lucifugus*). Our methodology
70 uses publically available and widely used software packages (Python and USEARCH) more
71 commonly used to construct *de novo* "taxa" (operational taxonomic units or "OTUs") in NGS
72 studies of microbial diversity (Caporaso et al. 2010; Edgar 2010; Edgar 2013; Schloss et al.
73 2009).

74 **MATERIALS AND METHODS**

75 Tissue samples were obtained from little brown bats (*M. lucifugus*) live-captured with
76 mist nets from summer roosting sites during June-September 2013 on the Chequamegon-Nicolet
77 National Forest in Wisconsin (n = 8) and the Superior National Forest in Minnesota (n = 7) by
78 US Forest Service employees as part of a larger habitat monitoring program. Because sampling
79 occurred on federal lands by federal employees on a species not federally listed, sampling
80 permits and Institutional Animal Care and Use Committee (IACUC) review were not required.
81 Samples were collected in a humane and ethical manner outlined by the American Society of
82 Mammologists for the safe handling and sampling of bats (Sikes et al. 2011). Mist nets were

83 operated by trained and experienced employees, at dusk into night, monitored continuously when
84 bats were active, and handled carefully when removing bats to avoid injury to animals. Tissue
85 was extracted from plagiopatagioum membranes with 4 mm sterile biopsy punches (Wilmer and
86 Barratt 1996). A biopsy was selected from a region of the wing with few or no visible blood
87 vessels to reduce bleeding; one biopsy was collected from each wing. Tissue was stored in 2.0
88 mL tubes and shipped on ice packs to the US Forest Service Northern Research Station in
89 Rhinelander, WI. To kill *P. destructans* or other potentially infectious agents, shipping
90 containers and contents were decontaminated with 10% bleach solution (National WNS
91 decontamination protocol v 06.25.2012;
92 https://www.whitenosesyndrome.org/topics/decontamination). Paper and any material that could
93 not withstand bleach treatment were UV irradiated for 10 minutes. Once received, samples were
94 stored at -20° C in 100% ethanol and DNA was extracted using the DNeasy® Blood and Tissue
95 Kit (Qiagen, Valencia, CA), following the manufacturer's protocols.

96 *Next-generation sequencing on the Ion Torrent PGM platform*

97 Using a combination of BLAST searches and Ensemble tools, 16 putative MHC II DRB
98 paralogs were identified from the little brown bat reference genome (Ensembl 73) (Flicek et al.
99 2013), and primers were designed in a conserved region of the genome surrounding exon 2
100 (Supplemental Figure S1). The forward primer was designed based on the EX2F primer used by
101 Richman et al. (2010) for other *Myotis* spp. and extended into exon 2. Although the area was
102 relatively conserved, 2 primers were designed that differed by 2 base pairs in an attempt to avoid
103 amplification bias. The reverse primer sequences from Richman et al. (2010) were not conserved
104 in little brown bats so a reverse primer was designed using a conserved block of sequence that
105 extended slightly into the other end of exon 2 of MHC II DRB locus. To generate amplicons that

106 are compatible with the Ion Torrent PGM platform, fusion primers were synthesized that
107 contained Ion P1 and Ion A adapter sequences; additionally, the reverse primer was barcoded
108 (according to protocols distributed from the manufacturer) to allow for multiplex sequencing.
109 The forward primers were synthesized as below (Ion P1 adapter is italicized): Mylu IntF_GC: 5'-
110 *CCTCTCTATGGGCAGTCGGTGAT*|TGTCCCCGCRGCGCATTTCCTG-3' and Mylu
111 IntF_AA: 5'-*CCTCTCTATGGGCAGTCGGTGAT*|TGTCCCCGCAGCAAATTTCCTG-3'. The
112 reverse primer contained the Ion A adapter (italicized), the key sequence, a 10 bp barcode
113 sequence (bolded), followed by a single 'A' linker sequence: Mylu Int2R: 5'-
114 *CCATCTCATCCCTGCGTGTCTCCGAC*|TCAG|**CTAAGGTAAC**|A|GGGTGCTCCTCACAGC
115 CCTGTG-3'.

116 MHC II DRB loci were amplified in two separate reactions with the reverse primer (Mylu
117 Int2R) and one of the forward primers (Mylu Int1F_GC or Mylu Int1F_AA). To minimize the
118 risk of PCR error from the use of multiple primers at the same site (Kanagawa 2003), each
119 reaction was run separately and the products were combined after PCR. All reactions contained
120 5U of Platinum *Taq* Polymerase (Invitrogen), 2 µL of 10X PCR Buffer, 0.4 µM each primer, 0.2
121 mM dNTPs, 2 mM $MgCl_2$ and 2-20 ng of template DNA in a total volume of 20 µL.
122 Thermalcycling conditions were: 94^0 C for 4 min followed by 28 cycles of 30 s at 94^0 C, 45 s at
123 58^0 C, and 90 s at 72^0 C. PCR products were visualized on a 1.4% agarose gel stained with
124 ethidium bromide and for each individual bat the two products were pooled equally according to
125 band intensity on the gel. The pooled PCR products were then gel-purified by 0.8% CloneWell
126 gels using the E-gel system and subsequently quantified with the Qubit High Sensitivity DNA
127 assay (Thermo-Fisher). The purified PCR products were pooled in equimolar concentrations,
128 attached to the Ion Sphere Particles (ISPs) using the Ion PGM Template OT2 400 kit, and

129 sequenced on a 314v2 chip using Ion PGM 400 bp Sequencing Kit according to manufacturer's
130 recommendations (Thermo-Fisher).

131 *Ion Torrent Data Processing*

132 Raw sequence data were obtained from the Ion Torrent Server using the BaseCaller
133 option [--disable-all-filters] to keep the barcode sequences of the reads intact and disable 3'
134 quality filtering in order to prevent trimming of the reverse primer sequence. The unaligned
135 BAM file from the Ion Torrent Server was converted to FASTQ using BedTools (Quinlan &
136 Hall 2010). The reads were then processed using custom python scripts that are available online
137 (https://github.com/nextgenusfs/mhc_cluster) (Figure 1). The first script, "mhc-
138 process_reads.py", functions to trim forward and reverse primer sequences from the read as well
139 as re-label the header with the corresponding barcode name. Because our primer design resulted
140 in sequencing from the opposite translation orientation of MHC II DRB exon 2, "mhc-
141 process_reads.py" also reverse complements (optional) each read for downstream processing
142 (Figure 1). The output of "mhc-process_reads.py" is a FASTQ file containing only those reads
143 that are full length (i.e. valid barcode and both primer sequences – with barcodes and primers
144 trimmed from the reads).

145 The second step of our pipeline is to run the "mhc-OTU_cluster.py script" (Figure 1).
146 This script runs a modified UPARSE algorithm that does the following: (1) quality filters the
147 data by removing reads that have an expected error > 1.0 (Edgar & Flyvbjerg 2015), (2) filters
148 reads using a DNA Hidden Markov Model of DRB-exon2 using HMMER version 3.1b
149 (hmmer.org) (DNA HMM model for little brown bats was created using a MUSCLE (Edgar
150 2004) alignment of the exon 2 region of the 16 MHC II DRB paralogs identified from *in silico*

151 interrogation of the genome reference), (3) trims or pads (with N's) each read to a set length, (4)
152 runs the UPARSE algorithm (Edgar 2013) using a default clustering identity setting of 99%, and
153 (5) optionally the script can translate the OTUs into amino acid sequences (Figure 1). The output
154 of "mhc-OTU_cluster.py" is composed of several files including: multi-FASTA file containing
155 OTUs, multi-FASTA file with translated OTUs, and an OTU table. In order to determine the
156 functional variation of the DNA based OTUs, the "mhc-OTU_cluster.py script" can 3-frame
157 translate each OTU to its corresponding amino acid sequence. The amino acid sequences are
158 then aligned to the MHC_II_beta HMM model (pfam00969) using HMMer3, the region that
159 produces a significant alignment is extracted for each OTU, and exact duplicates are
160 concatenated to output unique protein sequences. These python scripts are customizable; for
161 example, they can be used with primer sets for other MHC loci and can be easily adapted to
162 other organisms by updating the HMM models used for PCR contaminant filtering. USEARCH
163 version 8.0.1623 (Edgar 2010) was used for all the analysis in this manuscript and the reads were
164 trim/padded to a set length of 320 base pairs.

165 **RESULTS**

166 A total of 875,516 raw reads were obtained from the Ion PGM (314v2 chip) sequencing
167 run and after filtering sequences that did not have a valid barcode sequence or both
168 forward/reverse primers, 62,009 sequences remained. Low quality sequences were removed by
169 discarding reads with expected errors > 1.0 (Edgar & Flyvbjerg 2015), which resulted in the
170 retention of 21,057 reads. Preliminary processing of the data indicated that some of the sequence
171 data were derived from non-target contaminating sequences presumably due to non-specific
172 amplification during the initial PCR reactions. Therefore, a DNA HMM model was used to
173 quickly filter out contaminating sequences, and this resulted in removal of ~20% of the

174 sequences. The remaining 16,898 sequences were subsequently trimmed or padded to 320 base
175 pairs and clustered using UPARSE (Edgar 2013) at 99% identity. These parameters would allow
176 for sequences that have more than 4 differences across 320 base pairs to be clustered into
177 discrete OTUs. The number of reads per individual bat that passed all 3 quality control criteria
178 ranged from 1,165 – 15,556 (average ± SD: 4,133.9 ± 3,875.6). The UPARSE algorithm
179 produced 24 OTUs from the 15 individual bats analyzed in this study. We subsequently used
180 presence/absence of reads for each OTU to determine allelic diversity for each individual bat,
181 which ranged from 1 – 12 OTUs per individual bat (average ± SD: 6.8 ± 3.5) (Figure 2A).
182 Translation of these 24 OTUs to amino acid sequences indicated that 20 were unique across
183 MHC II DRB exon 2, indicating that there are at least 20 different functional variants of exon 2
184 from these 15 individual bats (Figure 2B). Three of the unique translated OTUs had internal stop
185 codons and thus represented either frame-shift mutations or non-functional alleles. Sequence data
186 is available for download thru the NCBI SRA under the SRP067348 accession number.

187 Interestingly, no MHC II DRB alleles were present in all 15 individuals sampled; the
188 allele with the highest abundance, MyLu-DRB_3, was found in 14 individuals. Moreover, 5
189 alleles with the lowest abundance were observed in single bats, which represents ~20% of the
190 alleles discovered (MyLu-DRB_8, 14, 17, 22, and 24). Three of the low frequency alleles were
191 found together in one individual bat (MN012; Figure 2A). Taken together with the range of
192 MHC II DRB alleles detected per individual, these data suggest that a larger sampling effort is
193 needed to adequately characterize the MHC II DRB alleles in little brown bats. However, despite
194 the small sample size, there is considerable variation in the amino acid sequence of the MHC II
195 DRB alleles discovered with our methodology (Figure 3).

196 **DISCUSSION**

197 We extended the use of an algorithm (UPARSE) designed to filter NGS data to the
198 analysis of MHC genes in little brown bats. Previous NGS MHC pipelines have used
199 complicated filtering parameters to differentiate between "artifact" sequences and "real"
200 sequences; however, UPARSE is very efficient at removing low quality reads through a
201 combination of error filtering and removal of low abundance reads (removing singletons) prior to
202 clustering into OTUs (Edgar 2013). Our method offers a conservative estimate of MHC II DRB
203 diversity, as sequences are clustered together based on a 99% identity threshold. This is notably
204 different from traditional Sanger sequencing methodology and other NGS MHC pipelines, where
205 sequences divergent by a single base pair have been referred to as an independent allele.
206 However, clustering highly similar MHC II DRB alleles may be more advantageous in
207 downstream analyses because genetic differences among highly similar alleles may not translate
208 into amino acid substitutions in the antigen-binding site, and thus these alleles would likely be
209 functionality identical in terms of their interactions with pathogens. For example, we identified
210 24 alleles on a DNA level using a clustering threshold of 99%, while 20 were unique on a protein
211 level.

212 There are some caveats to our technique, as amplicon sequencing on the Ion Torrent
213 PGM is currently limited to regions less than 400 base pairs and thus only allows for
214 characterization of small regions of the MHC II DRB locus (exon 2). As a result, functional
215 frame-shift mutations are difficult to properly characterize with this technique because the
216 sequences in the surrounding regions remain unknown. Moreover, in this study we have only
217 used a single priming site for amplification of MHC II DRB alleles and it has previously been
218 demonstrated that primers situated in conserved regions can be highly biased to a specific locus
219 or allele and can fail to effectively co-amplify all multigene copies (Burri et al. 2014). Given that

220 NGS approaches can sequence each PCR amplicon thousands of times, allelic dropout, or the
221 failure to detect an allele due to low throughput, is not as problematic with NGS protocols
222 compared to traditional Sanger sequencing approaches (Lighten et al. 2014). While we were
223 successful in this study with only a single priming site in amplifying MHC II DRB loci from
224 little brown bats, careful evaluation of additional conserved primers is warranted for future
225 studies.

226 The immune systems of bats remain poorly understood (Baker et al. 2013) and there are
227 few studies of bat MHC diversity (Del Real-Monroy et al. 2014; Mayer & Brunner 2007;
228 Richman et al. 2010; Schad et al. 2012). We have provided the first description of MHC II DRB
229 diversity among little brown bats. Little brown bats have a relatively high number of MHC II
230 DRB loci (up to 12 identified herein) compared to other mammals. Humans and other old world
231 primates have among the highest number of characterized MHC II DRB genes with 9, 4 of which
232 are functional (Klein et al. 2007). The number of DRB loci in bats varies considerably among
233 species (Schad et al. 2012). Richman et al. (2010) recovered only 2 – 4 unique sequences (1 – 2
234 loci) per bat in *M. velifer* and *M. vivesi*, while Schad et al. (2012) found 10 functional DRB loci
235 in an individual greater sac-winged bat (*Saccopteryx bilineata*). Due to our limited sample size
236 (n=15), a definitive characterization of MHC II DRB copy number is not attainable, but the
237 potential exists for little brown bats to have among the highest number of MHC DRB loci among
238 mammals and considerable MHC DRB diversity.

239 Because MHC is involved in the activation of the adaptive immune system, expansion of
240 MHC genes may have evolved to recognize and respond to a large number of pathogens (Hughes
241 and Nei 1989; Hedrick 1991). Little brown bats also have highly diverse immunoglobulin genes
242 (Bratsch et al. 2011), which indicates that an increased immune gene repertoire may be

243 characteristic of this species. Although such adaptations to efficiently recognize pathogens may
244 be highly developed in some species of bats (Schad et al. 2012), recent research suggests bats
245 may have features that suppress their immune response to infection (Baker et al. 2013; Wynne &
246 Wang 2013; Zhang et al. 2013). The precipitous decline of North American bats due to WNS has
247 prompted numerous investigations into how bats respond to this fungal pathogen (e.g., Meteyer
248 et al. 2011; Moore et al. 2011). One question that is particularly interesting is why North
249 American bats experience mass mortality when exposed to the fungus causing WNS, while
250 European bats do not experience mass mortality in response to this disease (Zukal et al. 2014).
251 Here we identified 20 unique MHC II DRB alleles at the amino acid level, which may be
252 indicative of a varied immune response to pathogens (perhaps WNS) among individual bats.
253 Low MHC diversity may be related to infection susceptibility and survival of bottlenecked
254 populations (Radwan et al. 2010); therefore, future characterization of MHC II DRB diversity in
255 North American little brown bat populations may offer insight into the susceptibility of these
256 bats to WNS.

257 **ACKNOWLEDGEMENTS**

258 We are grateful to Timothy Catton, David Grandmaison, Brian Heeringa, Kari Kirschbaum, and
259 Nicole Shutt (USFS) for assistance in obtaining samples. Anita Foss, Ben Prom, and Mark Banik
260 (USFS) provided laboratory support; Sue Lietz (USFS) delivered GIS support, while Jacqueline
261 Frair, ESF Roosevelt Wild Life Station (SUNY), contributed programmatic support.

262 **LITERATURE CITED**

263 Babik W, Taberlet P, Ejsmond MJ, and Radwan J. 2009. New generation sequencers as a tool for
264 genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology*

265 *Resources* 9:713-719. 10.1111/j.1755-0998.2009.02622.x
266 Baker ML, Schountz T, and Wang LF. 2013. Antiviral immune responses of bats: a review.
267 *Zoonoses Public Health* 60:104-116. 10.1111/j.1863-2378.2012.01528.x
268 Blehert DS, Hicks AC, Behr M, Meteyer CU, Berlowski-Zier BM, Buckles EL, Coleman JT,
269 Darling SR, Gargas A, Niver R, Okoniewski JC, Rudd RJ, and Stone WB. 2009. Bat
270 white-nose syndrome: an emerging fungal pathogen? *Science* 323:227.
271 10.1126/science.1163874
272 Bratsch S, Wertz N, Chaloner K, Kunz TH, and Butler JE. 2011. The little brown bat, *M.*
273 *lucifugus*, displays a highly diverse VH, DH and JH repertoire but little evidence of
274 somatic hypermutation. *Developmental & Comparative Immunology* 35:421-430.
275 10.1016/j.dci.2010.06.004
276 Burri R, Promerova M, Goebel J, and Fumagalli L. 2014. PCR-based isolation of multigene
277 families: lessons from the avian MHC class IIB. *Molecular Ecology Resources* 14:778-
278 788. 10.1111/1755-0998.12234
279 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena
280 AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE,
281 Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh
282 PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, and Knight R. 2010. QIIME
283 allows analysis of high-throughput community sequencing data. *Nature Methods* 7:335-
284 336. 10.1038/nmeth.f.303
285 Del Real-Monroy M, Martinez-Mendez N, and Ortega J. 2014. MHC-DRB exon 2 diversity of
286 the Jamaican fruit-eating bat (*Artibeus jamaicensis*) from Mexico. *Acta Chiropterologica*
287 16:301-314. 10.3161/150811014x687260

288 Dionne M, Miller KM, Dodson JJ, and Bernatchez L. 2009. MHC standing genetic variation and
289 pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal*
290 *Society of London Series B: Biological Sciences* 364:1555-1565. 10.1098/rstb.2009.0011
291 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
292 throughput. *Nucleic Acids Research* 32:1792-1797. 10.1093/nar/gkh340
293 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
294 26:2460-2461. 10.1093/bioinformatics/btq461
295 Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads.
296 *Nature Methods* 10:996-998. 10.1038/nmeth.2604
297 Edgar RC, and Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-
298 generation sequencing reads. *Bioinformatics*. 10.1093/bioinformatics/btv401
299 Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P,
300 Coates G, Fairley S, Fitzgerald S, Gil L, Garcia-Giron C, Gordon L, Hourlier T, Hunt S,
301 Juettemann T, Kahari AK, Keenan S, Komorowska M, Kulesha E, Longden I, Maurel T,
302 McLaren WM, Muffato M, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E,
303 Riat HS, Ritchie GR, Ruffier M, Schuster M, Sheppard D, Sobral D, Taylor K, Thormann
304 A, Trevanion S, White S, Wilder SP, Aken BL, Birney E, Cunningham F, Dunham I,
305 Harrow J, Herrero J, Hubbard TJ, Johnson N, Kinsella R, Parker A, Spudich G, Yates A,
306 Zadissa A, and Searle SM. 2013. Ensembl 2013. *Nucleic Acids Research* 41:D48-55.
307 10.1093/nar/gks1236
308 Frank CL, Michalski A, McDonough AA, Rahimian M, Rudd RJ, and Herzog C. 2014. The
309 resistance of a North American bat species (Eptesicus fuscus) to White-nose Syndrome
310 (WNS). *PloS One* 9:e113958. 10.1371/journal.pone.0113958

311 Galan M, Guivier E, Caraux G, Charbonnel N, and Cosson JF. 2010. A 454 multiplex
312 sequencing method for rapid and reliable genotyping of highly polymorphic genes in
313 large-scale studies. *BMC Genomics* 11:296. 10.1186/1471-2164-11-296
314 Gargas A, Trest MT, Christensen M, Volk TJ, and Blehert DS. 2009. *Geomyces destructans* sp.
315 nov. associated with bat white-nose syndrome. *Mycotaxon* 108:147-154.
316 doi:10.5248/108.147
317 Haig SM. 1998. Molecular contributions to conservation. *Ecology* 79:413-425.
318 Hedrick PW, Whittam TS, and Parham P. 1991. Heterozygosity at individual amino acid sites:
319 Extremely high levels for HLA-A and –B genes. *Proceedings of the National Academy of*
320 *Science USA* 88: 5897-5901
321 Hedrick PW, Parker KM, and Lee RN. 2001. Using microsatellite and MHC variation to identify
322 species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology*
323 10:1399-1412.
324 Hill AVS. 2001. The genomics and genetics of human infectious disease susceptibility. *Annual*
325 *Review of Genomics and Human Genetics* 2: 373-400.
326 Hughes AL, and Nei M. 1989. Nucleotide substitution at major histocompatibility complex class
327 II loci: evidence for overdominant selection. *Proceedings of the National Academy of*
328 *Sciences* 86:958-962.
329 Kanagawa T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR).
330 *Journal of Bioscience and Bioengineering* 96:317-323. 10.1016/S1389-1723(03)90130-7
331 Kelley J, Walter L, and Trowsdale J. 2005. Comparative genomics of major histocompatibility
332 complexes. *Immunogenetics* 56:683-695. 10.1007/s00251-004-0717-7
333 Klein J, Sato A, and Nikolaidis N. 2007. MHC, TSP, and the origin of species: from

334 immunogenetics to evolutionary genetics. *Annual Review of Genetics* 41:281-304.
335 10.1146/annurev.genet.41.110306.130137
336 Lighten J, van Oosterhout C, and Bentzen P. 2014. Critical review of NGS analyses for de novo
337 genotyping multigene families. *Molecular Ecology* 23:3957-3972. 10.1111/mec.12843
338 Lorch JM, Meteyer CU, Behr MJ, Boyles JG, Cryan PM, Hicks AC, Ballmann AE, Coleman
339 JTH, Redell DN, Reeder DM, and Blehert DS. 2011. Experimental infection of bats with
340 *Geomyces destructans* causes white-nose syndrome. *Nature* 480:376-378.
341 10.1038/nature10590
342 Mayer F, and Brunner A. 2007. Non-neutral evolution of the major histocompatibility complex
343 class II gene DRB1 in the sac-winged bat *Saccopteryx bilineata*. *Heredity* 99:257-264.
344 10.1038/sj.hdy.6800989
345 Meteyer CU, Valent M, Kashmer J, Buckles EL, Lorch JM, Blehert DS, Lollar A, Berndt D,
346 Wheeler E, White CL, and Ballmann AE. 2011. Recovery of little brown bats (*Myotis*
347 *lucifugus*) from natural infection with *Geomyces destructans*, white-nose syndrome.
348 *Journal of Wildlife Diseases* 47:618-626.
349 Miller HC, Miller KA, and Daugherty CH. 2008. Reduced MHC variation in a threatened tuatara
350 species. *Animal Conservation* 11:206-214. 10.1111/j.1469-1795.2008.00168.x
351 Moore MS, Reichard JD, Murtha TD, Zahedi B, Fallier RM, and Kunz TH. 2011. Specific
352 alterations in complement protein activity of little brown myotis (*Myotis lucifugus*)
353 hibernating in white-nose syndrome affected sites. *PloS One* 6:e27430.
354 10.1371/journal.pone.0027430
355 Murphy KM, Travers P, and Walport M. 2007. Janeway's immunobiology. Garland Science,
356 New York.

357 O'Brien SJ, Roelke ME, Marker L, Newman A, Winkler CA, Meltzer D, Colly L, Evermann JF,
358 Bush M, and Wildt DE. 1985. Genetic basis for species vulnerability in the cheetah.
359 *Science* 227:1428-1434.
360 Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire PA, Filteau M, Hebert
361 FO, Maaroufi H, and Bernatchez L. 2013. Nonparallelism in MHCIIbeta diversity
362 accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus*
363 *clupeaformis*) species pairs as revealed by next-generation sequencing. *Molecular*
364 *Ecology* 22:3833-3849. 10.1111/mec.12358
365 Quinlan AR, and Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
366 features. *Bioinformatics* 26:841-842. 10.1093/bioinformatics/btq033
367 Radwan J, Biedrzycka A, and Babik W. 2010. Does reduced MHC diversity decrease viability of
368 vertebrate populations? *Biological Conservation* 143:537-544.
369 10.1016/j.biocon.2009.07.026
370 Richman AD, Herrera ML, Ortega-Garcia S, Flores-Martinez JJ, Arroyo-Cabrales J, and
371 Morales-Malacara JB. 2010. Class II DRB polymorphism and sequence diversity in two
372 vesper bats in the genus *Myotis*. *International Journal of Immunogenetics* 37:401-405.
373 10.1111/j.1744-313X.2010.00941.x
374 Savage AE, and Zamudio KR. 2011. MHC genotypes associate with resistance to a frog-killing
375 fungus. *Proceedings of the National Academy of Sciences of the United States of America*
376 108:16705-16710. 10.1073/pnas.1106893108
377 Schad J, Dechmann DK, Voigt CC, and Sommer S. 2012. Evidence for the 'good genes' model:
378 association of MHC class II DRB alleles with ectoparasitism and reproductive state in the
379 neotropical lesser bulldog bat, *Noctilio albiventris*. *PloS One* 7:e37101.

380 10.1371/journal.pone.0037101
381 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
382 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, and
383 Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-
384 supported software for describing and comparing microbial communities. *Applied and*
385 *Environmental Microbiology* 75:7537-7541. 10.1128/AEM.01541-09
386 Sikes RS, Gannon WL, and Mammalogists AS. 2011. Guidelines of the American Society of
387 Mammalogists for the use of wild mammals in research. *Journal of Mammalogy* 92:235-
388 253. 10.1644/10-Mamm-F-355.1
389 Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology
390 and conservation. *Front Zool* 2:16. 10.1186/1742-9994-2-16
391 Sommer S, Courtiol A, and Mazzoni CJ. 2013. MHC genotyping of non-model organisms using
392 next-generation sequencing: a new methodology to deal with artefacts and allelic
393 dropout. *BMC Genomics* 14:542. 10.1186/1471-2164-14-542
394 Spurgin LG, and Richardson DS. 2010. How pathogens drive genetic diversity: MHC,
395 mechanisms and misunderstandings. *Proceedings: Biological Sciences* 277:979-988.
396 10.1098/rspb.2009.2084
397 Sutton JT, Nakagawa S, Robertson BC, and Jamieson IG. 2011. Disentangling the roles of
398 natural selection and genetic drift in shaping variation at MHC immunity genes.
399 *Molecular Ecology* 20:4408-4420. 10.1111/j.1365-294X.2011.05292.x
400 Wilmer JW, and Barratt E. 1996. A non-lethal method of tissue sampling for genetic studies of
401 chiropterans. *Bat Research News* 37: 1-3.
402 Wynne JW, and Wang LF. 2013. Bats and viruses: friend or foe? *PLoS Pathogens* 9:e1003651.

403 10.1371/journal.ppat.1003651
404 Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichon M, and Radwan J. 2010. 454
405 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex
406 in the collared flycatcher. *BMC Evolutionary Biology* 10:395. 10.1186/1471-2148-10-395
407 Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, Wynne JW, Xiong Z, Baker
408 ML, Zhao W, Tachedjian M, Zhu Y, Zhou P, Jiang X, Ng J, Yang L, Wu L, Xiao J, Feng
409 Y, Chen Y, Sun X, Zhang Y, Marsh GA, Crameri G, Broder CC, Frey KG, Wang LF, and
410 Wang J. 2013. Comparative analysis of bat genomes provides insight into the evolution
411 of flight and immunity. *Science* 339:456-460. 10.1126/science.1230835
412 Zukal J, Bandouchova H, Bartonicka T, Berkova H, Brack V, Brichta J, Dolinay M, Jaron KS,
413 Kovacova V, Kovarik M, Martinkova N, Ondracek K, Rehak Z, Turner GG, and Pikula J.
414 2014. White-nose syndrome fungus: a generalist pathogen of hibernating bats. *PloS One*
415 9:e97224. 10.1371/journal.pone.0097224

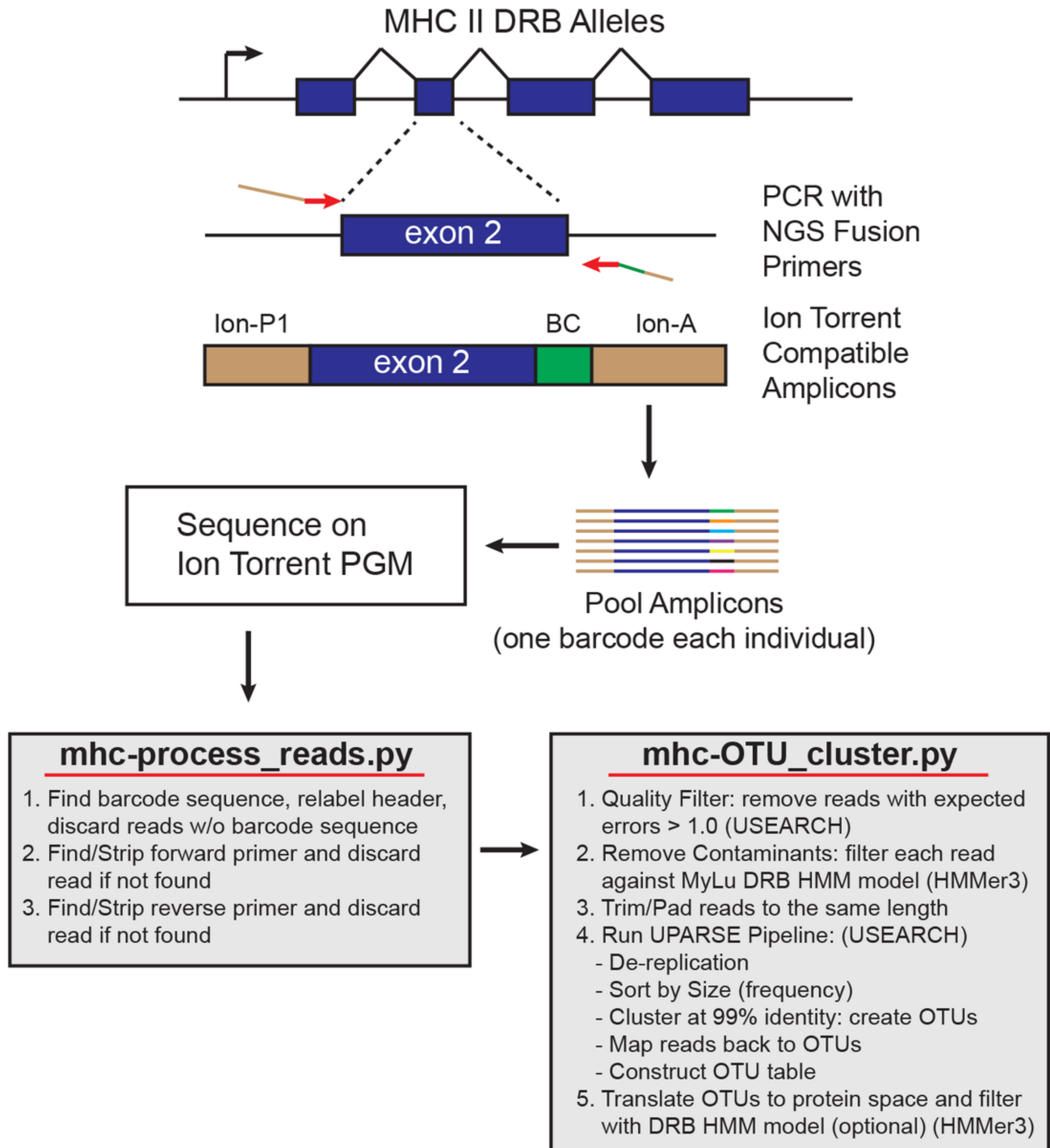
PeerJ Preprints NOT PEER-REVIEWED

1

Sequencing and clustering approach to MHC genotyping

Diagram of the methodology used to characterize MHC II DRB alleles from little brown bats using the Ion Torrent PGM semi-conductor sequencing platform. Barcoded Ion Torrent PGM compatible fusion primers are used to amplify a pool of PCR products from each individual bat (1 barcode per individual). The barcoded amplicons are then pooled in equal concentrations and sequenced on the Ion Torrent PGM. The raw sequences are obtained from the Torrent Server and processed with 2 python scripts. The first script, "mhc-process_reads.py", functions to find only those reads that contain a valid barcode sequence and contain both forward/ reverse primers. The second script, "mhc-OUT_cluster.py", quality trims the reads, filters the reads for contamination, and then runs the UPARSE clustering algorithm at a default setting of 99% identity. The result of the pipeline is multi-FASTA sequences for OTUs (DNA and amino acid) and an OTU table that can be used to genotype each individual using presence/absence metrics.

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016



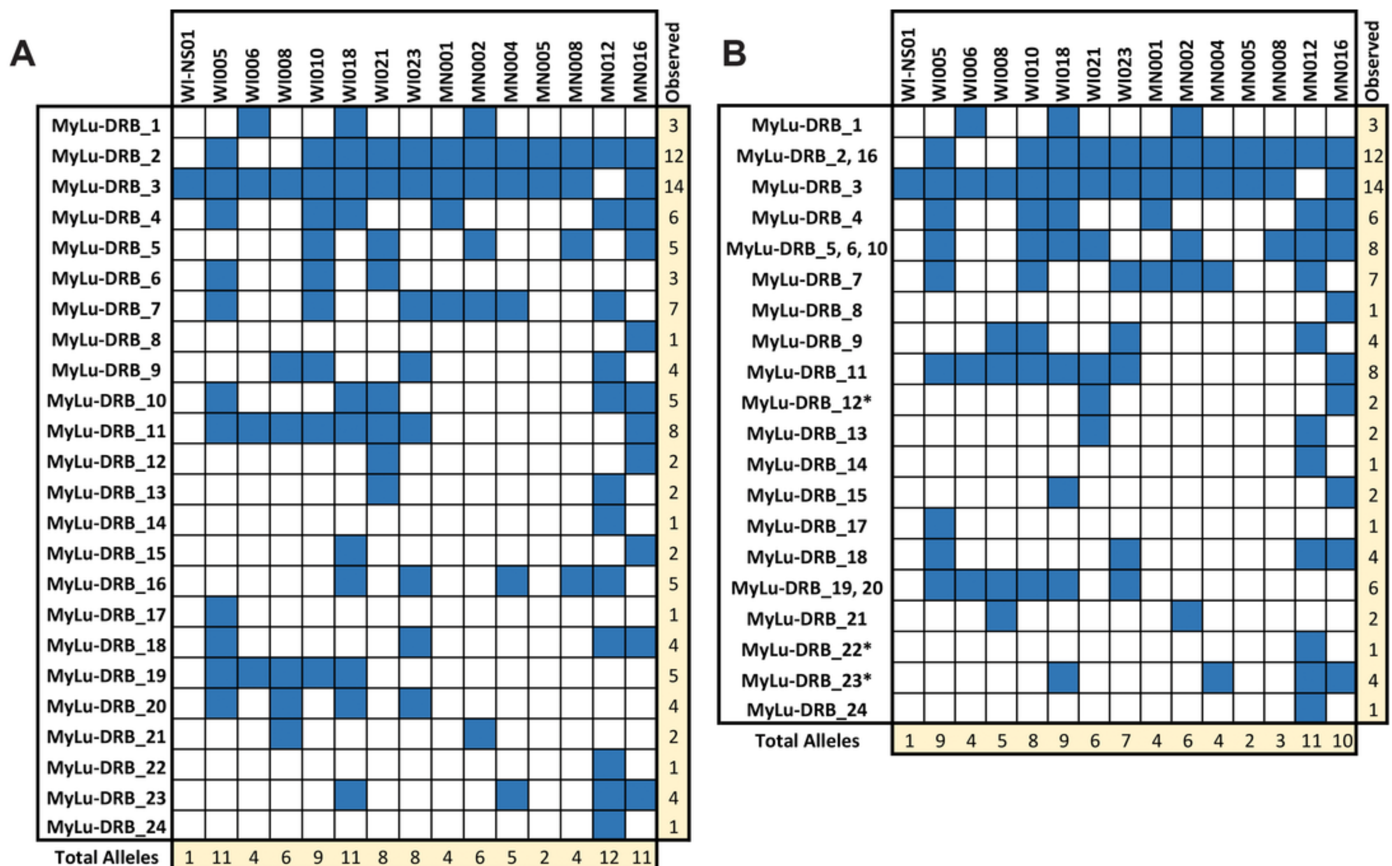
PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016

PeerJ Preprints NOT PEER-REVIEWED

2

DNA based Operational Taxonomic Unit Table

Figure 2: (A) Graphical representation of an OTU table using presence/absence metrics illustrates the distribution of DNA-based MHC II DRB alleles among the 15 little brown bats in this study (presence indicated by blue square). Twenty-four MHC II DRB OTUs were discovered with our method with the most abundant allele (MyLu-DRB_3) being found in 14 of 15 individuals from WI and MN. The range of alleles found per individual varied from 1 – 12. (B) Translation of the MHC II DRB OTUs to amino acid sequence resulted in 20 unique protein sequences, which have a similar distribution pattern per individual as the DNA-based OTUs. Sequences marked with an asterisks (*) were truncated at the amino acid level in exon 2 and thus representing either frame-shift mutations or are non-functional alleles.



PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1662v1 | CC-BY 4.0 Open Access | rec: 21 Jan 2016, publ: 21 Jan 2016

3

Protein sequence alignments

Figure 3. Multiple sequence alignment of protein sequences from the 20 MHC II DRB alleles illustrate s the functional variability seen in exon 2 from little brown bats. Sequences were aligned with MUSCLE and visualized using Geneious 7.1.8 (BioMatters). The Sequence Logo displays the level of conservation for each residue and completely conserved residues are colored gray.