



Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen

Lighten, Jackie; Papadopulos, Alexander S. T.; Mohammed, Ryan S.; Ward, Ben J.; Paterson, Ian; Ballie, Lyndsey; Bradbury, Ian R.; Hendry, Andrew P.; Bentzen, Paul; van Oosterhout, Cock

Nature Communications

DOI:

[10.1038/s41467-017-01183-2](https://doi.org/10.1038/s41467-017-01183-2)

Published: 01/01/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Lighten, J., Papadopulos, A. S. T., Mohammed, R. S., Ward, B. J., Paterson, I., Ballie, L., Bradbury, I. R., Hendry, A. P., Bentzen, P., & van Oosterhout, C. (2017). Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nature Communications*, [1294]. <https://doi.org/10.1038/s41467-017-01183-2>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 Evolutionary genetics of immunological supertypes reveals 2 two faces of the Red Queen

3

4 Jackie Lighten,¹ Alexander S.T. Papadopoulos,² Ryan S. Mohammed,³ Ben J. Ward,⁴ Ian
5 Paterson,⁵ Lyndsey Baillie,⁶ Ian R. Bradbury,⁷ Andrew P. Hendry,^{8,9} Paul Bentzen^{5*} and Cock
6 van Oosterhout^{1*}

7

8 *1. School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, Norfolk, UK*

9 *2. Molecular Ecology and Fisheries Genetics Laboratory, Environment Centre Wales, School of*
10 *Biological Sciences, Bangor University, Bangor LL57 2UW, UK*

11 *3. Department of Life Sciences, The University of the West Indies, St Augustine, Trinidad and*
12 *Tobago*

13 *4. Earlham Institute, Norwich Research Park Innovation Centre, Colney Lane, Norwich, NR4 7UZ,*
14 *UK,*

15 *5. Marine Gene Probe Laboratory, Department of Biology, Dalhousie University, 1355 Oxford*
16 *Street, Halifax, Nova Scotia, B3H 4R2, Canada*

17 *6. Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British*
18 *Colombia, V6T 1Z4, Canada*

19 *7. Fisheries and Oceans Canada, 136 Crosbie Road, St. John's, A1B 3K3, Newfoundland, Canada*

20 *8. McGill University, 859 Sherbrooke Street West, Montreal, H3A 0C4, Montreal, Quebec,*
21 *Canada*

22 *9. Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, H3A 0C4,*
23 *Montreal, Quebec, Canada*

24 ** Jointly supervising author*

25

26 Correspondence and requests for materials should be addressed to J.L. (email:

27 Jackielighten@gmail.com) and C.v.O. (email: c.van-oosterhout@uea.ac.uk).

28

29 **Editor's summary**

30 Host-parasite coevolution can lead to arms races favouring novel immunogenetic alleles or the
31 maintenance of diversity in a balanced polymorphism. Here, Lighten *et al.* combine data on MHC
32 diversity across three guppy species and simulations to show that polymorphisms of immunogenetic
33 supertypes may persist even as alleles within supertypes are involved in an arms race.

34 **Abstract**

35 Red Queen host-parasite co-evolution can drive adaptations of immune-genes by positive
36 selection that erodes genetic variation (Red Queen Arms Race), or result in a balanced
37 polymorphism (Red Queen Dynamics) and long-term preservation of genetic variation (trans-
38 species polymorphism). These two Red Queen processes are opposite extremes of the co-
39 evolutionary spectrum. Here we show that both Red Queen processes can operate
40 simultaneously, analyzing the Major Histocompatibility Complex (MHC) in guppies (*Poecilia*
41 *reticulata* and *P. obscura*), and swamp guppies (*Micropoecilia picta*). Sub-functionalization of
42 MHC alleles into “supertypes” explains how polymorphisms persist during rapid host-parasite
43 co-evolution. Simulations show the maintenance of supertypes as balanced polymorphisms,
44 consistent with Red Queen Dynamics, whereas alleles within supertypes are subject to positive
45 selection in a Red Queen Arms Race. Building on the Divergent Allele Advantage hypothesis, we
46 show that functional aspects of allelic diversity help to elucidate the evolution of polymorphic
47 genes involved in Red Queen co-evolution.

48

49 **Introduction**

50 Co-evolution is defined as the reciprocal, adaptive genetic changes between two or more
51 interacting species¹. Unlike adaptive evolution to the abiotic environment, during co-evolution
52 the selection pressures constantly change because adaptations in one species provoke counter-
53 adaptations in the co-evolving species. Rather than climbing a fitness peak in a nearly fixed
54 adaptive landscape, the landscape itself is evolving in response to selection by antagonistic
55 species climbing their respective fitness peaks. In host-parasite co-evolution, this amounts to

56 species gaining no fitness advance despite continuous adaptations, making antagonistic co-
57 evolution a zero-sum game. Van Valen² named this dynamic co-evolutionary process after the
58 character of the Red Queen in Lewis Carroll's *'Through the Looking Glass'*, who said: "... it takes
59 all the running you can do, to keep in the same place". Here, we investigate the population
60 genetic processes underpinning co-evolutionary change in the Major Histocompatibility
61 Complex (MHC) of three species of Poeciliid fish to understand Red Queen processes of host-
62 parasite co-evolution.

63

64 Co-evolution can result in stable or dynamic polymorphisms with cyclic or chaotic
65 fluctuations in allele frequencies, i.e. Red Queen Dynamics¹ (Fig.1a). The evolutionary force
66 operating to produce these fluctuations is balancing selection^{3,4}. In contrast, host-parasite co-
67 evolution can also result in the successive fixation of favourable alleles, i.e. the Red Queen
68 Arms Race¹ (Fig.1b). Underpinning the Arms Race is positive selection, and unlike Red Queen
69 Dynamics, genetic polymorphisms are transient with alleles rapidly replacing each other¹.
70 Genetic polymorphism for some immune-genes can be shared by species that have diverged by
71 millions of years, and this phenomenon is known as trans-species polymorphism (TSP)^{5,6}. TSP is
72 consistent with Red Queen Dynamics (balancing selection)⁷, but not with a Red Queen Arms
73 Race. These two Red Queen processes are opposite extremes of a co-evolutionary spectrum⁸,
74 and it is difficult to consolidate how both processes can operate simultaneously in the same
75 gene in a population⁹. According to Deborah Charlesworth, "*For MHC genes, frequently*
76 *observed high diversity and trans-specific polymorphism rule out a high turnover rate and, thus,*
77 *arms-race scenarios*"¹⁰.

78

79 TSP has been observed in numerous immune genes^{6,11,12}. A well-studied example of
80 balancing selection and TSP is the vertebrate MHC, in which genetic polymorphisms (allelic
81 lineages) can be preserved over millions of years. The most ancient case of TSP reported is for
82 MHC class I, with allelic lineages shared between paddlefish (*Polyodon spathula*) and Chinese
83 sturgeon (*Acipenser sinensis*); two fish species that diverged 187 million years ago¹³. The

84 strongest signal of TSP in the MHC is in the small exonic sections that encode the Peptide
85 Binding Region (PBR), which forms the adaptive interface of MHC proteins and bind the
86 epitopes of parasite antigens⁶. Remarkably, these codons experience the most intense positive
87 selection, evidenced by the elevated ratio of non-synonymous to synonymous substitutions
88 ($dN/dS > 1$)⁴. In addition, the MHC in large metapopulations is typified by allelic turnover¹⁴⁻¹⁶,
89 which is indicative of positive selection. Like the MHC, plant resistance genes (R genes) also
90 show both signals of selection. For example, R genes of different *Arabidopsis* species can evolve
91 rapidly though positive selection¹⁷, yet at the same time, the signal of TSP in these genes
92 suggests that balancing selection may help to maintain some functional variation^{17,18}.

93

94 To study Red Queen co-evolution, we genotyped MHC class IIb of 1,675 fish across three
95 species and two genera, guppies (*P. reticulata* and *P. obscura*) and swamp guppies (*M. picta*),
96 from 59 populations in 39 rivers/lakes across Trinidad, Tobago, Barbados and Hawaii. We
97 detected 539 alleles that were grouped in 15 supertypes (STs) based on the similarity in the
98 physicochemical properties of their PBR amino acids¹⁹. At the macro-evolutionary scale, these
99 STs were shared between genera diverged by >20 million years²⁰, and such TSP is strong
100 evidence of balancing selection and Red Queen Dynamics⁷. However, at the micro-evolutionary
101 scale, we observed large population genetic differentiation of the MHC alleles, which suggests
102 adaptive evolution by positive selection and a Red Queen Arms Race.

103

104 Here we build on the Divergent Allele Advantage (DAA) hypothesis²¹, which proposes that
105 diverged alleles are selectively favored, resulting in balancing selection that can maintain
106 genetic polymorphisms^{22,23}. We combine the DAA with contemporary “epitope space
107 theory”^{24,25}, classifying MHC alleles into STs with the aim to delineate different modes of
108 selection acting on the MHC (i.e. MHC alleles versus MHC STs). In addition, we present a
109 computer model to simulate the population genetics of MHC alleles and STs in an
110 epitope/paratope space. This concept is illustrated in Fig.2. The loss of an allele in a population
111 by positive selection acting on an alternative allele (or genetic drift) leads to a hole in the
112 paratope space. This hole can be exploited by parasites with matching epitope that are selected

113 to avoid host immune recognition. Novel allelic variants that can cover this paratope hole will
114 be introduced to the genepool by mutation, recombination, or migration, and these alleles will
115 be positively selected as they offer immune protection against the now common parasite strain
116 with the given epitope. Such “paratope holes” appear over space and time in host populations
117 because not all alleles (and STs) are present in an individual. In turn, the parasite’s epitope co-
118 evolves to exploit those holes. Our simulations show fluctuations in frequency of STs in
119 response to parasite-mediated balancing selection, consistent with Red Queen Dynamics.
120 However, at the same time, the extant immune alleles that constitute STs are replaced by bouts
121 of positive selection acting on novel alleles, a pattern predicted by the Red Queen Arms Race
122 (Fig. 2). Crucially, despite the transient nature of individual alleles, STs appear to be anchored in
123 the functional epitope/paratope space, resulting in TSP. We discuss how our study may be
124 relevant to other highly polymorphic immune/resistance genes involved in host-parasite co-
125 evolution in both animals and plants.

126

127 **Results**

128 **MHC alleles and supertypes**

129 To assess the relationship among immune alleles and supertypes (STs), we PCR amplified and
130 Illumina MiSeq sequenced the PBR of MHC IIb in 1,694 fish (*P. reticulata*, *P. obscura* and *M.*
131 *picta*) from 59 populations. Of these, 1,675 (98.87%) could be confidently genotyped using our
132 previously published workflow^{26,27}, resulting in 539 MHC IIb alleles (Supplementary Data 1). The
133 total number of alleles observed within an individual (A_i) ranged from one to nine (mean (\pm SD)
134 $A_i = 3.25 \pm 1.19$), with mean A_i varying among populations from 1.00 – 4.76 (Supplementary
135 Data 2). Among four sequencing runs, 233 replicate amplicons were sequenced across 103
136 individuals (2-4 independent PCR products were sequenced for each sample), and genotyping
137 repeatability was 99.83%.

138

139 To estimate functional diversity of the MHC, observed alleles were classified into STs
140 based on the shared physicochemical properties of the amino acids at the PBR which are under
141 positive selection¹⁹. Alleles clustered into 15 STs (Supplementary Fig. 1-2, Supplementary Data
142 1), and the number of STs within an individual (ST_i) ranged from one to seven (mean (\pm SD) =
143 2.79 ± 0.95 , Supplementary Data 2). Across all individuals ST_i and A_i were positively correlated
144 (Linear regression $P < 0.001$, $R^2 = 0.73$, Supplementary Fig. 3). In all but 16 individuals (<1%),
145 each ST was represented by a maximum of two alleles, suggesting that each ST is specific to a
146 single MHC IIb locus in guppies. Furthermore, given that there are more STs than there are
147 MHC IIb loci in guppies, it appears that alleles from multiple STs are segregating at the same
148 locus. To validate the robustness of estimates of functional diversity, we also inferred the
149 number of STs in a subsample for the dataset (820 individuals), comprising approximately 50
150 per cent of randomly drawn individuals per population. In this subsample, 407 alleles were
151 identified (76% of all alleles) that clustered into 15 STs (Supplementary Fig.1d), which supports
152 the robustness of well-defined functional clusters of alleles. The total number of alleles within
153 an ST ranged from 16 (ST-1) to 55 (ST-9) (mean 35.93 ± 11.79 , Supplementary Data 1 and 3). STs
154 were, on average, found in 17.75% (± 18.62) of individuals, however, ST-9 was observed in 81.64
155 % of individuals.

156

157 Alleles of ST-9 were present in 95% of populations. Other STs were present in 17 – 45
158 populations (29% - 76%). The number of ST-9 alleles within populations ranged from 1-15
159 (mean 4.38). Nine unique PBR sequences (10 alleles) observed in *M. picta* were distributed
160 among six STs shared with guppies (Supplementary Fig.2). Across both species, there was no
161 correlation between the number of unique ST-9 alleles and the total ST-9 frequency within
162 populations (Linear regression $P = 0.303$, $R^2 = 0.01$), or the number of unique ST-9 PBR amino acid
163 sequences and the total ST-9 frequency (Linear regression $P = 0.162$, $R^2 = 0.03$, Supplementary
164 Fig.4). However, the number of ST-9 alleles was significantly correlated with differences in the
165 number of microsatellite alleles between populations (Pearson correlation: $r = 0.65$; $P < 0.001$),
166 indicating that ST-9 alleles are subject to genetic drift.

167

168 For all STs the functional redundancy (S_r) of their alleles was calculated. S_r is defined as
169 the proportion of unique alleles with identical PBR amino acid sequences within a ST
170 (Supplementary Fig.5, Supplementary Data 3). We hypothesised that this redundancy is subject
171 to drift when present in the same gene pool. Across populations, however, this redundancy
172 would contribute to genetic differentiation. ST-9 displayed the highest redundancy ($S_r=3.23$),
173 with 55 alleles translating to 17 unique PBR amino acid sequences, and lowest level of amino
174 acid differentiation among PBR amino acid sequences (mean number of amino acids differences
175 among PBRs = 3.79). Similar degrees of redundancy and cumulative ST-9 frequency were
176 observed in *M. picta* (Supplementary Data 4). ST-6 displayed the lowest PBR redundancy (S_r
177 =1.22) and highest within ST differentiation (mean number of amino acids differences among
178 PBRs =7.75), while still comprising a relatively high number of unique alleles (44). Although ST-6
179 was observed in a range of populations across different geographic regions, it was notably more
180 common in southern Trinidad, and comparatively rare in the North Slope (Fig. 2, Supplementary
181 Fig.6). ST-3 was even more localized in distribution, with high frequencies (>0.20) only in North
182 Slope populations. In the North Slope, ST-3 was represented by just 16 alleles, with an overall S_r
183 of 1.455, which is less than half of that observed in ST-9 (3.235). Moreover, in each population,
184 one or two unique alleles tended to dominate the cumulative frequency of this ST suggesting
185 that allele frequencies of this ST are subject to positive selection, or drift because of functional
186 redundancy. In turn, this could explain why ST-9 alleles are strongly correlated to microsatellite
187 numbers. This pattern holds when analysed across all MHC alleles and all STs; genetic
188 differentiation (D_{est}) of microsatellites was significantly correlated with MHC alleles but not with
189 STs (see below).

190

191 To evaluate the evidence of TSP at the level of MHC alleles and STs, we examined two
192 species of guppy found in Caroni drainage/North Slope (*P. reticulata*) (n=790) and the
193 Oropouche drainage/north east Trinidad (*P. obscura*) (n=250)²⁸. Both species share 40 (31%)
194 alleles and 14 (93%) STs. The Caroni and Oropouche lineage share 28 (26%) alleles and 12
195 (80%) STs, despite at least 600,000 years of divergence with little to no gene flow²⁹. (In this

196 analysis, the introgressed Turure population³⁰ was excluded). With up to three guppy generations
197 per year this equated to an evolutionary divergence of over ~1,800,000 guppy generations. We
198 further examined MHC diversity shared between both guppy species (*P. reticulata* and *P.*
199 *obscura*) (n=1,313) and *M. picta* (n=5). Despite >20 million years of divergence²⁰ (~60 million
200 guppy generations), their MHC alleles fell into STs shared with both *P. reticulata* and *P.*
201 *obscura*. (Fig. 3, Supplementary Data 2). This is remarkable given that they did not have a single
202 allele in common (Jost's $D_{est} = 1$). Isolated, ecologically distinct guppy populations have few
203 alleles in common whilst sharing a large proportion of their STs (Fig. 3 and Supplementary
204 Fig.6). On average, each allele was observed in only 2.08 (\pm SD 2.81) (3.52%) populations and
205 321 out of the 537 alleles (59.7%) were private (i.e. observed in only a single population).
206 Conversely, each ST was found on average in 31.7 (\pm 11.2) (52.1%) populations, and there were
207 no private STs (Supplementary Data 2-4). Next we compared MHC diversity to microsatellite
208 diversity to quantify deviations in the MHC variation from the pattern expected under neutral
209 evolution. Populations were highly differentiated at microsatellite alleles ($D_{est} = 0.741 \pm$ SD
210 0.007), yet the MHC alleles showed an even higher level of population differentiation ($D_{est} =$
211 $0.88 \pm$ SD 0.003) (Supplementary Fig.7, Supplementary Data 5 and 6. See Supplementary Data 8
212 for microsatellite genotypes). From the total of 1225 pairwise comparisons between 50 guppy
213 populations, 1014 (82.8%) comparisons showed a higher level of differentiation for the MHC
214 alleles than for microsatellite alleles. When correcting for non-independence and including
215 each population only once, we found that on average only 4.3 comparisons out of 25
216 independent pairwise populations comparisons showed a higher D_{est} for microsatellite than for
217 MHC alleles (binomial test; $P=0.002$). The inflated level of genetic differentiation of MHC alleles
218 is consistent with the effects of positive selection acting on these immune alleles, resulting in
219 rapid evolutionary change and high population differentiation. In contrast, the population
220 differentiation based on STs ($D_{est} = 0.388 \pm$ SD 0.014) was significantly lower than that based on
221 both the microsatellites, and MHC alleles (binomial tests; $P = 7.83 \times 10^{-5}$, and $P = 1.75 \times 10^{-7}$,
222 respectively) (Supplementary Fig. 7, Supplementary Data 5-7). This suggests that balancing
223 selection is acting on ST variation, which homogenizes ST frequencies across populations. A
224 Mantel test with Holm P value correction revealed that D_{est} estimates of microsatellites and
225 MHC alleles were significantly correlated (correlation=0.10, $P=0.012$), yet microsatellite

226 differentiation was not significantly correlated with that of MHC supertypes (correlation=0.06,
227 $P=0.151$). Conversely, population differentiation estimates based on MHC alleles were highly
228 correlated with those of MHC STs (correlation=0.43, $P<0.001$). This further supports the
229 hypothesis that MHC allelic variation is significantly governed by demographic processes (e.g.
230 genetic drift), while MHC ST variation is less affected by such processes and under strong
231 balancing selection. Simulations show that the distribution of ST variation across populations is
232 not just an artifact of lumping alleles into groups (Supplementary Fig.8). In other words, the
233 simulations show that the observed ST distribution across populations is too uniform to be
234 explained by a random process such a genetic drift, but suggests that balancing selection is
235 homogenizing the ST diversity across the guppy populations in Trinidad, Tobago, Barbados,
236 Hawaii, and *M. picta*. Such uniformity in the frequency spectrum of STs across populations is a
237 hallmark of balancing selection³¹, and/or the presence of a consistent selection pressure (e.g. a
238 ubiquitous parasite). Gyrodactylus species are the most prevalent multicellular ecto-parasites in
239 natural guppy populations³², and although infections of these parasites have been correlated to
240 the MHC, the presence/absence of these worms was associated with a different ST^{31,33}.

241

242 Six pairwise ST combinations (out of a total of 105) show significant linkage
243 disequilibrium (LD) after Bonferroni correction, as evidenced by the relative excess of these STs
244 in individuals (T-test: $T\geq 3.64$, $p\leq 0.001$), whereas one other combination (ST1 and ST10) shows
245 evidence of repulsion (T-test: $T=-4.59$, $p<0.001$) (Supplementary Fig.9, Supplementary Table 1).
246 Although various processes can cause LD (e.g. demographic fluctuations, epistatic interactions,
247 Wahlund effects), the observation that the same STs are in LD across different populations and
248 species suggests that these STs are physically linked on the same haploblock. Furthermore,
249 there were significantly fewer than expected individuals with two copies of the same ST, i.e.
250 “homozygous STs” (one sample T-test; $T=-10.98$, $p<0.0001$), (Supplementary Fig.10). This
251 suggests that the polymorphism of STs is maintained by a form of balancing selection.

252

253 In summary, population genetic analysis show that ST variation is subject to balancing
254 selection, as is evidenced by (1) the sharing of STs among species (TSP), (2) the relative uniform
255 ST distribution across populations, (3) the lack in correlation with microsatellite differentiation,
256 and (4) the deficiency of “homozygous STs”. In contrast, the population dynamics of MHC
257 alleles is correlated with that of microsatellite alleles and seems to be governed at least in part
258 by drift, as well as the effects of positive selection. Indeed, the higher level of population
259 genetic differentiation of MHC alleles compared the (neutral) microsatellite alleles, and the
260 absence of allele sharing (despite ST sharing) between genera is consistent with positive
261 selection and local adaptations to parasites.

262

263 **Agent Based Model of co-evolution**

264 We built an Agent Based Model utilizing “epitope space theory”^{24,25} to study the evolution of
265 immune genes in a host-parasite system (see Methods). In these simulations, co-evolution led
266 to the formation and maintenance of eight equidistant STs in the finite epitope space (Fig. 4a-
267 b). (STs were not “pre-programmed” in the model; they appeared when the simulated
268 population approached a mutation- selection-drift equilibrium). We also simulated the null
269 model of no parasite selection ($s=0$), which resulted in just a single ST with one (or a few) alleles
270 that blink in-and-out of existence due to recurrent mutations. We observed significant
271 fluctuations in ST frequency in response to changes in parasite frequency over time (Fig. 4c),
272 consistent with balancing selection (Red Queen Dynamics). Importantly, successive frequency
273 peaks of a given ST can comprise different allele spectra (Fig. 4e, Supplementary Fig.11), which
274 is a pattern indicative of bouts of positive selection for alternative alleles (Red Queen Arms
275 Race). Nevertheless, despite allelic turnover of alleles within ST, the relative position of each ST
276 remained stable in the epitope/paratope space (Fig. 4a, b & d). A regression analysis shows that
277 the Euclidean distance between the centroids of STs did not change significantly over time in
278 populations that had attained a mutation-selection-drift equilibrium (Regression: $F_{1,23999}=0.52$,
279 $P=0.469$). Indeed, the changes in allele composition and frequencies within STs merely resulted
280 in a slight wobble in the position of STs in the epitope/paratope space when simulated over a
281 macro-evolutionary timescale (Fig. 4a & b, Supplementary Fig.13). Furthermore, the alleles

282 belonging to a given ST did not stray significantly from their ST centroid position over time
283 (Supplementary Fig.12). Hence, despite the turbulence of the Red Queen Arms Race, the ST can
284 become a TSP that coalesces deeply so that it might be shared between diverged species (Fig.
285 4a & b, Supplementary Fig.12).

286

287 Finally, to assess whether demographic processes significantly affect ST diversity, we
288 estimated the effect of strong genetic drift on both ST and allelic diversity by simulating
289 population bottlenecks (Fig. 4e). Simulations show that although drift can greatly reduce the
290 number of alleles, the number of STs remains comparatively constant. However, in populations
291 with a very small effective population size ($N_e=100$), the number of ST does go down, and each
292 ST is represented by only one (or very small number) of alleles.

293

294 Discussion

295 We genotyped MHC class IIb of 1,675 individuals of three species of guppies (*P. reticulata* and
296 *P. obscura*) and swamp guppies (*M. picta*) from 59 populations in 39 rivers/lakes across
297 Trinidad, Tobago, Barbados and Hawaii. We detected 539 alleles that could be grouped in 15
298 supertypes (STs) based on similarities of the physicochemical properties of their Peptide
299 Binding Region (PBR). The MHC IIb locus is commonly duplicated in vertebrates as a
300 requirement to increase the immunogenetic repertoire in light of the multiple parasites that
301 can infect an individual⁴. The Birth and Death³⁴, and Accordion Model³⁵ of multigene evolution,
302 as well as empirical data³⁶ suggest that copy number variation can evolve rapidly, which implies
303 that ancient whole genome duplication that occurred in teleost fish ~350 million years ago³⁷ is
304 likely to have little impact on MHC evolution. Hence, we argue that the findings of this study
305 are not affected by the whole genome duplication, and widely applicable for other vertebrates.
306 Remarkably, despite being completely differentiated in terms of their alleles, the STs were
307 shared between genera that are diverged by >20 million years²⁰. Such trans-species
308 polymorphism (TSP) is a hallmark of the MHC¹¹, and the evolutionary force maintaining this
309 diversity is balancing selection³⁸. However, at the micro-evolutionary scale, we observed large

310 genetic differentiation (expressed in Jost's D) at the MHC alleles (but not STs), and the level of
311 genetic differentiation even exceeded that of (neutral) microsatellites. Such a high level of
312 genetic differentiation is evidence of spatiotemporal variation in natural selection augmenting
313 the effects of genetic drift¹⁵. This interpretation is consistent with MHC studies on other
314 species; for example, populations of the New Zealand Hochstetter's frog (*Leiopelma*
315 *hochstetteri*) all shared the same MHC supertypes despite positive selection driving high
316 population differentiation at the MHC allelic level¹⁶. Altogether, these data lead us to the
317 following question: how can we explain the signal of both positive selection (rapid allelic
318 turnover) and balancing selection (TSP of STs) at the MHC?

319

320 Brockhurst *et al.*⁹ defined three broad classes of Red Queen co-evolution distinguished
321 by the modes of selection operating and the genetic architecture of co-evolving traits. Their
322 paper discusses how both balancing and positive selection occur within the Red Queen
323 framework, and they suggest that these modes of co-evolution often operate simultaneously in
324 different genes. These different modes of selection result in distinctly different allele frequency
325 dynamics (allele oscillations or recurrent bouts of positive selection)⁹. In our study, we
326 observed a relative uniform ST distribution across populations, as well as a deficiency of
327 "homozygous STs"; population genetic signatures consistent with balancing selection. On the
328 other hand, the population genetic differentiation of MHC alleles exceeded that of
329 microsatellites, which is consistent with spatiotemporal variation in selection and a high
330 turnover rate of alleles due to an arms race. However, the current opinion is that the TSP of the
331 MHC rules out a high turnover rate and arms-race scenarios¹⁰.

332

333 Consistent with Brockhurst *et al.*⁹, our computer model shows that both balancing
334 selection and positive selection can operate simultaneously, and remarkably, they can operate
335 in synchrony on a single gene (i.e. the MHC). In our simulations, the polymorphisms of alleles
336 cluster into groups or STs that each performs a distinct immunological function, i.e. a particular

337 area in the epitope/paratope grid. We hypothesize that the balanced polymorphism at each
338 MHC locus is generated by selection on alleles of different STs that are not functionally
339 equivalent. Because an allele of a given ST cannot perform the function of an allele belonging to
340 another ST, the two alleles cannot substitute each other at a locus without causing some
341 dysfunction that leads to a fitness cost. In other words, the replacement of alleles in recurrent
342 bouts of positive selection takes place within STs, whereas different STs are maintained by
343 balancing selection driven by the necessity to broadly cover the epitope space (when viewed at
344 the level of the population, not the level of an individual). The simulations show that despite
345 the rapid evolution of alleles, STs show very little net evolutionary change (Fig.4). This can be
346 understood when realizing that a significant shift of a ST in one direction (e.g. due to the loss of
347 one of its alleles) would expose a hole in the paratope between the STs at the population level
348 (Fig.2). Such holes can be exploited by parasites, allowing them to infect hosts more efficiently,
349 thus resulting in a reversal of this change, for example by a mutation, migration or
350 recombination that replaces the lost allele. Each ST thus “wobbles” in a given position in the
351 epitope/paratope space, which can explain the phenomenon of TSP, and the “trench warfare”
352 hypothesis in plant resistance genes¹⁸. Importantly, although we simulated a fixed epitope
353 space for simplicity, alterations in parasite community structure will result in dynamic change in
354 the shape and size of the epitope space. Rather than climbing a fitness peak in a nearly fixed
355 epitope space, the space itself is evolving rapidly in response to selection by antagonistic
356 parasites attempting to climb their respective fitness peaks.

357

358 Note that there is space between the alleles of different STs, which is not presented in
359 the simplified schematic of Fig. 2. However, in our computer simulation model, alleles are able
360 to bind parasites even if their epitope is not exactly matching the parasite’s paratope – and this
361 is in line with empirical data³⁹ and the DAA hypothesis²¹. In our model, the distance between
362 the allele and parasite in the epitope/paratope space is used to calculate the probability of
363 binding. Hence, even with the holes between the STs, parasites are bound by the alleles. If the
364 distance between STs would be larger, there would be more space for the parasites to exploit,

365 i.e. areas in the epitope/paratope grid where they would be bound less efficiently. In
366 shorthand, we have referred to this as “parasites exploiting the hole in the epitope space”. We
367 propose that this could explain the observation of TSP of STs.

368

369 Our empirical data and simulations are consistent with the Divergent Allele Advantage
370 (DAA) hypothesis of immunogenetic evolution²¹, and our computer simulations build on this.
371 Wakeland *et al.*²¹ hypothesised that the preservation of diverse allelic lineages reflects the
372 selective advantage of maintaining a broad spectrum of MHC functionality (heterozygote
373 advantage), which has been supported by simulations of binding prediction⁴⁰, and in the
374 genetics of natural populations⁴¹. Moreover, they hypothesised that balancing selection not
375 only operates on the presence/absence of STs (“immune void overdominance”) but also
376 separately on the alleles within each supertype (which in their terminology was referred to as
377 an “allelic lineage”). Our simulations refine the DAA model, showing that positive selection (and
378 genetic drift) operate on alleles independent of balancing selection operating on STs. As such,
379 we show that both Red Queen processes can operate simultaneously, even in a single locus.
380 Furthermore, Wakeland *et al.*²¹ observed that although the majority of differences among
381 alleles within each lineage were attributed to point mutations, recombination among variants
382 also contributed to variation among alleles^{21,42}. Note, however, that we did not simulate
383 recombination in our computer model (because that approach enabled us to define STs and
384 allocate alleles based on their co-ancestry). However, sequence exchange through
385 recombination between alleles of different STs could dramatically shift the paratope of the
386 recombinant alleles, more so than any single mutation in our model. Consequently, the MHC in
387 natural systems may be less preserved than in our simulations. Given that recombination
388 (including gene conversion and micro-recombination) is thought to play an important role in
389 MHC of some species^{36,42}, it would be interesting to examine the effects of recombination on
390 the effects of TSP in future simulation studies of MHC STs.

391

392 Our study assumes that the alleles of multiple STs segregate at a single MHC locus.
393 Some species, however, may possess loci with alleles belonging to just a single ST. Such species
394 cannot maintain a balanced polymorphism, but they can nevertheless be polymorphic for their
395 MHC if they possess multiple duplicated MHC loci. This genetic architecture is likely to benefit
396 species that undergo severe inbreeding. For example, the self-fertilizing fish, *Kryptolebias*
397 *marmoratus* was found to have 3.9 MHC STs per individual after more than 10 generations of
398 selfing in the laboratory, which was similar to the number found in the natural population⁴³.
399 Although MHC gene duplication can preserve the MHC polymorphism in the face of severe drift
400 and inbreeding, this genomic architecture may also incur a possible fitness cost, such as
401 delimiting T-cell diversity and reducing the efficiency of pathogen recognition⁴⁴.

402
403 The existence of MHC STs has been recognized by immunologists since the mid-1990s⁴⁵,
404 and our study expands the evolutionary genetic implications of such sub-functionalization. The
405 human MHC class I, or human leukocyte antigen (HLA) alleles are traditionally clustered and
406 defined into nine different STs, and although different methodologies have been employed to
407 classify STs, e.g.^{39,46}, the classification of alleles into STs is broadly consistent across these
408 methods. Each ST is characterized by a supermotif that reflects the broad main anchor motif.
409 The majority of HLA STs demarcate groups of alleles with non-overlapping repertoires, although
410 the binding repertoire does overlap in a small proportion of the alleles that span multiple STs³⁹.
411 Similarly, up to 62% of foreign peptides have been shown to be bound by more than one ST³⁹.
412 Nevertheless, despite some fuzziness in peptide binding and allele classification, the HLA ST
413 classification has been effectively used to identify T-cell epitopes from many disease targets,
414 and STs show specific disease associations, explaining variation in susceptibility and disease
415 outcome (reviewed in ref.³⁹). The population genetic patterns of STs infers that broad MHC/HLA
416 functionality has been driven by pathogen mediated selection not only in humans⁴⁷, but also in
417 populations of non-model vertebrates^{31,48,49}. Interestingly, when we employed a computer
418 simulation model to study the evolutionary genetics of the MHC, alleles also clustered into
419 groups with unique immunological function. These clusters resemble STs, and crucially, these

420 were not “pre-programmed” in the model, but they appeared because of the effects of
421 selection, mutation and drift simulated in the Agent Based Model. In other words, the
422 similarities in peptide binding specificities of the simulated alleles within STs were the result of
423 common ancestry and balancing selection, as has been found in the MHC ⁵⁰.

424

425 A puzzling observation about MHC gene evolution is that some studies report that MHC
426 diversity is primarily affected by drift⁵¹⁻⁵³, whereas others show that the high polymorphism is
427 maintained by balancing selection^{54,55}. This contradiction is reconciled when realising that
428 genetic drift (as well as selection) acts on the alleles, whereas balancing selection acts on the
429 immunological function of the alleles, defined by their ST. Indeed, a previous study of
430 Galápagos Mockingbirds inferred a significant effect of genetic drift on the number of alleles in
431 island populations but not the number of STs⁵⁶. Our simulations confirm that the effects of drift
432 are most noticeable after bottlenecks in populations with high allelic variation, when each ST is
433 represented by multiple (functionally more-or-less equivalent) alleles, like the guppy’s ST-9
434 alleles. Our simulations demonstrate that drift during a bottleneck can erode such allelic
435 diversity within STs with less effect on the number of STs. This is also supported by our
436 empirical data, which shows that although there are large differences in the number of MHC
437 alleles between populations (coefficient of variation, CV=0.60), the number of STs is more
438 similar (CV=0.38). In addition, the diversity of MHC alleles (but not STs) in populations is
439 correlated with microsatellite diversity, which shows how drift is a significant force at this level
440 of polymorphism, a finding echoed by other MHC studies (e.g. ref.⁵¹). These observations
441 support an important prediction of Wakeland *et al.*²¹, in that divergent allele advantage will
442 augment rare-allele advantage and protect rare-allelic lineages (or STs) from extinction.
443 Although the total number of alleles may be rapidly reduced in a ST by genetic drift, as it
444 becomes increasingly rare, this ST is likely to be saved from extinction by the functional benefits
445 it provides in recognizing pathogens that are adapted to avoid recognition by other, more
446 common STs.

447

448 Finally, our model can also explain why natural selection is unable to remove the large
449 number of disease-causing mutations in the human MHC (HLA) that result in over 100 heritable
450 disorders⁵⁷. Our empirical data show there is a significant deficiency of individuals with two
451 allelic copies of the same ST. As a consequence of this homozygote ST deficiency, recessive
452 deleterious mutations are rarely exposed to purifying selection, potentially resulting in the
453 buildup of a ‘sheltered load’ in a Muller’s Ratchet type process⁵⁸ in each ST lineage. We
454 hypothesize that each ST may thus accumulate a unique ‘sheltered load’ of recessive
455 deleterious mutations over time⁵⁸, which could explain the large number of heritable disorders
456 associated to the human MHC⁵⁷.

457

458 The implications of this study are likely to be relevant also to other genes that show high
459 levels of allelic polymorphism despite being involved in a Red Queen Arms Race. Prime
460 examples of such genes are Killer cell Immunoglobulin-like Receptor (KIR) genes, which show
461 extensive polymorphism⁵⁹, plant Resistance genes (R genes) that are engaged in “trench
462 warfare” and stuck in an evolutionary stalemate¹⁸, self-incompatibility S-loci in flowering plants
463 that display patterns of diversity consistent with Red Queen Dynamics and TSP⁶⁰, and possibly
464 also some avirulence and effector genes (see ref. ¹² for more examples). The identification of
465 sub-functionalized groups of alleles into STs is likely to help evolutionary genetic studies of such
466 genes, and we believe that by combining supertype theory and Wakeland *et al.*’s DAA
467 hypothesis, new light can be shed on complex observations associated with Red Queen co-
468 evolution in immune genes.

469

470 **Methods**

471 **Sampling**

472 Guppies (*Poecilia reticulata*, n genotyped=1,425 and *Poecilia obscura*, n genotyped=250) were
473 collected between 2008 and 2012 from 59 populations distributed among 39 rivers/stream and
474 a lake, across Trinidad, Tobago, Barbados, and Hawaii. Only guppies collected in the Oropouche

475 drainage and North East Trinidad were considered *P. obscura*²⁸. Each fish was euthanized in MS-
476 222, and then preserved in 100% ethanol. Individuals from one population of the related
477 swamp guppy (*Micropoecilia picta*, n=5) were also sampled from Trinidad. Fish were collected
478 with written approval from the Director of Fisheries Division, Ministry of Agriculture, Land and
479 Marine Resources, Trinidad and Tobago.

480

481 **Molecular methods**

482 DNA was extracted from 3 to 5 dried scales or pectoral fins using a glassmilk-binding protocol⁶¹.
483 Samples were genotyped at 10 polymorphic microsatellite loci using *P. reticulata* specific
484 primers^{62–64}. DNA was amplified via PCR in 5µl volumes comprising 10–50ng DNA, 0.5µl 10x
485 ThermoPol PCR buffer (20 mM Tris-HCl, 10 mM KCl, 10 mM (NH₄)₂SO₄, 0.1 % Triton X-100),
486 200µM dNTP, 200µM fluorescently labelled forward primer, 200µM reverse primer, and 0.5U
487 *Taq* DNA polymerase (New England BioLabs). PCR amplification consisted of the following: 4
488 min at 95°C, 30 cycles of 30s at 95°C, 30s at locus specific annealing temperature
489 (Supplementary Table 2), 30s at 72°C, and 3 min at 72°C. PCRs were carried out in Eppendorf
490 Mastercycler ep thermal cyclers. Microsatellite PCR products were visualized by electrophoresis
491 on 8% denaturing polyacrylamide gels run on a LI-COR IR2 DNA analyzer at 50°C. All gels
492 included positive control samples, redundant samples, and a molecular weight size standard
493 ladder. All analyses were conducted in the R statistical package⁶⁵ unless otherwise stated.
494 Microsatellite genotypes were checked using *Micro-Checker v.2.2.3*⁶⁶. All 10 loci were checked
495 for selection using *LOSITAN*⁶⁷, which suggested that one of the ten loci (Pret-46) did not
496 conform to expectations under neutrality. Moreover, inclusion of this locus resulted in
497 inferences of population structure that were bio-geographically implausible and contradictory
498 to those previously reported in guppies²⁹, and so was removed.

499

500 A 209-base pair (bp) fragment of MHC IIb, encompassing all but three codons predicted to
501 comprise the PBR was amplified using the degenerate primer pair DABdegFb–
502 GTGTCTTARCTCSHCTGARC⁶⁸, and DABdegRei–CTCACCTGATTAKYYAG²⁶. Each primer was
503 uniquely modified on the 5' end with a 10-bp multiplex identifier (MID; Roche Diagnostics

504 Technical Bulletin TCB No.005-2009), and samples were amplified using a unique combination
505 of forward and reverse MID-labelled primers, which allowed recovery of the amplicons per
506 individual after demultiplexing. PCRs contained 0.2 mM dNTPs (New England Biolabs),
507 0.5M forward primer, 0.5M reverse primer, 19 Phusion HF buffer, 6% DMSO, ~1–3 ng genomic
508 DNA and 0.4 U Phusion DNA Polymerase (Finnzymes). PCRs were performed in Mastercycler
509 Eppgradient S (96well), or ep384 thermocyclers (Eppendorf), using the following parameters:
510 98°C for 3 min; 30 cycles of 98°C 15 s, 57°C 40s, 72°C 60 s; 10 min at 72°C, then held at 10°C. PCR
511 amplicons were pooled and prepared for 150-bp paired-end Illumina MiSeq (Illumina, Inc., San
512 Diego, CA, USA) sequencing using the vendor's TruSeq library protocol.

513

514 To infer genotypes from MHC sequence we used ultra-deep sequencing and error-correction to
515 identify a sequencing breakpoint (or a Degree of Change - DOC) between the cumulative depth
516 distribution of alleles and sequencing artefacts²⁶. The approach assumes that sequence errors
517 attain significantly less sequencing depth than true alleles when compared within an amplicon.
518 This approach provided accurate and repeatable genotype estimates of co-amplified loci. The
519 application of ultra-deep sequencing is an effective approach to mitigate the effects of random-
520 allele drop out⁶⁹, where one allele may fail to be efficiently amplified. Indeed, we previously
521 observed very high genotyping repeatability (low allele drop-out) among samples within a
522 sequencing run (100%) and between sequencing runs (83.7%)²⁶. Moreover, repeatability in this
523 study rose to 99.83% when samples (including replicate PCRs) were sequenced among four
524 independent Illumina MiSeq sequence runs. This meant that very little random-allele drop out
525 occurred when using the same primers and ultra-deep sequencing as previously described²⁶.
526 We also previously showed that this genotyping approach and PCR primers produced no
527 detectable amplification and genotyping bias (which could impeded genotyping accuracy) even
528 below a total amplicon depth of 100x²⁶. We were therefore confident that the population level
529 variation in the total number of alleles per individual (or gene copy number variation) reflected
530 real biological processes and not genotyping artefact.

531 **Supertype classification**

532 We grouped MHC IIb alleles into functional supertypes (ST)s by analysis of amino acid
533 polymorphism at the guppy specific PBR²⁶. The PBR is the adaptive interface of pathogen
534 recognition leading to the host immune response⁷⁰, and it is under positive selection.
535 Therefore, PBR diversity should reflect functional differences among alleles. The PBR was
536 inferred previously from amino acids that showed an elevated posterior probability of positive
537 selection, based on dN/dS ratios under a Bayesian population genetics framework²⁶. The alleles
538 of all three-guppy species were pooled for PBR identification, and justified as follows: (1) many
539 alleles were shared between *P. reticulata* and *P. obscura*, (2) these species share the same
540 infecting parasite species, (3) the small number of *M. picta* samples did not allow for a separate
541 analysis, (4) we aimed to identify codons under selection among species, which would allow
542 phylogenetic analysis among homologous codons, and (5) experimentally validated
543 crystallography evidence confirmed shared PBR amino acid sites across taxonomically diverse
544 species e.g.⁷¹. The PBR of each allele was numerically characterized based on the
545 physicochemical properties of each amino acid¹⁹, based on five metric descriptors:
546 z1(hydrophobicity), z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects)⁷². We produced a
547 matrix with rows representing each allele and columns representing z1 to z5 for each amino
548 acid of the PBR, concatenated in sequence. Using this matrix, alleles were clustered by
549 Discriminant Analysis of Principle Components (DAPC) with the *adagenet* package^{73,74} in R⁶⁵.
550 Supertype classification avoided introducing missing data points into correlation analyses
551 where the presence of alleles varied greatly among populations, but the presence of STs were
552 more consistent (see Results). Bayesian Information Criterion (BIC) values were used to explore
553 different clustering solutions. The optimal number of supertypes was defined as the minimal
554 number of clusters after which the BIC increases as indicated by BIC values as a function of
555 cluster number⁷⁴. After identifying the optimal number of clusters we applied DAPC, and the
556 supertype clusters were visualized using the first two PCs. To validate the robustness of ST
557 inference we repeated the analysis for a subset of the samples, including ~50% of the
558 individuals per population (820 in total), comprising 407 alleles (76% of total), which also

559 confirmed 15 STs.

560

561 **Population and supertype genetic diversity**

562 Population differentiation was calculated using Jost's D , which is more appropriate than other
563 statistics given that MHC loci in guppies are highly diverse, exist in multiple copies, and because
564 the locus affiliation of alleles is unknown⁷⁵. This is the case with many MHC population genetic
565 studies and poses an issue when estimating population differentiation, as many software
566 programs require the designation of alleles to loci. The arbitrary allocation of alleles to loci may
567 severely bias population genetic estimates. As such we used custom scripts modified from
568 *SPADE* R package⁷⁶, which estimates Jost's D by comparing population level allelic frequencies
569 (i.e. allele pool unassigned to loci), which is appropriate when estimating population
570 differentiation in MHC datasets. Like F_{ST} , at complete differentiation, D_{est} equals unity, and D_{est}
571 equals zero when the allele frequencies of populations are identical. Differentiation was
572 estimated independently using MHC allele, MHC ST, and microsatellite allele frequencies. We
573 tested whether population differentiation estimates based on STs were lower because of
574 simply reducing the amount of diversity being compared among populations (i.e. a random
575 bioinformatic artifact) or reflected a real biological phenomenon driven by balancing selection.
576 Given the observation of 15 STs, we therefore randomly distributed the observed alleles in 15
577 artificial groups. In this procedure, the MHC allele genotypes remained the same as the
578 observed ones, but each individual now comprised a random ST genotype. We compared
579 estimates of ST population differentiation computed using the empirical data to the random
580 distribution of estimates. To achieve this, code was written which implements the following
581 procedure: For 1000 iterations, (a) clear the supertype designations across all alleles in the
582 empirical data, (b) without replacement, randomly reallocate the alleles in the total gene pool
583 to the same number of groups as observed STs, which receive the same number of total alleles
584 as observed in the empirical data. Each individual now comprised an MHC allelic genotype that
585 is unchanged from the empirical data but a random MHC ST genotype, (c) count the number of
586 individual randomized occurrences of each supertype present in each population. This results in
587 a matrix of supertype counts per population, where each row is a supertype, and each column

588 is a population, (d) calculate the pairwise Jost's D statistic for the matrix generated by step c,
589 and then the average of each column i.e. compute the mean Jost's D statistic for each
590 population.

591

592 Pairwise values of D_{est} were compared among microsatellites, MHC alleles, and MHC STs using a
593 Mantel test in the *ape* package⁷⁷ with 10,000 iterations, and Holm corrected *P*-values for
594 multiple comparisons. For each ST, we calculated (1) the total number of MHC alleles
595 (nucleotide variants), (2) the number of PBR amino acid sequences, (3) the mean distance
596 among PBR amino acid sequences (number of differences) within an ST, (4) the mean frequency
597 of an ST in populations, (5) the mean number of alleles within each ST per population, and (6)
598 the level of PBR redundancy among the constitutional MHC alleles within an ST (S_r). S_r was
599 calculated as the number of alleles within an ST divided by the number of unique PBR amino
600 acid sequences in the same ST. Sequence similarity and dendrograms of MHC alleles and PBR
601 sequences were inferred in *MEGA 5*⁷⁸, and edited in *FIGTREE*
602 (www.tree.bio.ed.ac.uk/software/figtree/).

603

604 We also analysed whether there was a relative deficiency of individuals with two copies of
605 the same ST ("homozygote ST"), which would be expected if STs were under balancing
606 selection. Given that 1600 out of 1675 individuals (99%) possessed two or fewer copies of the
607 same ST, we assumed for this analysis that STs are locus-specific. Also, we refer to individuals
608 with two copies of the same ST as guppies with a "homozygote ST" genotype. We calculated for
609 each ST the deviation between the observed and expected number of pairwise combinations of
610 "homozygote ST", and we summed this across all populations (n=55). On average, populations
611 possessed 7.96 STs, so the total number of comparisons was $55 \times 7.96 = 438$. The deviation
612 from zero was tested using a paired T-test.

613

614 **Linkage disequilibrium (LD) and Hardy Weinberg Equilibrium**

615 Because locus affiliations of MHC alleles and STs are not known when multiple duplicated loci
616 are co-amplified using degenerate primers, we were unable to use published software that
617 calculates linkage disequilibrium (LD) between pair-wise combinations of loci. Therefore, the
618 relative excess of associations between certain STs within individual genotypes was analysed to
619 examine evidence of LD, which is a statistically robust approach to detect patterns of
620 association. LD (i.e. an excess of observed combinations between certain STs relative to the
621 expected count) is indicative of physical linkage between loci on haploblocks, although
622 demographic effects and cryptic population substructure can also create significant LD.
623 Furthermore, balancing selection acting on two (or more) loci will also reinforce LD through
624 epistasis, as increased LD would reduce the segregation load. In our LD analysis, we calculated
625 the deviation between the observed and expected frequencies of all possible pairs of STs within
626 individuals of a population. Given that we identified 15 STs, the maximum number of pairwise
627 comparisons of two STs equals 105 ($N = (15 \times 14)/2 = 105$), although this number was generally
628 smaller because not all STs were present in each population. To calculate the observed
629 frequency of a ST pair, all pairwise combinations of those two STs within an individual were
630 counted, and this was summed across all individuals within the population. This value was then
631 divided by the total number of pairwise combinations of STs within all individuals of that
632 population. The expected frequency of a ST pair was calculated by first establishing the
633 frequencies of the two STs in the population, and then calculating the product of the two ST
634 frequencies. Finally, the deviation between the observed and expected frequency of each ST
635 pair was tested using a paired T-test. The p-value was corrected for multiple comparison using a
636 sequential Bonferroni correction. The deviations between the observed and expected
637 frequencies were visualized in XY-graphs, with the expected frequency of the ST pair on the X-
638 axis and the observed value on the Y axis. Values above the line $X=Y$ indicate a relative excess
639 (and hence, LD), and values below this line a shortage of the ST pair in a population (consistent
640 with repulsion).

641

642 **Computer simulations of immune gene evolution**

643 We developed an Agent Based Model to study the evolution of immune genes in a host-
644 parasite system, examining whether trans-species polymorphism (TSP) of STs can evolve in a
645 Red Queen Arms Race. Rather than using a strict population genetic model, in which alleles and
646 genotypes are assigned fitness values, this model was based on “epitope space theory”^{24,25} that
647 supports a finite epitope space in which parasite antigens and host immune recognition
648 molecules co-evolve. We analysed the adaptive evolutionary change in epitope recognition of
649 immune alleles and STs (i.e. their paratope) during host parasite co-evolution. We aimed to
650 construct the most basic model that (1) would result in antagonistic host-parasite co-evolution,
651 and (2) in which we could quantify the resulting adaptive evolutionary change in phenotype
652 over time. Hence, rather than using a strict population genetic model, we modelled the
653 paratope of immune alleles and the epitope of parasites in a 2-D grid with size 1000 x 1000,
654 which fits with current antigen/epitope modelling theory^{24,25}. The relative position of each
655 immune allele and the parasites in this space thus determines the selection coefficient acting
656 on each immune allele, and the fitness of an individual is proportional to the Euclidian distance
657 between the antigen and immune allele in the epitope/paratope space. The adaptive
658 evolutionary change in phenotype of alleles and STs was quantified by tracking changes in their
659 position within this space over a large period of evolutionary time. Analysing the phenotypic
660 change enabled us to study TSP. Furthermore, by analysing fluctuations in immune allele
661 frequencies, we could study the population genetic characteristics of the model.

662

663 The simulation began with both host and parasite alleles randomly distributed across
664 the epitope/paratopes space. Hosts were diploid with one or three immune loci. In the main
665 text, we show the results of a single locus model, and in Supplementary Fig.12 we show a 3-
666 locus model without recombination (i.e. loci were completely linked). Parasites were haploid,
667 and each host was infected by one parasite every generation. The minimum Euclidean distance
668 was calculated between an individual’s immune alleles and one randomly drawn parasite
669 representing the infection. Depending on this distance, the parasite was either recognized (in a
670 resistant host) or not (in a susceptible host). Fitness was relative so that 50% of all parasites
671 died (on resistant hosts). The other 50% of parasites (on susceptible hosts) reproduced clonally

672 one individual offspring. The epitope of parasite offspring mutated, causing a change in X or Y
673 coordinates by one unit within the grid. Parasite infection on the susceptible hosts reduced
674 host fitness by 0.25. (Hence, host generation time was ≥ 4 times longer than the parasite
675 generation time). Host with zero fitness died. Resistant host gained 0.25 fitness units, and
676 individuals with one fitness units reproduced offspring that all started with 0.25 fitness units.
677 Hosts reproduced sexually, producing gametes with one parental immune allele each. This
678 immune allele mutated with probability μ , which caused it to change its X or Y coordinates by
679 one unit within the grid. Shown are the results with a high mutation rate ($\mu=0.1$), which
680 effectively accelerates evolutionary time in the model. With $\mu=0.1$, the evolutionary time is
681 accelerated by a factor 3.1×10^6 , assuming a base mutation rate of 10^{-9} per base per
682 generation, and 16 PBR codons with a total of 32 replacement sites (i.e. the 1st and 2nd codon
683 positions of the PBR). Gametes of reproducing hosts united randomly to produce the next
684 generation of diploid offspring. This resulted in a Poisson distribution of offspring per parent
685 (mean=variance=unity).

686

687 Alleles were individually labelled at the start of the simulations, which enabled us to
688 track the ancestry of extant alleles and define STs. In the model, an ST is defined as all the
689 alleles that belonged to the same ancestral allele that was identified by its unique label, i.e. an
690 ST is a group of alleles that coalesced with each other at the start of the simulations. We opted
691 for this approach in our simplified model because without recombination between alleles, the
692 ancestry of alleles in our model is unambivalent, which enabled us to identify STs. The position
693 of alleles and parasites in the grid was recorded every time step, and the centroid of an ST was
694 found by calculating the mean X and Y coordinate of the alleles belonging to that ST. We
695 examined the evidence of TSP by analysing the rate of adaptive evolutionary change (i.e.
696 change in the position of centroid) of STs over time. For each ST, its nearest neighbour (after
697 the burn-in at generation $t=1000$) was determined. To study the population genetics of host-
698 parasite co-evolution, we also recorded the changes in frequency of immune alleles of one ST
699 over time. Finally, to study the effects of drift on allele and ST variation, we analysed the effect
700 of population bottlenecks, simulating host population sizes $N=10^4$, 10^3 and 10^2 , and recording

701 the number of alleles and STs present in the host population.

702

703 In nature, the epitope space faced by a host population is vast and constantly changing,
704 and in order to make it more tractable, we simulated a large but finite epitope space, which is
705 consistent with the current understanding of the antigen/epitope modelling theory^{24,25}. Setting
706 boundaries does not generate a “magic number” of stable supertypes; simulations show that
707 this number depends on the strength of parasite selection, the amount of genetic drift (and
708 effective population size), as well as the number of MHC loci simulated. This is also consistent
709 with divergent allele advantage²¹ and MHC evolution in nature; ST variation of HLA-B differs
710 geographically among (human) populations, suggesting that selection on STs reflects local
711 adaptation to different parasite communities⁴⁷. In addition, we did not simulate multiple
712 parasite infections, and therefore, natural systems may demonstrate more complex
713 interactions among host and parasite communities than simulated here.

714

715 **Code availability.** The code for host-parasite co-evolutionary simulations was developed in
716 Minitab and it is available from GitHub along with R scripts to perform supertype analyses
717 (https://github.com/Ward9250/Supertypes_RedQueen_TSE).

718

719 **Data availability**

720 The datasets generated during and/or analysed during the current study are available in the
721 NCBI data base (www.ncbi.nlm.nih.gov) in the form of MHC allelic sequences: accessions
722 KF321642.1 - KF321728.1 (PopSet: 544451456), and KT003989.1 - KT004363.1 (PopSet:
723 1033321404). All other data generated or analysed during this study are included in this
724 published article (and its supplementary information files).

725

726 **Ethical approval**

727 All applicable international, national, and/or institutional guidelines for the care and use of

728 animals were followed.

- 729 1. Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R.
730 Biological and biomedical implications of the co-evolution of pathogens and their hosts.
731 *Nat. Genet.* **32**, 569–577 (2002).
- 732 2. van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).
- 733 3. Spurgin, L. G. & Richardson, D. S. How pathogens drive genetic diversity: MHC,
734 mechanisms and misunderstandings. *Proc. Biol. Sci.* **277**, 979–88 (2010).
- 735 4. Bernatchez, L. & Landry, C. MHC studies in nonmodel vertebrates : what have we
736 learned about natural selection in 15 years ? *J. Evol. Biol.* **16**, 363–377 (2003).
- 737 5. Klein, J., Sato, A., Nagl, S. & O’huigin, C. Molecular trans-species polymorphism. *Annual*
738 *Review of Ecology and Systematics* **29**, 1–21 (1998).
- 739 6. Těšický, M. & Vinkler, M. Trans-Species Polymorphism in Immune Genes: General
740 Pattern or MHC-Restricted Phenomenon? *J. Immunol. Res.* **2015**, 838035 (2015).
- 741 7. Garrigan, D. & Hedrick, P. W. Perspective: Detecting adaptive molecular polymorphism:
742 Lessons from the MHC. *Evolution (N. Y.)* **57**, 1707 (2003).
- 743 8. Gandon, S., Buckling, A., Decaestecker, E. & Day, T. Host-parasite coevolution and
744 patterns of adaptation across time and space. *J. Evol. Biol.* **21**, 1861–1866 (2008).
- 745 9. Brockhurst, M. A. *et al.* Running with the Red Queen: the role of biotic conflicts in
746 evolution. *Proc. Biol. Sci.* **281**, 1–30 (2014).
- 747 10. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome
748 regions. *PLoS Genetics* **2**, 379–384 (2006).
- 749 11. Klein, J., Sato, A. & Nikolaidis, N. MHC, TSP, and the origin of species: from
750 immunogenetics to evolutionary genetics. *Annu. Rev. Genet.* **41**, 281–304 (2007).
- 751 12. Azevedo, L., Serrano, C., Amorim, A. & Cooper, D. N. Trans-species polymorphism in
752 humans and the great apes is generally maintained by balancing selection that modulates
753 the host immune response. *Hum. Genomics* **9**, 21 (2015).
- 754 13. Wang, D., Zhong, L., Wei, Q., Gan, X. & He, S. Evolution of MHC class I genes in two
755 ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*).

- 756 *FEBS Lett.* **584**, 3331–3339 (2010).
- 757 14. Fraser, B. A., Ramnarine, I. W. & Neff, B. D. Temporal variation at the MHC class IIb in
758 wild populations of the guppy (*Poecilia reticulata*). *Evol. Int. J. Org. Evol.* **64**, 2086–2096
759 (2010).
- 760 15. McMullan, M. & van Oosterhout, C. Inference of selection based on temporal genetic
761 differentiation in the study of highly polymorphic multigene families. *PLoS One* **7**, e42119
762 (2012).
- 763 16. Lillie, M. *et al.* Selection on MHC class II supertypes in the New Zealand endemic
764 Hochstetter's frog. *BMC Evol. Biol.* **15**, 63 (2015).
- 765 17. Wang, J., Zhang, L., Li, J., Lawton-Rauh, A. & Tian, D. Unusual signatures of highly
766 adaptable R-loci in closely-related Arabidopsis species. *Gene* **482**, 24–33 (2011).
- 767 18. Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. Dynamics of disease
768 resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* **400**, 667–71 (1999).
- 769 19. Doytchinova, I. A., Guan, P. & Flower, D. R. Identifying Human MHC Supertypes Using
770 Bioinformatic Methods. *J. Immunol.* **172**, 4314–4323 (2004).
- 771 20. Meredith, R. W., Pires, M. N., Reznick, D. N. & Springer, M. S. Molecular phylogenetic
772 relationships and the coevolution of placentotrophy and superfetation in Poecilia
773 (Poeciliidae: Cyprinodontiformes). *Mol. Phylogenet. Evol.* **59**, 148–157 (2011).
- 774 21. Wakeland, E. *et al.* Ancestral polymorphisms of MHC class II genes: divergent allele
775 advantage. *Immunol Res* **9**, 115–22 (1990).
- 776 22. Sommer, S. The importance of immune gene variability (MHC) in evolutionary ecology
777 and conservation. *Front. Zool.* **2**, 1–18 (2005).
- 778 23. Richman, a. Evolution of balanced genetic polymorphism. *Mol. Ecol.* **9**, 1953–63 (2000).
- 779 24. Recker, M. *et al.* The generation of influenza outbreaks by a network of host immune
780 responses against a limited set of antigenic types. *Proc. Natl. Acad. Sci. U. S. A.* **104**,
781 7711–7716 (2007).
- 782 25. Prechl, J. A generalized quantitative antibody homeostasis model: regulation of B-cell
783 development by BCR saturation and novel insights into bone marrow function. *Clin.*
784 *Transl. Immunol.* **6**, e130 (2017).

- 785 26. Lighten, J., van Oosterhout, C., Paterson, I. G., McMullan, M. & Bentzen, P. Ultra-deep
786 Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for
787 copy number variation in the guppy (*Poecilia reticulata*). *Mol. Ecol. Resour.* **14**, 753–767
788 (2014).
- 789 27. Lighten, J., van Oosterhout, C. & Bentzen, P. Critical review of NGS analyses for de novo
790 genotyping multigene families. *Mol. Ecol.* **23**, 3957–3972 (2014).
- 791 28. Schories, S., Meyer, M. K. & Scharl, M. Description of poecilia (acanthophaelus)
792 obscura n. sp., (teleostei: Poeciliidae), a new guppy species from western trinidad, with
793 remarks on p. wingei and the status of the ‘endler’s guppy’. *Zootaxa* 35–50 (2009).
794 doi:10.1007/s00705-010-0823-9
- 795 29. Willing, E.-M. *et al.* Genome-wide single nucleotide polymorphisms reveal population
796 history and adaptive divergence in wild guppies. *Mol. Ecol.* **19**, 968–84 (2010).
- 797 30. Shaw, P. W., Carvalho, G. R., Seghers, B. H. & Magurran, A. E. Genetic Consequences
798 Of an Artificial Introduction Of Guppies (*Poecilia reticulata*) In N-Trinidad. *Proc. R. Soc.*
799 *London Ser. B-Biological Sci.* **248**, 111–116 (1992).
- 800 31. Fraser, B. A. & Neff, B. D. Parasite mediated homogenizing selection at the MHC in
801 guppies. *Genetica* **138**, 273–278 (2010).
- 802 32. Stephenson, J. F., van Oosterhout, C., Mohammed, R. S. & Cable, J. Parasites of
803 Trinidadian guppies: evidence for sex- and age-specific trait-mediated indirect effects of
804 predators. *Ecology* **96**, 489–498 (2015).
- 805 33. Lighten, J. Elucidating patterns of Major Histocompatibility complex polymorphism in the
806 Trinidadian guppy (*Poecillia reticulata*) using Next Generation Sequencing. (Dalhousie
807 University, 2015).
- 808 34. Nei, M., Gu, X. & Sitnikova, T. Evolution by the birth-and-death process in multigene
809 families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7799–7806
810 (1997).
- 811 35. Klein, J, Ono, H., D, K. & C, O. U. The Accordion Model of MHC evolution. *Prog.*
812 *Immunol.* **8**, 137=143 (1993).
- 813 36. Eimes, J. a *et al.* Rapid loss of MHC class II variation in a bottlenecked population is
814 explained by drift and loss of copy number variation. *J. Evol. Biol.* **24**, 1847–56 (2011).

- 815 37. Glasauer, S. M. K. & Neuhauss, S. C. F. Whole-genome duplication in teleost fishes and
816 its evolutionary consequences. *Molecular Genetics and Genomics* **289**, 1045–1060
817 (2014).
- 818 38. Takahata, N. A simple genealogical structure of strongly balanced allelic lines and trans-
819 species evolution of polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2419–23 (1990).
- 820 39. Sidney, J., Peters, B., Frahm, N., Brander, C. & Sette, A. HLA class I supertypes: a
821 revised and updated classification. *BMC Immunol.* **9**, 1 (2008).
- 822 40. Lenz, T. L. Computational prediction of mhc ii-antigen binding supports divergent allele
823 advantage and explains trans-species polymorphism. *Evolution (N. Y.)*. **65**, 2380–2390
824 (2011).
- 825 41. Lenz, T. L., Wells, K., Pfeiffer, M. & Sommer, S. Diverse MHC IIB allele repertoire
826 increases parasite resistance and body condition in the Long-tailed giant rat
827 (*Leopoldamys sabanus*). *BMC Evol. Biol.* **9**, 269 (2009).
- 828 42. Spurgin, L. G. *et al.* Gene conversion rapidly generates major histocompatibility complex
829 diversity in recently founded bird populations. *Mol. Ecol.* **20**, 5213–25 (2011).
- 830 43. Ellison, a *et al.* Maintaining functional major histocompatibility complex diversity under
831 inbreeding: the case of a selfing vertebrate. *Proc. Biol. Sci.* **279**, 5004–13 (2012).
- 832 44. Vidović, D. & Matzinger, P. Unresponsiveness to a foreign antigen can be caused by self-
833 tolerance. *Nature* **336**, 222–225 (1988).
- 834 45. Sidney, J., Grey, H. M., Kubo, R. T. & Sette, A. Practical, biochemical and evolutionary
835 implications of the discovery of HLA class I supermotifs. *Immunol. Today* **17**, 261–266
836 (1996).
- 837 46. Hertz, T. & Yanover, C. Identifying HLA supertypes by learning distance functions. in
838 *Bioinformatics* **23**, (2007).
- 839 47. Dos Santos Francisco, R. *et al.* HLA supertype variation across populations: new insights
840 into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms.
841 *Immunogenetics* **67**, 651–663 (2015).
- 842 48. Sepil, I., Lachish, S., Hinks, A. E. & Sheldon, B. C. Mhc supertypes confer both
843 qualitative and quantitative resistance to avian malaria infections in a wild bird population.

- 844 *Proc. Biol. Sci.* **280**, 20130134 (2013).
- 845 49. Schwensow, N., Fietz, J., Dausmann, K. H. & Sommer, S. Neutral versus adaptive
846 genetic variation in parasite resistance: importance of major histocompatibility complex
847 supertypes in a free-ranging primate. *Heredity (Edinb)*. **99**, 265–77 (2007).
- 848 50. Sette, A. *et al.* Class I molecules with similar peptide-binding specificities are the result of
849 both common ancestry and convergent evolution. *Immunogenetics* **54**, 830–41 (2003).
- 850 51. Sutton, J. T., Nakagawa, S., Robertson, B. C. & Jamieson, I. G. Disentangling the roles of
851 natural selection and genetic drift in shaping variation at MHC immunity genes. *Mol. Ecol.*
852 **20**, 4408–4420 (2011).
- 853 52. Strand, T. M. *et al.* Can balancing selection on MHC loci counteract genetic drift in small
854 fragmented populations of black grouse? *Ecol. Evol.* **2**, 341–53 (2012).
- 855 53. Santonastaso, T. *et al.* The effects of historical fragmentation on major histocompatibility
856 complex class II β and microsatellite variation in the Aegean island reptile, *Podarcis*
857 *erhardii*. *Ecol. Evol.* **In press**, (2017).
- 858 54. Aguilar, A. *et al.* High MHC diversity maintained by balancing selection in an otherwise
859 genetically monomorphic mammal. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3490–4 (2004).
- 860 55. Van Oosterhout, C. *et al.* Balancing selection, random genetic drift, and genetic variation
861 at the major histocompatibility complex in two wild populations of guppies (*Poecilia*
862 *reticulata*). *Evol. Int. J. Org. Evol.* **60**, 2562–2574 (2006).
- 863 56. Vlček, J. *et al.* Balancing selection and genetic drift create unusual patterns of MHCII
864 β variation in Galápagos mockingbirds. *Mol. Ecol.* **25**, 4757–4772 (2016).
- 865
- 866 57. Miretti, M. M. *et al.* A high-resolution linkage-disequilibrium map of the human major
867 histocompatibility complex and first generation of tag single-nucleotide polymorphisms.
868 *Am. J. Hum. Genet.* **76**, 634–46 (2005).
- 869 58. van Oosterhout, C. A new theory of MHC evolution: beyond selection on the immune
870 genes. *Proc. Biol. Sci.* **276**, 657–65 (2009).
- 871 59. Middleton, D. & Gonzelez, F. The extensive polymorphism of KIR genes. *Immunology*
872 **129**, 8–19 (2010).

- 873 60. Llaurens, V. *et al.* Does frequency-dependent selection with complex dominance
874 interactions accurately predict allelic frequencies at the self-incompatibility locus in
875 *Arabidopsis halleri*? *Evolution* **62**, 2545–57 (2008).
- 876 61. Elphinstone, M. S., Hinten, G. N., Anderson, M. J. & Nock, C. J. An inexpensive and high-
877 throughput procedure to extract and purify total genomic DNA for population studies. *Mol.*
878 *Ecol. Notes* **3**, 317–320 (2003).
- 879 62. Watanabe, T., Yoshida, M., Nakajima, M. & Taniguchi, N. Isolation and characterization
880 of 43 microsatellite DNA markers for guppy (*Poecilia reticulata*). *Mol. Ecol. Notes* **3**, 487–
881 490 (2003).
- 882 63. Paterson, I. G., Crispo, E., Kinnison, M. T., Hendry, A. P. & Bentzen, P. Characterization
883 of tetranucleotide microsatellite markers in guppy (*Poecilia reticulata*). *Mol. Ecol. Notes* **5**,
884 269–271 (2005).
- 885 64. Shen, X., Yang, G. & Liao, M. Development of 51 genomic microsatellite DNA markers of
886 guppy (*Poecilia reticulata*) and their application in closely related species. *Mol. Ecol.*
887 *Notes* **7**, 302–306 (2006).
- 888 65. R Development Core Team, R. R: A Language and Environment for Statistical
889 Computing. *R Foundation for Statistical Computing* **1**, 409 (2011).
- 890 66. van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. Micro-checker:
891 Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol.*
892 *Notes* **4**, 535–538 (2004).
- 893 67. Beaumont, M. A. & Nichols, R. A. Evaluating Loci for Use in the Genetic Analysis of
894 Population Structure. *Proceedings of the Royal Society B: Biological Sciences* **263**,
895 1619–1626 (1996).
- 896 68. Llaurens, V., McMullan, M. & van Oosterhout, C. Cryptic MHC polymorphism revealed
897 but not explained by selection on the class IIB peptide binding region. *Mol. Biol. Evol.* **29**,
898 1631–1644 (2012).
- 899 69. Sommer, S., Courtiol, A. & Mazzoni, C. J. MHC genotyping of non-model organisms
900 using next-generation sequencing: a new methodology to deal with artefacts and allelic
901 dropout. *BMC Genomics* **14**, 542 (2013).
- 902 70. Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility

- 903 complex class I loci reveals overdominant selection. *Nature* **335**, 167–70 (1988).
- 904 71. Zhang, N. *et al.* Crystal Structure of Swine Major Histocompatibility Complex Class I SLA-
905 1*0401 and Identification of 2009 Pandemic Swine-Origin Influenza A H1N1 Virus
906 Cytotoxic T Lymphocyte Epitope Peptides ▽. *J. Virol.* **85**, 11709–11724 (2011).
- 907 72. Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M. & Wold, S. New chemical
908 descriptors relevant for the design of biologically active peptides. A multivariate
909 characterization of 87 amino acids. *J. Med. Chem.* **41**, 2481–91 (1998).
- 910 73. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers.
911 *Bioinformatics* **24**, 1403–5 (2008).
- 912 74. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a
913 new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94
914 (2010).
- 915 75. Jost, L. GST and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026
916 (2008).
- 917 76. Ma, K., Hsieh, T. & Chao, A. spadeR: species prediction and diversity estimation in R. R
918 package version 1.0. *spadeR: species prediction and diversity estimation in R* (2014).
919 Available at: <http://chao.stat.nthu.edu.tw/blog/software-download>.
- 920 77. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R
921 language. *Bioinformatics* **20**, 289–290 (2004).
- 922 78. Tamura, K. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis using Maximum
923 Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol.* **28**,
924 1530–4 (2011).

925

926 **Acknowledgements**

927 This research was funded by a British Ecological Society Large Research Grant [5875-6919]
928 awarded to JL, a Special Research Opportunity grant [356373-07] from the Natural Sciences and
929 Engineering Research Council of Canada (NSERC) awarded to AH and PB, a NSERC Discovery
930 Grant [249682-12] to PB, and grants from The Leverhulme Trust [RPG-2013-305],

931 and Biotechnology and Biological Sciences Research Council (BBSRC) [BB/N02317X/1] awarded
932 to CvO, as well as support by the Earth & Life Systems Alliance (ELSA). The authors thank Karl
933 Phillips for helpful comments on an earlier version of the manuscript.

934 **Author contributions**

935 Jackie Lighten and Cock van Oosterhout conceived the study, developed the model and wrote
936 the paper. DNA lab work was conducted by Ian Paterson and Lyndsey Baillie. Alexander S.T.
937 Papadopoulos, Ryan S. Mohammed, Ben J. Ward, Ian R. Bradbury, Andrew P. Hendry and Paul
938 Bentzen provided samples/analytical tools/reagents. All authors reviewed and agreed on the
939 final manuscript.

940 **Competing financial interests.**

941 The authors declare no competing financial interests.

942

943 **Figure Legends**

944 **Figure 1. Allele frequency changes driven by Red Queen co-evolution.** (a) Red Queen
945 Dynamics: A dynamic equilibrium of co-evolution acting on host immune alleles (blue) and
946 pathogen virulence alleles (red). This is maintained by balancing selection acting on existing
947 polymorphism over relatively short evolutionary time scales. (b) Red Queen Arms Race:
948 recurrent bouts of positive selection in host and pathogen operating on new polymorphisms
949 arising through mutation (adapted from¹).

950

951 **Figure 2. Diagrammatic representation of adaptive evolution of immune supertypes.**

952 Supertypes (STs) evolve an epitope/paratope space within a population that results in a
953 balanced polymorphism during Red Queen co-evolution. For simplicity in depicting interactions,
954 we visualize all alleles present in the population. Depending on the number of MHC loci,
955 individuals possess just a subset of these alleles and STs. Immune alleles (dots) that are
956 phylogenetically related are presented in coloured networks (blue, green, red and orange).
957 Alleles of the same ST cover an area in the epitope/paratope space depicted by coloured

958 ellipses. (Non-focal STs are depicted in grey) **(a)** STs in a gene pool have evolved to cover the
 959 entire epitope space with little overlap. **(b)** The loss of one allele from the population (e.g. due
 960 to drift or positive selection on an alternative allele) opens a hole in the paratope space (black
 961 area) that becomes exploited by parasites with matching epitope. **(c)** Selection favours new
 962 alleles with a paratope that covers the hole, but only rarely are these substitutions made by
 963 alleles from a different ST (red allele covering the hole left by the loss of the blue allele). This
 964 causes the STs to ‘wobble’ in the epitope/paratope space. Nevertheless, changes in the
 965 paratope of STs are restricted by the presence of neighbouring STs, effectively resulting in a
 966 form of balancing selection. Hence, STs remain conserved over evolutionary time, despite the
 967 Red Queen Arms Race and the high turnover of their constituent alleles (see Results).

968

969 **Figure 3. The geographic distribution of MHC supertypes in guppy populations across Trinidad**

970 **and other oceanic islands.** Rivers in the mountainous Northern Range comprise three major
 971 regions: The North Slope (green), Caroni Drainage (light blue), and Oropouche drainage
 972 (orange). Separate drainages in the Northern range are shown in grey. Rivers in the relatively
 973 flat regions towards the south are shown in dark blue. See Supplementary Fig.6 for abbreviated
 974 population names and populations in region B. ST9 is observed across 95% of populations and
 975 maintained in similar frequencies, despite wide variation in the frequencies and presence of the
 976 55 ST-9 alleles. Importantly, *Micropoecilia picta* shares STs with the guppy, despite allelic
 977 divergence. ST-9 is similarly high in frequency in *M. picta*, as well as on Barbados and Hawaii,
 978 where (except for Tobago) guppies were introduced by humans. Scale – 30km.

979

980 **Figure 4. Computer simulations of host parasite co-evolution.** The centroid position of each
 981 supertype (ST) is indicated by the black dots and red crosses for parasites. Shown are
 982 simulations of a one locus model with $N=10,000$ host population size, and mutation rate $\mu=0.1$.
 983 Eight STs evolve. **(a)** Each ST shows little change within the epitope/paratope space (XY
 984 coordinate) over time (Z-axis), which is consistent with trans-species polymorphism. **(b)** A two-
 985 dimensional view shows that STs “wobble” in epitope/paratope space over time. **(c)** A dynamic

986 co-evolutionary equilibrium between ST and parasite frequencies is driven by negative
987 frequency dependent selection, consistent with the Red Queen Dynamics. **(d)** The number of
988 alleles is more strongly affected by demographic processes (population sizes of $N=10^4$, 10^3 , and
989 10^2) than the number of STs. **(e)** Frequencies of alleles within a single ST show a classical Red
990 Queen Arms Race with alleles replacing one another over time, consistent with recurrent bouts
991 of positive selection. STs remain preserved over evolutionary time, despite the Red Queen Arms
992 Race and the high turnover of their constituent alleles (see Results, *cf.* Divergent Allele
993 Advantage hypothesis by Wakeland *et al.*²¹).

994

995