



HAL
open science

Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird

Hila Levy, Steven Fiddaman, Juliana Vianna, Daly Noll, Gemma Clucas, Jasmine Sidhu, Michael Polito, Charles Bost, Richard Phillips, Sarah Crofts, et al.

► To cite this version:

Hila Levy, Steven Fiddaman, Juliana Vianna, Daly Noll, Gemma Clucas, et al.. Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. *Molecular Biology and Evolution*, 2020, 37 (6), pp.1708-1726. 10.1093/molbev/msaa040 . hal-02988644

HAL Id: hal-02988644

<https://hal.science/hal-02988644>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article (Discoveries)

Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird

Authors: Hila Levy*¹, Steven R. Fiddaman*¹, Juliana A. Vianna², Daly Noll^{2,3}, Gemma V. Clucas^{4,5}, Jasmine K. H. Sidhu¹, Michael J. Polito⁶, Charles A. Bost⁷, Richard A. Phillips⁸, Sarah Crofts⁹, Gary D. Miller¹⁰, Pierre Pistorius¹¹, Francesco Bonnadonna¹², Céline Le Bohec^{13,14}, Andrés A. Barbosa¹⁵, Phil Trathan⁸, Andrea Raya Rey^{16,17,18}, Laurent A.F. Frantz¹⁹, Tom Hart¹, Adrian L. Smith¹

* These authors have provided equal contribution and share first authorship.

[1] Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom; [2] Pontificia Universidad Católica de Chile, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackenna 4860, Macul, Santiago, Chile; [3] Instituto de Ecología y Biodiversidad, Universidad de Chile, Departamento de Ciencias Ecológicas, Santiago, Chile; [4] Cornell Atkinson Center for a Sustainable Future, Cornell University, Ithaca, New York 14850, USA; [5] Cornell Lab of Ornithology, Cornell University, Ithaca, New York 14850, USA; [6] Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA; [7] Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 du CNRS-Université de La Rochelle, Villiers-en-Bois, 79630, France ; [8] British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, United Kingdom; [9] Falklands Conservation, PO Box 26, Stanley, Falkland Islands, FIQQ 1ZZ, United Kingdom; [10] Microbiology and Immunology, PALM, University of Western Australia, Crawley, WA, 6009, Australia; [11] DST/NRF Centre of Excellence at the Percy FitzPatrick Institute for African Ornithology, Department of Zoology, Nelson Mandela University, Port Elizabeth, 6031, South Africa; [12] CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, 1919 route de Mende, 34293 Montpellier cedex 5, France; [13] Université de Strasbourg, CNRS, IPHC UMR 7178, 23 rue Becquerel, F-67000 Strasbourg, France; [14] Centre Scientifique de Monaco, Département de Biologie Polaire, 8 quai Antoine 1er, MC 98000 Monaco, Principality of Monaco ; [15] Museo Nacional de Ciencias Naturales, Departamento de Ecología Evolutiva, CSIC, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain; [16] Centro Austral de Investigaciones Científicas – Consejo Nacional de Investigaciones Científicas y Técnicas (CADIC-CONICET), Bernardo Houssay 200, Ushuaia, Tierra del Fuego, Argentina; [17] Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Yrigoyen 879, Ushuaia, Argentina; [18] Wildlife Conservation Society, Amenábar 1595, Office 19, C1426AKC CABA, Buenos Aires, Argentina [19] School of Biological and Chemical Sciences, Fogg Building, Mile End Rd, Bethnal Green, London E1 4DQ, United Kingdom

1
2
3 35 All sequence data generated in this study were deposited in GenBank under accession numbers
4 36 MN394222 - MN394376 (*TLR4*), MN313018 - MN313169 (*TLR5*), MN312870 - MN313017 (*TLR7*),
5 37 and MN566362 - MN566421 (mitochondrial control region, HVR1).
6
7

8 38 Abstract

9
10
11 39 Over evolutionary time, pathogen challenge shapes the immune phenotype of the host to better
12 40 respond to an incipient threat. The extent and direction of this selection pressure depends on the local
13 41 pathogen composition, which is in turn determined by biotic and abiotic features of the environment.
14 42 However, little is known about adaptation to local pathogen threats in wild animals. The Gentoo
15 43 penguin (*Pygoscelis papua*) is a species complex that lends itself to the study of immune adaptation
16 44 because of its circumpolar distribution over a large latitudinal range, with little or no admixture
17 45 between different clades. In this study, we examine the diversity in a key family of innate immune
18 46 genes - the Toll-like receptors (*TLRs*) - across the range of the Gentoo. The three *TLRs* that we
19 47 investigated present varying levels of diversity, with *TLR4* and *TLR5* greatly exceeding the diversity
20 48 of *TLR7*. We present evidence of positive selection in *TLR4* and *TLR5*, which points to pathogen-
21 49 driven adaptation to the local pathogen milieu. Finally, we demonstrate that two positively selected
22 50 co-segregating sites in *TLR5* are sufficient to alter the responsiveness of the receptor to its bacterial
23 51 ligand, flagellin. Taken together, these results suggest that Gentoo penguins have experienced distinct
24 52 pathogen-driven selection pressures in different environments, which may be important given the role
25 53 of the Gentoo as a sentinel species in some of the world's most rapidly changing environments.
26
27
28
29
30
31
32
33
34

35 54 Introduction

36 55 All organisms are challenged by pathogens in their surrounding environments, but it is clear that the
37 56 pathogen pressure can vary by location. Similarly to free-living metazoans, a latitudinal species
38 57 richness gradient has been identified in several parasitic and pathogenic taxa, which may be driven by
39 58 temperature and other abiotic and biotic factors (Rohde and Heap 1998; Guernier, et al. 2004; Dionne,
40 59 et al. 2007). Given this gradient in pathogen pressure, it follows that natural selection on the host will
41 60 favour distinct immune phenotypes in different environments, as suggested by MHC II genetic
42 61 diversity patterns in Humboldt penguins associated with higher pathogen diversity in lower latitudes
43 62 (Sallaberry-Pincheira, et al. 2015; Sallaberry-Pincheira, et al. 2016). In our study, we sought to test
44 63 the hypothesis that pathogen-driven selection can drive distinct patterns of host immune system
45 64 genotype and phenotype, using the Gentoo penguin (*Pygoscelis papua* ssp.) as a model species
46 65 complex.
47
48
49
50
51
52
53
54
55

56 66 The Gentoo penguin complex (Vianna, et al. 2017; Clucas, et al. 2018) is ideally suited for
57 67 investigating pathogen-driven selection on the immune system. Firstly, it has a circumpolar range,
58 68 spanning the largest latitudinal range of any penguin species, between 46-66° S, with breeding

1
2
3 69 colonies in most of the Southern Ocean's sub-Antarctic islands, as well as the islands off Tierra del
4 70 Fuego in South America, South Georgia, the Scotia Arc, and the Western Antarctic Peninsula
5
6 71 (Stonehouse 1970). Secondly, population monitoring of the species has shown it to be growing at the
7
8 72 southern end of its range (Lynch, et al. 2012), with highly fluctuating changes over time in colonies in
9
10 73 the South Atlantic and Indian Oceans (Lescroel and Bost 2006; Trathan, et al. 2007). Thirdly, the
11
12 74 Gentoo penguin is a highly philopatric seabird known to remain close to its breeding colonies year
13
14 75 round (Trivelpiece, et al. 1987; Wilson, et al. 1998; Clausen and Putz 2003; Hinke, et al. 2017),
15
16 76 limiting gene flow across breeding regions (Levy, et al. 2016; Vianna, et al. 2017; Clucas, et al.
17
18 77 2018).

18 78 Furthermore, across its range, Gentoo penguins overlap with (and occasionally co-occur in mixed
19
20 79 colonies with) King (*Aptenodytes patagonicus*), Magellanic (*Spheniscus magellanicus*), Macaroni
21
22 80 (*Eudyptes chrysolophus*), and Southern Rockhopper penguins (*Eudyptes chrysocome*) in sub-
23
24 81 Antarctic colonies, as well as congeneric Adélie (*P. adeliae*) and Chinstrap penguins (*P. antarcticus*)
25
26 82 in its Antarctic range. Gentoo penguin colonies are also frequented by a number of flying birds with
27
28 83 vast ranges, including albatrosses and petrels (Order: Procellariiformes), as well as predator-
29
30 84 scavengers like skuas (genus *Stercorarius*) and sheathbills (genus *Chionis*) that could introduce and/or
31
32 85 spread novel avian pathogens. Levels of human interaction also vary across the range, from
33
34 86 permanent settlements with livestock present near colonies in the Falkland/Malvinas Islands, to
35
36 87 seasonal or year-round scientific research stations in Sub-Antarctic and Antarctic colonies, and an
37
38 88 increasing presence of Antarctic tourism. Differences in sympatric interactions with other species
39
40 89 across the range of the Gentoo is likely to result in different pathogen challenges and therefore
41
42 90 different selective pressures.

39 91 To investigate genetic diversity across the immune system, many immunogenetic studies on penguins
40
41 92 have focused on the major histocompatibility complex (MHC; Tsuda, et al. 2001; Bollmer, et al.
42
43 93 2007; Knafler, et al. 2012; Sallaberry-Pincheira, et al. 2016). Increasingly, however, the Toll-like
44
45 94 receptors (TLRs) are recognised as important monogenic determinants of disease resistance
46
47 95 phenotypes, and are therefore important operands for natural selection (Grueber, et al. 2014). Toll-like
48
49 96 receptors are the best-studied family of pattern-recognition receptors in the vertebrate innate immune
50
51 97 system, representing the front line of detection of pathogen challenge (Kawai and Akira 2006). TLRs
52
53 98 respond to highly conserved microbe-associated molecular patterns (MAMPs) that are structurally
54
55 99 conserved in large groups of pathogens. Upon binding of a MAMP, TLRs undergo dimerization and
56
57 100 initiate an intracellular signalling cascade that culminates in the production of anti-pathogen effector
58
59 101 molecules (Akira, et al. 2001; Botos, et al. 2011).

57 102 Vertebrates have six major families of TLRs which are typically conserved across evolutionary time
58
59 103 to retain specificity for a particular MAMP or family of MAMPs. In most avian species, there are ten
60

1
2
3 104 recognized Toll-like receptors (Roach, et al. 2005; Boyd, et al. 2007; Brownlie and Allan 2011). Of
4 105 these, TLR4 and TLR5 respond to the bacterial agonists lipopolysaccharide and flagellin,
5 106 respectively, while TLR7 responds to single-stranded RNA of viruses in the endosomal compartment
6 107 (Chow, et al. 1999; Gewirtz, et al. 2001; Lund, et al. 2004).

7
8
9
10 108 To investigate TLR diversity across the range of the Gentoo penguin complex, we sequenced the full
11 109 coding sequences of *TLR4*, *TLR5* and *TLR7*, as opposed to targeted portions of certain exons as in
12 110 previous studies (Dalton, et al. 2016a). These three genes represent bacterial- and viral-sensing Toll-
13 111 like receptors that are present in almost all vertebrates. Samples (n = 155) were obtained from a broad
14 112 geographic range across the range of the species (**Figure 1**), representing the largest geospatial scale
15 113 of any immunogenetics study outside of humans. We describe patterns of diversity in TLRs that have
16 114 a clear spatial component, and provide evidence that some of the diversity in TLR4 and TLR5 is
17 115 driven by positive selection between different locations. We also demonstrate that two of the
18 116 positively selected residues in TLR5 yield a phenotypic difference in the response of the receptor to
19 117 flagellin, providing further evidence that Gentoo penguins have experienced differential pathogen-
20 118 driven selection pressures in different environments.

119 Results

120 Amplification of TLR genes in the Gentoo penguin

121 Through successful amplification by PCR or whole-genome sequencing, we were able to confirm that
122 Gentoo penguins have clear homologs of *TLR4*, *TLR5*, and *TLR7*, finding no evidence of gene loss or
123 pseudogenization, as has been reported in other avian lineages for *TLR5* (Velová, et al. 2018) or in
124 African penguins for *TLR7* (Dalton, et al. 2016a).

125 The length of the *P. papua* *TLR4* coding sequence matches the longest reported length (2550 bp/849
126 aa) in other bird species. For *TLR5*, there is a start codon that yields an open reading frame in line
127 with the length of previously published *TLR5* sequences (2589 bp/862 aa; Velová, et al 2018), but the
128 ORF continues upstream of the putative start codon, yielding a complete ORF that is 2643 bp/880 aa,
129 which is 54 bp longer than other reported avian *TLR5* sequences. Both the longer and shorter ORF
130 respond to flagellin in our *in vitro* system (data not shown), suggesting both could be functional *in*
131 *vivo*. The length of the *TLR7* coding exon, at 3126 bp/1042 aa falls within the reported range of avian
132 coding sequence lengths.

133 Diversity Indices and Population Differentiation

134 Mitochondrial Hypervariable Region (HVR1)

135 All colonies with more than two sampled individuals presented high levels of mitochondrial
136 hypervariable region 1 haplotype diversity ($H_d = 0.60-1.00$; **Figure 2** and **Supplementary Table S1**)

1
2
3 137 as obtained in DnaSP 6.12.10 (Rozas, et al. 2017). Differentiation between colonies at this locus was
4 138 significant and in line with previous data for this species (see **Supplementary Table S2D**), with four
5 139 clearly differentiated clades obtained through population-level analyses in Arlequin v3.5.1.3
6 140 (Excoffier and Lischer 2010): (1) a southern clade consisting of colonies South of the Polar Front in
7 141 South Georgia, the South Orkneys, the South Shetlands, and Western Antarctic Peninsula; (2) a South
8 142 American/Falklands/Malvinas clade, (3) a Kerguelen clade and (4) a North Indian Ocean clade
9 143 (Marion and Crozet Islands).

14 144 Our BEAST 2 phylogenetic analysis of HVR1 showed support for a division within the Gentoo
15 145 penguin complex occurring approximately 3.36 Mya (1.72-4.88 Mya), when the North Indian Ocean
16 146 clade diverges from all others. The Kerguelen lineage diverges from the Atlantic lineages 2.36 Mya
17 147 (1.14-3.51 Mya), and the populations North and South of the Polar Front within the Atlantic Ocean
18 148 appear to diverge 1.19 Mya (0.51-1.75 Mya; **Figure 3**). Within each clade, no clear site-specific
19 149 mtDNA structure was noted.

25 150 TLR4

26 151 For *TLR4*, 13 polymorphic sites were identified, with a total of 21 distinct phased haplotypes coding
27 152 for 9 unique protein variants (**Supplementary Table S1**). Indian Ocean colonies (CR, MAR, COU,
28 153 and MO) presented the highest levels of haplotype diversity ($H_d = 0.66-0.90$), while the South
29 154 American colony in Tierra del Fuego (MT) and the Falkland/Malvinas Islands (CB and BR) had very
30 155 low diversity ($H_d = 0.00-0.05$; **Figure 2**). Among Gentoo penguin colonies south of the Polar Front,
31 156 diversity was highest in the central colonies of SIG, COP, SP, and BO ($H_d = 0.51-0.63$), decreasing
32 157 sharply in BI in South Georgia ($H_d = 0.00$), as well as southward down the Western Antarctic
33 158 Peninsula (GGV and JP; $H_d = 0.08-0.26$).

34 159 Comparing pairwise genetic distances (F_{st} and Φ_{st}) for in Arlequin *TLR4* (**Figure 5** and
35 160 **Supplementary Table S2A**), the Indian Ocean populations of Crozet and Marion differed
36 161 significantly ($F_{st} > 0.3$, $p < 0.01$) or near-significantly (corrected $p \sim 0.01$) from all other colonies in
37 162 the Atlantic. Crozet and Marion did not differ significantly from each other ($F_{st} = 0.001$, $p = 0.40$),
38 163 while some haplotypes were shared between CR/MAR and the Kerguelen Island colonies of COU and
39 164 MO, yielding some non-significant pairings among the four Indian Ocean colonies. In the Atlantic,
40 165 one haplotype, also present in the Indian Ocean in lower frequencies, dominated across all the
41 166 colonies (**Figure 4**). The central sites in the Scotia Arc (COP/SP on King George Island, Signy Island,
42 167 and BO on the northern end of the Western Antarctic Peninsula) exhibited greater diversity than other
43 168 Atlantic sites and contained private alleles. The overall pattern of significance for F_{st} and Φ_{st}
44 169 comparisons can be seen in **Figure 5**. Not surprisingly, colonies within the same island group that are
45 170 in close proximity to each other (COU/MO, CB/BR, COP/SP) did not differ significantly, despite
46 171 variations in sample sizes.

1
2
3 172 Hierarchical population structure was detected for *TLR4* across Gentoo penguin colonies using
4 Arlequin (AMOVA, global $F_{st} = 0.32$, $p < 0.0001$). The proportion of variation resulting from
5 173 differences among groups was 24.01% ($F_{cr} = 0.24$, $p = 0.003$) when colonies were placed into four
6 174 groups, coinciding with the four mtDNA clades: (1) Marion/Crozet archipelagos; (2) Kerguelen Is.;
7 175 (3) Falkland/Malvinas and Tierra del Fuego; and (4) south of the Polar Front in the Scotia Arc and
8 176 Antarctic Peninsula. However, F_{cr} increased to 28.19% ($p = 0.007$) when Bird Island was grouped
9 177 separately from other Southern Gentoo penguin colonies, a pattern also suggested in genomic-level
10 178 analyses (Clucas, et al. 2018).
11 179

16 180 *TLR5*

17 181 *TLR5* was the most diverse TLR locus analysed. Twenty polymorphic sites were identified, with a
18 182 total of 46 distinct phased haplotypes coding for 32 unique protein variants (**Supplementary Table**
19 183 **S1**). Five Gentoo penguin colonies north of the Polar Front (COU/MO in Kerguelen, CB/BR in
20 184 Falklands/Malvinas and MT in Tierra del Fuego) exhibited the highest diversity measures (6-17
21 185 haplotypes, $H_d = 0.77$ -0.92, **Figure 2**), despite differences in sample size. This is unexpected given
22 186 that Martillo Island is the smallest known population of this species ($N_c = 12$ breeding pairs, Ghys, et
23 187 al. 2008), yet still maintained high diversity (5 unique protein variants in a sample of $n = 5$) at this
24 188 locus. Interestingly, Crozet and Marion Island colonies exhibited substantially lower genetic variation
25 189 at this locus with only two haplotypes ($H_d = 0.13$ -0.44), though these were shared with the Kerguelen
26 190 colonies ($H_d = 0.86$ -0.87; **Figure 4**). Southern colonies exhibited moderate diversity ($H_d = 0.51$ -0.65),
27 191 with the exception of SP in the South Shetland Islands and the Western Antarctic Peninsula at the
28 192 edge of the range, with only two or three haplotypes ($H_d = 0.12$ -0.38). Strikingly, Atlantic colonies
29 193 south of the Polar Front were dominated by one haplotype found less frequently (10-28%) in the
30 194 northern Atlantic colonies and completely absent from all Indian Ocean colonies. The second-most
31 195 prevalent haplotype in southern colonies was completely absent from all northern colonies.

32 196 Pairwise F_{st} and Φ_{st} values obtained in Arlequin revealed significant clustering by clade in terms of
33 197 genetic distance at the *TLR5* locus (**Figure 5** and **Supplementary Table S2B**). Within three of the
34 198 four clades mentioned above, there were no significant differences (CR/MAR, COU/MO, and
35 199 CB/BR/MT). The only significantly differentiated within-clade pairs lay in Gentoo penguin colonies
36 200 south of the Polar Front, where Jougla Point differed significantly from COP (highest diversity among
37 201 Southern Gentoo penguins) and BI (at the opposite geographic edge of this clade's range). All other
38 202 colonies differed significantly from all colonies outside their clade (F_{st} range 0.1-0.88, $p < 0.01$), with
39 203 CR/COU being near-significant ($F_{st} = 0.13$, $p = 0.02$). This was reflected in AMOVA results, where a
40 204 four-clade grouping presented a proportion of variation resulting from differences among groups of
41 205 32.70% ($F_{cr} = 0.327$, $p < 0.0001$), and isolating Jougla Point increased the F_{cr} to 56.56% ($p <$
42 206 0.0001).

207 TLR7

208 *TLR7* was the least diverse TLR locus, with 9 polymorphic sites, 10 phased haplotypes, and 8 unique
209 protein variants present across the study colonies (**Figure 2** and **Supplementary Table S1**). One
210 haplotype was predominant in all colonies (frequency of 70-100%) and 7 of the 10 haplotypes were
211 private alleles, only found in single colonies (**Figure 4**). In the Indian Ocean, Kerguelen colonies
212 (COU/MO) had relatively more diversity ($H_d = 0.34-0.50$) than the Crozet and Marion Island colonies
213 ($H_d = 0.00$), which had only one haplotype. In the Atlantic, South Georgia (BI) had only one
214 haplotype ($H_d = 0.00$), while all northern Atlantic colonies (CB/BR/MT) and SP in the South Shetland
215 Islands presented two haplotypes ($H_d = 0.06-0.47$). Other southern colonies contained 3-4 haplotypes
216 ($H_d = 0.38-0.60$), while the southernmost colony at Jougla Point (JP) on the Western Antarctic
217 Peninsula exhibited the most unique haplotypes ($H = 5$; $H_d = 0.38$) and contained two private alleles.
218 Unsurprisingly, only 1 of 91 pairwise comparisons between colonies were significant in terms of F_{st}
219 and 3/91 for Φ_{st} ($p < 0.01$), with no pattern to this differentiation (**Figure 5** and **Supplementary**
220 **Table S2C**). Different AMOVA groupings yielded an among-group variation (F_{cr}) no higher than
221 2.85% (four-clade grouping), further highlighting the lack of structure in this locus.

222 Effect of Population Size on Diversity

223 Census population size was not significantly correlated to TLR haplotype diversity (See
224 **Supplementary Figure S1** and **Supplementary Tables S3 and S4**: *TLR4* $p = 0.067$; *TLR5* $p = 0.75$;
225 *TLR7* $p = 0.64$).

226 Isolation by Distance

227 Significant isolation by distance, using shortest distances by sea between colonies in a Mantel's test,
228 was detected in both *TLR4* ($r = 0.515$, $p = 0.001$) and *TLR5* ($r = 0.593$, $p = 0.001$; **Supplementary**
229 **Table S5A**). mtDNA HVR1 was less strongly correlated to isolation by distance ($r = 0.312$), though
230 marginally significant ($p = 0.015$). On the other hand, the lack of diversity and structure in the *TLR7*
231 locus yielded no significant correlation across the range of Gentoo penguin colonies sampled.

232 Analysis of Positive Selection

233 We investigated the *P. papua* *TLR4*, *TLR5* and *TLR7* genes for evidence of positive selection, which
234 could be an indicator of adaptation to local pathogen environments across the natural range of the
235 species. Neutrality tests (Tajima's D and Fu's F_s) did not yield observable patterns of significant
236 deviation from neutrality across the full length of these genes (**Supplementary Table S1**). Using a
237 codon-specific approach, the site models in the *codeml* package of programs in PAML v4.9 (Yang
238 1997; Yang 2007) were employed to test for signatures of positive selection in *P. papua* TLR loci.

239 Non-synonymous sites observed and analysed are graphically depicted in their relative positions on
240 the proteins in **Figure 6**.

241 As expected, the majority of codons (*TLR4*, 98.8%; *TLR5*, 98.4%) were predicted to be under
242 purifying selection with the ratio of non-synonymous to synonymous substitutions (dN/dS) being < 1 .
243 Interestingly, 1.2% (*TLR4*) or 1.6% (*TLR5*) of codons in the alignment were found to be positively
244 selected using M2a, and similar frequencies, 1.2% (*TLR4*) or 1.7% (*TLR5*), were found using M8. For
245 *TLR7*, 99.9% of sites were found to be under purifying selection, while the remaining 0.1% were
246 predicted to be under neutral selection. We investigated whether models that permit positive selection
247 were a significantly better fit to the multiple alignments than models where $dN/dS \neq 1$ by performing
248 likelihood ratio tests between pairs of models. For *TLR4* and *TLR5*, all model comparisons (M1a vs.
249 M2a, M7 vs. M8, and M8a vs. M8) significantly favoured the positive selection model compared to
250 the neutral model (*TLR4*, $p \leq 0.017$; *TLR5*, $p \leq 7.2 \times 10^{-23}$ for all comparisons), indicating that *P. papua*
251 *TLR4* and *TLR5* have likely undergone positive selection. In contrast, for *TLR7*, the data were not a
252 significantly better fit to the positive selection model compared to the neutral model ($p = 1$) so the null
253 hypothesis of codons being negatively- and neutrally-selected was not rejected.

254 For *TLR4* and *TLR5*, we then used the Bayes Empirical Bayes (BEB) algorithm to infer the posterior
255 probability that a particular codon has experienced positive selection. For *TLR4*, three codons were
256 predicted to have undergone positive selection at posterior probability of >0.90 under model M2a
257 (**Supplementary Tables S6A-C**). All but one of the *TLR4* polymorphic residues (12, 82, 236, 316
258 and 445) were located in the extracellular (LRR) domain, of which two were positively selected (12
259 and 236). The final positively selected site (659) was located in the transmembrane domain (**Figure**
260 **6**). Of the three selected sites, one (12V/A) is a relatively non-conservative change (see
261 **Supplementary Table S7** for amino acid distance metrics). One site in *TLR4* (12) has previously
262 been found to be under positive selection in birds (Velová, et al. 2018), while the remaining two (236
263 and 659) are novel selected sites in penguins (**Supplementary Table S8**).

264 For *TLR5*, there were 13 amino acid variants, of which nine were predicted to have undergone
265 positive selection at posterior probability of >0.90 in both M2a and M8 (**Supplementary Table S6D-**
266 **F**). Of these, four were located in the extracellular domain (10, 285, 442 and 535), one was in the
267 transmembrane domain (667), and four were located in the TIR (intracellular) domain (698, 747, 788
268 and 845; **Figure 6**). Two sites (442 and 698) have previously been reported as being positively
269 selected in other birds (Grueber, et al. 2014; Velová, et al. 2018), while the remaining seven sites are
270 novel selected sites in penguins. Interestingly, one site in the *TLR5* extracellular domain (285) is
271 adjacent to two residues known to be important for flagellin binding in Interface B (Yoon, et al. 2012;
272 Song, et al. 2017), and so could be important for ligand preference (**Supplementary Table S8**;
273 **Supplementary Figure S2**). Four positively selected sites in *TLR5* were non-conservative changes

1
2
3 274 (10C/Y, 442A/V, 667S/F and 788S/C). It is interesting to note that the non-conservative TLR5 667F/S
4
5 275 polymorphism is located in the transmembrane domain, a region that is typically constrained by the
6
7 276 physiochemical requirement to embed in the cell membrane. Furthermore, the homology-based
8
9 277 methods of amino acid substitution consequence prediction SIFT (Ng and Henikoff, 2003) and
10
11 278 PolyPhen-2 (Adzhubei, et al. 2013) both predict the S667F change to be of high functional
12
13 279 consequence (SIFT score = 0.01; PolyPhen-2 score = 0.998; **Supplementary Table S9**).
14
15 280 Transmembrane integrity is predicted to remain intact, despite the non-conservative polymorphism,
16
17 281 and the Phobius tool (Käll, et al. 2004) predicts the transmembrane domain is unchanged by the
18
19 282 polymorphism. While the vast majority (57/64, 89% of available sequences) of avian TLR5 have a
20
21 283 serine in this position in the transmembrane domain, only one other bird - the Northern Fulmar
22
23 284 (*Fulmarus glacialis*) - has a phenylalanine in this position (**Supplementary Figure S3**), indicating
24
25 285 there likely to have been a functional consequence to the S667F change in the Gentoo penguin.

26 **Functional Analysis of Selected TLR5 Residues**

27
28 287 While *in silico* methods can be useful indicators of protein residues under selection, functional study
29
30 288 is the only means to isolate the selected phenotype and its relevance. In order to assess whether key
31
32 289 selected sites identified in the positive selection analyses have functional consequences, we developed
33
34 290 an *in vitro* assay using transient expression of TLRs in a reporter cell line.

35
36 291 Since extracellular domain polymorphisms in TLRs are likely to give rise to preferences in ligand
37
38 292 type (Faber, et al. 2018), we focused on TMD/TIR domain polymorphisms that are likely to give rise
39
40 293 to differential signalling in response to the same agonist. Given that five *TLR5* polymorphisms were
41
42 294 located in the TMD/TIR domain, we tested the two polymorphisms with the highest posterior
43
44 295 probability of selection from the PAML analysis – residues 667 (TMD) and 845 (TIR).

45
46 296 Polymorphisms at these positions segregate well with geographical location: birds from Crozet and
47
48 297 Marion were all homozygous for the 667S/845I haplotype, while 87.3% (n = 71) of birds from
49
50 298 colonies South of the Polar Front were homozygous for the derived haplotype, 667F/845V (**Figure 7**).
51
52 299 In the Kerguelen Islands, the ancestral 667S/845I haplotype predominated, but variants at both
53
54 300 positions were present at lower frequencies. South American/Falklands/Malvinas colonies were the
55
56 301 most diverse at the positions of interest, where 46.7% (n = 14) of birds were heterozygous at one or
57
58 302 both of the sites. Overall, however, polymorphisms at these loci tended to co-segregate: 72.4% (n =
59
60 303 110) of birds were homozygote for either the ancestral (667S/845I) or derived (667F/845V)
304 haplotypes.

305
306 305 Given the strong tendency for the alleles at these positions to co-segregate at the extremes of the range
307
308 306 of *P. papua*, and the fact that these were the TIR/TMD polymorphisms with the highest likelihood of
309
310 307 positive selection, the functional consequences of altering both residues together were investigated.
311
312 308 FLAG-tagged constructs of both of the TLR5 TIR/TMD variants with the same LRR domain were

309 transiently expressed in *TLR5*^{-/-} HEK-Blue™ Null1 NF-κB reporter cells and protein expression levels
310 were normalised using an anti-FLAG ELISA. Cells expressing either constructs were then treated
311 with *Salmonella enterica* serovar Typhimurium-derived flagellin, or PBS control, and the NF- κB
312 response was measured. Cells expressing either construct responded to flagellin, demonstrating that
313 the *P. papua* TIR domain interacts efficiently with human adapter molecules. Interestingly, there was
314 a marked enhancement (~1.5 fold, $p = 0.04$) of the flagellin response in the variant that was
315 predominantly found in the Southern Gentoo compared to the variant found in the Indian Ocean
316 clades, suggesting that the derived haplotype 667F/845V has enhanced signalling capability compared
317 to the ancestral genotype 667S/845I (**Figure 7**). These data provide further evidence that *P. papua*
318 *TLR5* has undergone positive selection for different immune capabilities during the expansion of the
319 species below the Polar Front towards the Antarctic Peninsula.

320 Discussion

321 All vertebrates are subject to challenge by a plethora of pathogens that can exert strong selective
322 pressures on host populations. The innate immune system, and in particular the Toll-like receptors, is
323 responsible for both recognizing, and responding to, a pathogen threat by inducing inflammation and
324 priming the adaptive immune response. To investigate TLR adaptation in Gentoo penguins, we
325 sequenced the entire coding regions of *TLR4*, *TLR5* and *TLR7*, which recognize products from
326 bacterial and viral pathogens. Multiple individuals were sequenced from colonies at the extremes of
327 the species' range (~8000 km between the most distant colonies), providing an extensive geospatial
328 component to our analysis. We found spatially-associated patterns of diversity in the TLRs, although
329 greater diversity was observed in *TLR4* and *TLR5* compared to *TLR7*. Furthermore, clear evidence of
330 positive selection in both *TLR4* and *TLR5* was identified, which was further reinforced by the
331 demonstration that two of the *TLR5* TMD/TIR domain polymorphisms are sufficient to alter the
332 magnitude of responsiveness to flagellin. To our knowledge, no other TLR study outside of humans
333 has supported predictions of positive selection with confirmation of functional polymorphism.

334 Patterns of Diversification and Selection

335 Most studies of TLR genetic diversity in wild populations have investigated small, bottlenecked,
336 and/or endangered avian populations (Grueber, et al. 2012; Grueber, et al. 2013; Hartmann, et al.
337 2014; González-Quevedo, et al. 2015; Dalton, et al. 2016a; Dalton, et al. 2016b). In these populations,
338 drift, rather than selection, is suspected to have been the main driver of sampled diversity due to
339 recent bottlenecks or pronounced founder effects. Several studies have documented TLR diversity in
340 domesticated animals such as pigs (Darfour-Oduro, et al. 2016), cows (Novak, et al. 2019), and
341 chickens (Świdarská, et al. 2018), but large-scale studies on whole-gene TLR variation in a wild
342 population are lacking. Although study design can dramatically affect the diversity detected in
343 different studies it is noteworthy that with Gentoo penguins *TLR5* exhibited higher diversity than

1
2
3 344 *TLR4* whereas with domestic chicken breeds (n=110, 25 breeds; Świderská, et al. 2018) and grey
4 partridge (n=10; Vinkler, et al. 2015) *TLR4* was more diverse than *TLR5* (**Supplementary Table**
5 345 **S10**). The diversity of *TLR7* (compared with *TLR4* and/or *TLR5*) was relatively low in Gentoo
6 346 penguins, domestic chicken breeds and grey partridges (**Supplementary Table S10**).
7
8 347

9
10 348 To provide an internal reference for TLR diversity, we sequenced the mitochondrial hypervariable
11 349 region (HVR1) as a neutral marker in the same individuals. In line with previous studies, we found
12 350 evidence of at least four deeply divergent clades in *P. papua* based on HVR1 sequence (Vianna, et al.
13 351 2017; Clucas, et al. 2018; Pertierra, et al. in review). These more recent analyses support a revision,
14 352 first proposed by de Dinechin et al. (2012), of the previously-accepted two subspecies model that was
15 353 based on morphological characteristics (Stonehouse 1970). These four clades are likely to have much
16 354 greater divergence (millions of years) than what would be expected at the intraspecific level. We
17 355 found evidence of differentiation according to this underlying population structure in *TLR4* and *TLR5*
18 356 which further supports the argument for taxonomic revision of the species, with particular focus on
19 357 the classification of colonies in South Georgia and the Indian Ocean. Conversely, *TLR7* was highly
20 358 conserved across the species range and is clearly not subject to the same selection pressures as *TLR4*
21 359 and *TLR5*. Overall, our study highlights that the genetic differentiation across Gentoo penguin clades
22 360 is not just driven by drift, but by clear population-specific adaptations to the environment.
23
24
25
26
27
28
29
30

31 361 Diversity has been widely reported to vary between different families of TLRs, particularly
32 362 comparing extracellular and intracellular TLRs. Some authors have reported that TLRs that respond to
33 363 viral ligands are more likely to be under purifying selection, at least in mammals (Barreiro, et al.
34 364 2009; Wlasiuk and Nachman 2010; Wang, et al. 2016; Kloch, et al. 2018), although the pattern may
35 365 not be consistent in birds, with *TLR3* displaying the greatest number of non-synonymous variants of
36 366 the four TLRs tested in different chicken breeds (Świderská, et al. 2018), and *TLR7* diversity in the
37 367 house finch (*Carpodacus mexicanus*) far exceeding that of *TLR4* and *TLR5* (Alcaide and Edwards
38 368 2011). Consistent with the pattern observed in mammals (and also the lesser kestrel, *Falco naumanni*;
39 369 Alcaide and Edwards 2011), we observed overall nucleotide diversity measurements that were several
40 370 times higher for the extracellular/bacterial TLRs 4 and 5 (*TLR4*: $6.1 \times 10^{-4} \pm 0.5 \times 10^{-4}$; *TLR5*: $14.8 \times$
41 371 $10^{-4} \pm 0.6 \times 10^{-4}$) compared to the intracellular/viral *TLR7* ($1.0 \times 10^{-4} \pm 0.1 \times 10^{-4}$), indicating strong
42 372 purifying selection for maintenance of function in *TLR7*. In addition, we found no evidence of any
43 373 codons under selection in *TLR7*, compared to three and nine sites in *TLR4* and *TLR5*, respectively,
44 374 similar to the pattern of positively selected residues reported in several avian species (Alcaide and
45 375 Edwards 2011).
46
47
48
49
50
51
52
53
54
55

56 376 Within TLR sequences, levels of variation are not uniformly distributed across the domains of the
57 377 receptor. TLRs are type I integral membrane glycoproteins with highly conserved architecture across
58 378 large phylogenetic distances (Botos, et al. 2011). Typical TLR structure comprises an N-terminal
59
60

1
2
3 379 extracellular (or intraluminal for intracellular TLRs) leucine-rich repeat (LRR) domain for ligand
4 380 binding, a single transmembrane helix, and a C-terminal cytoplasmic signalling (Toll/interleukin-1
5 381 receptor, TIR) domain interacting with intracellular adapter proteins (Bell, et al. 2003). The leucine-
6 382 rich repeat domain directly binds microbe-derived ligands in all known vertebrate TLRs, with the
7 383 exception of TLR4 recognition of lipopolysaccharide via an accessory molecule, myeloid
8 384 differentiation factor, MD-2 (Park, et al. 2009). Since the LRR domain represents the interface
9 385 between host and pathogen, and pathogens exhibit variable MAMPs to evade detection (Andersen-
10 386 Nissen, et al. 2005), there is often an excess of diversity in the LRR domain compared to the TIR
11 387 domain (Świderská, et al. 2018; Velová, et al. 2018). In contrast, TIR domains interact with adapter
12 388 proteins such as MyD88 (myeloid differentiation primary-response protein 88) which are shared
13 389 between several TLRs, although a MyD88-independent pathway also facilitates TLR3 and TLR4
14 390 signalling (Akira and Takeda 2004). Unsurprisingly, TIR domains were found to be much more
15 391 highly conserved than their corresponding extracellular domains in a study of 366 vertebrate TLRs
16 392 from 96 species with the exception of *TLR10* (Mikami, et al. 2012). Within species, the same trend is
17 393 evident: in a study of *TLR3*, *4*, *5* and *7* diversity across domestic chicken breeds, only three of the 46
18 394 non-synonymous polymorphisms (two in *chTLR3* and one in *chTLR7*) were located in the TIR domain
19 395 (Świderská, et al. 2018).

20
21
22 396 In line with previous evidence for the asymmetric distribution of polymorphisms in TLR domains, we
23 397 identified an excess of polymorphisms in the LRR domain compared to the TIR domain in two of the
24 398 three TLRs studied (*TLR4*: 8 LRR vs. 0 TIR; *TLR7*: 6 LRR vs. 0 TIR). Interestingly, *P. papua TLR5*
25 399 contained a greater number of TIR domain polymorphisms than would be expected from other species
26 400 (11 LRR vs. 7 TIR), particularly given the LRR is over three-times the length of the TIR domain. Of
27 401 the seven *TLR5* TIR domain polymorphisms, five were non-synonymous substitutions, suggesting that
28 402 the TIR domain of *TLR5* has been under selection to modulate signalling intensity.

29
30
31 403 Somewhat surprisingly, we also identified non-synonymous polymorphic sites in the transmembrane
32 404 domains of both *TLR4* (659 Ala/Thr) and *TLR5* (667 Ser/Phe). The transmembrane domain is an
33 405 uncommon location for TLR polymorphisms, presumably because the region is highly constrained by
34 406 chemical and functional requirements. As such, the effects of polymorphisms in this region are often
35 407 large. For instance, the human *TLR1* 602S variant is associated with disrupted cell surface localization
36 408 of the receptor but is protective against pathology associated with leprosy. It is also noteworthy that
37 409 the Gentoo penguin *TLR5* transmembrane polymorphic site (667) identified in this study is highly
38 410 conserved elsewhere in avian phylogeny. Of the other birds with published *TLR5* sequences,
39 411 displayed in the alignment (**Supplementary Figure S3**), 57 (89%) have a serine (ancestral *P. papua*
40 412 genotype) at this position in the transmembrane domain, and only one other bird – the Northern
41 413 Fulmar (*Fulmarus glacialis*) – has a phenylalanine residue (derived *P. papua* genotype). The high
42 414 conservation of serine at this position in the protein points to a widespread pressure for maintenance

1
2
3 415 of function across avian phylogeny, and provides more evidence of a positively selected residue with
4
5 416 functional consequences.

6
7 417 **Functional polymorphisms in *TLR5* support positive selection**

8
9 418 We identified a number of positively selected codons in both *TLR4* and *TLR5*, making both of these
10
11 419 receptors candidates for further functional investigation. Polymorphisms in TLR LRR domains have
12
13 420 the potential to yield preferences for subtly different microbial ligands, such as LPS or flagellins from
14
15 421 different bacterial species (Nahori, et al. 2005; Faber, et al. 2018). However, very limited data are
16
17 422 available regarding which pathogens are present in the environments of each of the Gentoo clades,
18
19 423 and therefore elucidation of any differences in ligand preference will require further study. We did,
20
21 424 identify one positively selected site in TLR5 (285) that is adjacent to two important residues for
22
23 425 flagellin binding in Interface B (Yoon, et al. 2012; Song, et al. 2017). This site would be a good
24
25 426 candidate for functional investigation of changes in flagellin ligand preferences, but this would be
26
27 427 difficult in the absence of known flagellin variants in candidate *P. papua* pathogens. Instead, we
28
29 428 chose to investigate TIR and transmembrane domain polymorphisms for functional consequences
30
31 429 because these can yield signalling intensity differences in response to the same agonist (Faber, et al.
32
33 430 2018). Given that the TIR domain of TLR4 did not show any non-synonymous polymorphisms, we
34
35 431 focused on the TLR5 TIR/transmembrane domain, and in particular the two residues with the
36
37 432 strongest signature of positive selection (667 and 845). Site 667 was likely to be of significant
38
39 433 functional consequences because of its transmembrane location, non-conservative amino acid change
40
41 434 (serine to phenylalanine) and both SIFT and PolyPhen-2 predicting the change to be of high
42
43 435 importance.

44
45 436 The Gentoo penguin is reported to have undergone a circumpolar expansion, with ancestral
46
47 437 populations in the Indian Ocean seeding northern populations that expanded into the Atlantic, and
48
49 438 further expansions south of the Polar Front and to the West Antarctic Peninsula – the southernmost
50
51 439 extreme of the range (de Dinechin, et al. 2012; Peña, et al. 2014). It is interesting to note that one of
52
53 440 the ancestral Indian Ocean clades (Marion and Crozet archipelagos) is completely dominated by birds
54
55 441 of the 667S/845I genotype, while the most derived clade of Gentoo penguins south of the Polar Front
56
57 442 are almost entirely dominated by the 667F/845V genotype. These data may reflect an incipient
58
59 443 selective sweep of the 667F/845V genotype in Southern Gentoo penguins.

60
61 444 An alternative explanation for the reduction in diversity at residues 667 and 845 could be genetic
62
63 445 bottlenecks during the expansion of *P. papua* south of the Polar Front. However, evidence from
64
65 446 neutral markers in this study and a previous study (Levy, et al. 2016) reveal that neutral variation is
66
67 447 maintained in the Southern Gentoo penguin colonies at levels comparable with the northern Atlantic
68
69 448 Gentoo penguin clade up to the southernmost extreme of the range. In *TLR5*, we also saw no
70
71 449 correlation between census population size and haplotype diversity. These findings support the

1
2
3 450 hypothesis that a selective sweep, rather than a bottlenecking event, is responsible for the near-
4 451 fixation of the *TLR5* 667F/845V haplotype in the Southern Gentoo penguin. Perhaps more
5 452 importantly, the finding that the 667F/845V haplotype has enhanced signalling capability provides a
6 453 functional basis for selection of this *TLR5* haplotype.
7
8
9

10 454 **Potential drivers of selection**

11
12 455 Toll-like receptors, like other genes of the immune system, are subject to competing types of
13 456 selection. Balancing selection works to maintain diversity at a population level in response to the
14 457 diversity of pathogens in the environment, as was recently proposed in TLRs of the bank vole
15 458 (*Myodes glareolus*; Kloch, et al. 2018). In contrast, purifying selection may predominate (to retain
16 459 key functionality), which has been described in large-scale studies of human TLRs in different ethnic
17 460 backgrounds (Mukherjee, et al. 2014). Finally, positive selection may promote the fixation of novel
18 461 variants that confer a fitness advantage in the response to pathogens. In the present study, we found
19 462 evidence of positive selection in *TLR4* and *TLR5*, which likely indicates that the pathogen
20 463 composition differs substantially between distinct locations in the Gentoo penguin's range.
21
22
23
24
25
26

27 464 Spatial heterogeneity in the profile of pathogens that afflict Gentoo penguins would be a key driver
28 465 for the patterns of selection identified in the TLR variants. Latitudinal species diversity gradients have
29 466 been described for pathogens (and their hosts) (Rohde and Heap, 1998; Guégan, et al. 2008, Dionne,
30 467 et al. 2007), which might suggest fewer pathogens in Antarctic species. However, a diverse range of
31 468 pathogens are found in these environments (discussed below). Moreover, within the Gentoo range
32 469 there are diverse biotic and abiotic characters that exhibit spatial variation (Trathan, et al. 2007,
33 470 Barbosa, et al. 2009, Barbosa, et al. 2013, Lamont, et al. 2018; Chown, et al. 2015) and these factors
34 471 will affect the transmission of pathogens. Indeed, the regionalized selection of TLR alleles in different
35 472 sectors of the Gentoo range support the premise that different challenges are more prevalent or
36 473 pathogenic in different populations.
37
38
39
40
41
42
43

44 474 The dense colonial conditions and ubiquitous guano (faeces) that characterise Gentoo penguin
45 475 habitats provide ideal conditions for the transmission of a wide range of pathogens transmitted by
46 476 direct contact or faeces. Furthermore, penguins as a group are known to be highly susceptible to a
47 477 variety of infectious diseases, including, avian cholera (Jaeger, et al. 2018), avian pox (Kane, et al.
48 478 2012), avian malaria (Fix, et al. 1988; Grilo, et al. 2016) and aspergillosis (Flach, et al. 1990). A
49 479 number of infection associated mass mortality events have been documented in both wild and captive
50 480 penguin populations (Grimaldi, et al. 2015). However, little is known about the pathogens that exist in
51 481 sub-Antarctic and Antarctic regions, their prevalence, or their fitness costs on penguin populations.
52 482 Limited data are available on the prevalence of diseases in penguin populations (Clarke and Kerry
53 483 2000; Barbosa and Palacios 2009; Woods, et al. 2009; Grimaldi, et al. 2015), and most studies rely
54 484 upon short notes, observations, and case reports closely tied to obvious signs of disease and mass
55
56
57
58
59
60

1
2
3 485 mortality in well-studied and highly visited penguin colonies. Studies that survey the environmental
4 and host microbiomes to characterise pathogen presence in polar regions remain limited to sites near
5 486 major polar research stations, have small sample sizes, and/or do not cover large spatial and temporal
6 487 ranges (Zdanowski, et al. 2004; Dewar, et al. 2013; Fan, et al. 2013; Ma, et al. 2013; Dewar, et al.
7 488 2014).

8 489
9
10
11 490 The presence of Gram-negative bacteria exhibiting both lipopolysaccharides and flagella, including
12 491 *Campylobacter*, *Escherichia*, *Salmonella* and others, has been demonstrated in Gentoo penguin
13 492 colonies (Dimitrov, et al. 2009; Bonnedahl 2011; Barbosa, et al. 2013; González-Acuna, et al. 2013;
14 493 García-Peña, et al. 2017). However, it is not known whether any of these (or other bacterial
15 494 pathogens) vary across the Gentoo penguin's range, or may have played a role in the selection of
16 495 Gentoo penguin *TLR4* or *TLR5* variants.

17
18 496 Studies of single-stranded RNA viruses (which would typically be recognised by *TLR7*) are similarly
19 497 lacking in Gentoo penguins. Though single-stranded RNA viruses, including the causative agents of
20 498 Newcastle disease virus and avian influenza, have occasionally been detected in *Pygoscelis* penguins
21 499 through immunological assays and direct isolation (Morgan and Westbury 1988; Wallensten, et al.
22 500 2006; Neira, et al. 2017; Olivares, et al. 2019; Wille, et al. 2019), the fitness consequences of viral
23 501 infection on penguin populations are unknown. We know of only one case report from Signy Island
24 502 where evidence of a puffinosis outbreak (normally caused by Coronavirus) was described in Gentoo
25 503 penguin chicks (Mac Donald and Conroy 1971). The viral drivers behind the strong purifying
26 504 selection we observed in *TLR7* are unknown, but it could be that the ssRNA viruses that affect Gentoo
27 505 penguins are less diverse across the species range than flagellated Gram-negative bacteria.

28
29 506 Sympatric interactions with a diversity of migratory flying birds and their parasites may be important
30 507 contributors to pathogen diversity in Gentoo penguin colonies. Birds such as albatrosses, petrels,
31 508 shearwaters, sheathbills, shags, gulls, terns, and skuas are often observed in close proximity to Gentoo
32 509 penguin colonies and there are 46 species recognised in Antarctica alone (Lepage, et al. 2014). There
33 510 is some evidence that ectoparasites and blood parasites are transmitted between co-occurring bird
34 511 species (Barbosa et al. 2011; Levin, et al. 2013). It is plausible that cross-species transmission events
35 512 are important in cross-colony transmission and structuring the profile of pathogens afflicting
36 513 particular Gentoo penguins.

37
38
39 514 It remains unclear why the two functionally tested TMD/TIR residues in *TLR5* would confer
40 515 increased responsiveness to flagellin in Gentoo penguins south of the Polar Front. *TLR* signalling
41 516 must be tightly controlled and aberrant *TLR*-induced inflammation can lead to immune pathology,
42 517 toxic shock syndrome and death. It is unsurprising, therefore, that *TLR* polymorphisms have been
43 518 described that confer reduced sensitivity to their agonist and a state of tolerance. For instance, the
44 519 replacement of a highly conserved proline residue by a histidine at position 712 of *TLR4* confers

1
2
3 520 endotoxin resistance in certain strains of mice (Qureshi, et al. 1999). It could be that the exposure to
4
5 521 (or diversity of) pathogens is decreased for the Southern Gentoo penguin clade compared to other
6
7 522 clades of Gentoo penguins, and thus individuals can tolerate enhanced signalling to a prevailing
8
9 523 infection. Alternatively, the enhanced signalling could be a manifestation of the Southern Gentoo
10
11 524 penguin adapting to a particular pathogen that is present in the West Antarctic Peninsula and absent
12
13 525 elsewhere. The finding of adaptive changes in the Gentoo penguin immune system necessitates a
14
15 526 much better understanding of the pathogen threats faced by Gentoo penguins in order for their
16
17 527 significance to be realized.

18 528 **Concluding remarks**

19 529 This wide-ranging immunogenetic study of Toll-like receptors in wild Gentoo penguins reveals
20
21 530 differential selection and adaptation to local pathogen pressure. While the drivers behind the observed
22
23 531 patterns of diversity and selection remain unclear in the context of currently available data, it is clear
24
25 532 that the Gentoo penguin has undergone adaptation to local pathogen assemblages across its range.

26 533 Infectious disease threats to penguins are likely to become ever more severe in the coming decades
27
28 534 given the rapidly changing polar climate (Mayewski 2012; Lynch, et al. 2012). There is also evidence
29
30 535 of reverse zoonosis of enteric bacteria being transmitted from humans to sea bird species in Antarctica
31
32 536 (Cerdeña-Cuellar, et al. 2019), which could further increase transmission, especially in light of
33
34 537 increasing tourist and scientific research program presence in Antarctica. Although the Gentoo
35
36 538 penguin is not currently one of the 13 out of 18 penguin species with a conservation state of
37
38 539 threatened or near-threatened, certain sub-Antarctic populations have experienced sharp declines
39
40 540 (Crawford, et al. 2003; Lescroel and Bost 2006; Crawford, et al. 2009; Crawford, et al. 2014).
41
42 541 Consequently, the vulnerability of pathogen-naïve populations of penguins should not be
43
44 542 underestimated, nor should the importance of the Gentoo penguin as a sentinel species in the Southern
45
46 543 Ocean (Carpenter-Kling, et al. 2019).

47 544 Our findings have important implications for the conservation of not just Gentoo penguins, but also
48
49 545 many other vertebrate species, both in the wild and in captivity. Until now, most efforts to genetically
50
51 546 delineate conservation units have relied mostly on neutral markers. The ever-increasing availability of
52
53 547 genomic data allows targeted analysis of pathogen-recognition and other immune genes to assess
54
55 548 whether different populations possess specific functional adaptations to their environments and should
56
57 549 therefore be conserved separately. The approach used here, together with pathogen discovery and
58
59 550 surveillance systems, could better define conservation units in species that occupy varied habitats and
60
61 551 ecological niches in order to focus resources on potentially susceptible populations.

57 552 **Methods**

59 553 **Sample Collection**

1
2
3 554 This study used 155 blood samples from Gentoo penguins, previously obtained in the framework of
4 555 other projects. Samples were collected between 1999-2017 at the 14 sites shown in **Figure 1** (details
5 556 in **Supplementary Table S11**). To take blood, penguins were held with the flippers restrained and the
6 557 head placed under the arm of the handler, or they were wrapped in cushioned material covering the
7 558 head and preventing movement, to minimize stress during handling (Lemaho, et al. 1992). A second
8 559 handler took up to 1 mL blood from the brachial, intertarsal or jugular vein using a 25G or 23G needle
9 560 and 1 mL syringe or capillary, after cleaning the area with an alcohol swab. Total restraint time was
10 561 generally two to three minutes. The animal was then released at the edge of the colony and observed
11 562 to ensure it returned to its normal behavior. Blood was stored in 95% ethanol or Queen's Lysis Buffer
12 563 at -20 °C for transport at room temperature and subsequent storage at -20 °C upon arrival. All blood
13 564 samples were imported under the appropriate animal by-product import licenses.

14 565 Sampling was conducted under permits from each site's territorial government or governing agency.
15 566 These permits for animal handling were issued following independent institutional ethical review of
16 567 the sampling protocols, in accordance with Scientific Committee for Antarctic Research (SCAR)
17 568 guidelines.

28 569 **DNA extraction**

29 570 DNA for samples from MO, CB, BI, COP, and JP was extracted from blood samples using QIAGEN
30 571 DNeasy Blood and Tissue kits. The digestion step was modified to include 40 µL proteinase K and
31 572 extended to 3 hours for blood samples. Details of the modifications made to the protocols for tissue
32 573 samples are available in (Younger, Emmerson, et al. 2015; Younger, Clucas, et al. 2015). All these
33 574 samples were treated with 1 µL Riboshredder (Epicentre) to reduce RNA contamination and DNA
34 575 was visualized on a 1% agarose gel to confirm high molecular weight DNA was present. DNA
35 576 concentration and purity was measured on a Qubit and Nanodrop (ThermoFisher Scientific),
36 577 respectively. These samples are stored at the University of Oxford for future analysis. DNA from all
37 578 other sampling sites was isolated using a modified salt protocol (Aljanabi and Martinez 1997), with
38 579 details in Vianna, et al. (2017), stored at the Pontificia Universidad Católica de Chile for future
39 580 analysis.

48 581 **TLR Genotyping**

49 582 Primers for *TLR4*, 5, and 7 coding regions were designed using the primer design feature based on
50 583 Primer3 2.3.7 (Untergasser, et al. 2012) in Geneious v.11.0 (Biomatters, <http://www.geneious.com>).
51 584 Reference coding sequences for primer design were derived from the congeneric Adélie penguin
52 585 (*Pygoscelis adeliae*) reference genome (Genbank accession JMFP01000000) and unpublished Gentoo
53 586 penguin genomic data. *TLR4* amplifications for samples from MO, CB, BI, COP, and JP, were
54 587 conducted in 12 µL volumes (9 µL Qiagen *Taq* PCR Master Mix, 2 µL 10 µM forward and reverse
55 588 primer mix, and 1 µL of template DNA diluted 1:100). *TLR5* and *TLR7* amplifications for these

1
2
3 589 samples were conducted in 25 μ L volumes containing 5 μ L 5X Phusion High Fidelity (HF) Buffer
4 590 (New England Biolabs, UK), 0.5 μ L 10 mM dNTPs, 1.25 μ L of 10 μ M forward primer, 1.25 μ L of 10
5 591 μ M reverse primer, 2 μ L of template DNA diluted 1:100, 0.25 μ L of Phusion Hot Start Flex DNA
6 592 Polymerase (New England Biolabs, UK), and 14.75 μ L nuclease-free water. One GC-rich region in
7 593 *TLR7* required the use of Phusion GC buffer, rather than HF buffer for amplification. PCR products
8 594 were visualized on a 1% agarose gel stained with SYBR Safe. The primers and PCR reaction
9 595 conditions are fully detailed in the Supplementary Information (**Supplementary Table S12**). PCR
10 596 products were sequenced using Macrogen Europe's EZ-Seq (*TLR4*) or Eco-Seq (*TLR5* and *TLR7*)
11 597 services (<http://www.macrogen.com>, Netherlands) for purification and Sanger sequencing, using the
12 598 same PCR primers for sequencing.

13
14
15
16
17
18
19 599 For all remaining sampling sites, which underwent whole genome sequencing, a total of 100 ng of
20 600 genomic DNA was fragmented to an average of 350 base pairs to construct paired-end libraries using
21 601 the Illumina TruSeq Nano kit with the included indexed adapter and barcode. A total of six PCR
22 602 cycles were used for enrichment, purified with Sample Purification beads, quantified using a Qubit
23 603 fluorometer and then sequenced to ~20x coverage with 150 paired-end reads using an Illumina HiSeq
24 604 X platform at MedGenome (USA).

25
26
27
28
29 605 Sequences for each TLR coding region were assembled, edited, and aligned using Geneious v.11.0.
30 606 For *TLR4*, which has multiple exons along the coding region, exon sequences were extracted and
31 607 concatenated for further analysis. Heterozygous sites and single nucleotide polymorphisms (SNPs)
32 608 were detected by visually examining chromatograms. In cases of doubt, resequencing was
33 609 accomplished so that only high-quality reads from multiple sequencing runs were called as SNPs. All
34 610 heterozygous sites also had homozygous individuals within the dataset, and each gene had at least one
35 611 haplotype homozygous across the full length of the gene. All alleles were verified using a
36 612 combination of multiple independent Sanger sequencing runs and where available, the whole genome
37 613 sequencing data. The International Union of Pure and Applied Chemistry (IUPAC) code for
38 614 degenerate nucleotides was used to label heterozygous positions.

615 **mtDNA Genotyping**

39
40
41
42
43
44
45
46
47
48 616 For mitochondrial DNA, the hypervariable region of the mitochondrial control region (HVR1), also
49 617 known as Domain I, was amplified using the primers GPPAIR3F and GPPAIR3R (Clucas, et al.
50 618 2014) for samples from MO, CB, BI, COP, and JP. Amplifications were conducted in 25 μ L volumes
51 619 according to the manufacturer's instructions, using Phusion Hot Start Flex DNA Polymerase (New
52 620 England Biolabs, UK) and 2 μ L of template DNA diluted 1:100. Amplifications involved a two-step
53 621 PCR, with an initial cycle of 98 $^{\circ}$ C for 30 seconds, 40 cycles of 98 $^{\circ}$ C for 10 seconds and 72 $^{\circ}$ C for 20
54 622 seconds, followed by a 10-minute extension at 72 $^{\circ}$ C. PCR products were visualized on a 1% agarose
55 623 gel stained with SYBR Safe. PCR fragments were purified using an ethanol/sodium acetate

1
2
3 624 precipitation, and sequencing was performed using the Applied Biosystems BigDye Terminator v3.1
4 sequencing kit (Applied Biosystems) with the same PCR primers as sequencing primers.
5 625

6
7 626 Published mtDNA HVR1 sequences for the samples from BI (GenBank accessions KJ646314-
8 KJ646330, $n = 16$) and COP (KJ646361-KJ646382, $n = 21$) were included in the analysis for the
9 627 relevant individuals. mtDNA data was not available for the individuals from two Antarctic sites.
10 628 Though other individuals from those sites have sequence data available, we only included data from
11 629 individuals sequenced for both TLRs and HVR1.
12 630

13
14
15
16 631 Individual mtDNA fragments from all remaining sites (CR, MAR, COU, BR, MT, SIG, SP) were
17 632 amplified using the primers tRNAGlu and AH530 (Roeder, et al. 2002). These reactions included 0.4
18 633 μM of each primer, 1.5 mM 1X of PCR reaction buffer, MgCl_2 , 200 μM of each dNTP, and 1U of Taq
19 634 Polymerase Platinum (Invitrogen) in a two-phase touchdown program (Korbie and Mattick 2008): (1)
20 635 10 minutes at 95 °C, and 11 cycles of 95 °C for 15 seconds; a touchdown with an annealing
21 636 temperature of 60 °C–50 °C for 30 seconds, with one cycle per 1 °C interval, and 72 °C for 45
22 637 seconds; (2) 35 amplification cycles at 95 °C for 15 seconds, 50 °C for 30 seconds, and 72 °C for 45
23 638 seconds; and a final extension period of 30 minutes at 72 °C. The purification of these PCR products
24 639 and sequencing was carried out by Macrogen using an ABI PRISM 3730XL.
25 639
26
27
28
29
30
31

32 640 Only overlapping segments of the HVR1 sequences common to all samples were used by aligning and
33 641 editing with Geneious v11.0. Consensus sequences for the resulting 273 bp region of interest were
34 642 extracted for analysis.
35 642

36 37 38 643 **Population-Level Analyses**

39
40 644 Haplotypes were inferred for each of the diploid TLR loci using the PHASE algorithm (Stephens, et
41 645 al. 2001) implemented in DnaSP 6.12.10 (Rozas, et al. 2017) with 10,000 iterations and 1,000 burn-in
42 646 iterations. Phased haplotype data was used as input to determine standard genetic diversity measures
43 647 of each population, including number of polymorphic sites, haplotypes, haplotype diversity, and
44 648 nucleotide diversity, using DnaSP 6.12.10 and Arlequin v3.5.1.3 (Excoffier and Lischer 2010).
45 649 Minimum spanning haplotype networks were constructed and visualized using PopART 1.7 (Bandelt,
46 650 et al. 1999; Leigh and Bryant 2015) for each locus. DnaSP was also used to identify synonymous and
47 651 non-synonymous polymorphic sites and frequencies. Arlequin was used to calculate Tajima's D
48 652 (Tajima 1989), and Fu's F_s (Fu 1997). FSTAT v.2.9.3 (Goudet 1995) was used to calculate allelic
49 653 richness, taking into account differences in sample size.
50 653
51
52
53
54
55

56 654 Because TLR nucleotide diversity has been observed to have a correlation to population size in some
57 655 island bird populations (Gilroy, et al. 2017), we evaluated the relationship between Gentoo penguin
58 656 census population sizes (N_c) and haplotype diversity (H_d) at each locus. Gentoo penguins are
59
60

1
2
3 657 philopatric, but also show evidence of admixture within island groups and adjacent coastlines (Levy,
4 658 et al. 2016 and Vianna, et al. 2017). For this reason, the census population sizes selected for the
5 659 analysis were numbers of breeding pairs from the most recent available surveys of each archipelago or
6 660 region (in the case of the South Shetland Islands and Western Antarctic Peninsula). The N_c size and
7 661 source survey reference is available in **Supplementary Table S3**. Spearman's rank correlations and p
8 662 values were calculated for each diversity-size comparison.

9
10
11
12
13 663 For population differentiation comparisons, Arlequin was used to calculate pairwise F_{ST} distances
14 664 based on haplotype frequencies (Weir and Cockerham 1984), and pairwise Φ_{ST} s for the TLR and
15 665 mtDNA sequences (Excoffier, et al. 1992). FindModel (Posada and Crandall 1998) was used to find
16 666 the best fit substitution model for use in Arlequin. Φ_{ST} calculations for the TLR loci were obtained
17 667 using the Tamura and Nei substitution model (Tamura and Nei 1993), while mtDNA Φ_{ST} analysis was
18 668 carried out using the Kimura 2-Parameter model (Kimura 1980) with a gamma of 0.27. Analysis of
19 669 molecular variance (AMOVA) was used to compute hierarchical F-statistics, with 10,000
20 670 permutations, to evaluate likely patterns of genetic structure, seeking to identify the population
21 671 grouping that maximized the among-group variation (F_{CT}) and minimized the variation among
22 672 colonies within groups (F_{SC}) (Excoffier, et al. 1992). Significance of overall and pairwise genetic
23 673 distances were computed using 1,000,000 permutations. We used the SGoF+ method (Carvajal-
24 674 Rodriguez and de Una-Alvarez 2011) within the Myriads software (Carvajal-Rodriguez 2018) to
25 675 correct for multiple tests.

26
27
28
29
30
31
32
33
34 676 To test for isolation-by-distance, shortest distances by sea (during summer ice extent), in km, were
35 677 computed between each sampling location, using Google Earth v7.3.2.5776 (**Supplementary Table**
36 678 **S5B**), which were then related to pairwise F_{ST} in a Mantel test, implemented in Arlequin.

37 38 39 40 679 **Population Divergence and Phylogeography**

41
42 680 Phylogenetic reconstruction and estimates of divergence time were carried out using BEAST 2.5.2
43 681 (Bouckaert, et al. 2019). The evolutionary model for mtDNA analysis was selected using jModelTest
44 682 v. 2.1.10 (Darriba, et al. 2012), testing 88 candidate models and selecting the best fit using the Bayes
45 683 Information Criterion (BIC). All 88 models were within the 100% confidence interval, with HKY+G
46 684 selected for further divergence analyses (Hasegawa, et al. 1985). A total of 15 Adélie penguin (*P.*
47 685 *adeliae*) and 15 Chinstrap penguin (*P. antarcticus*) mitochondrial HVR1 GenBank sequences (Clucas,
48 686 et al. 2014), aligned and cropped to the equivalent size of the Gentoo sequences to avoid bias, were
49 687 included in the analysis as outgroups.

50
51
52
53
54
55 688 The most recent common ancestor prior was set for the *Pygoscelis* genus at 7.6 Mya (Subramanian, et
56 689 al. 2013), derived from the fossil calibration for *Pygoscelis grandis* (Walsh and Suarez 2006), with a
57 690 normal distribution, and standard deviation (σ) of 1.3 Mya. A strict molecular clock with a starting
58
59
60

1
2
3 691 prior of 1.0 and a Yule speciation process for branching rates, with uniform priors for birth and clock
4 692 rates of 1.0 was applied. Four independent runs of 30 million MCMC chains were performed, logging
5 693 parameters every 3,000 steps. The four independent runs were then combined using LogCombiner
6 694 v.2.5.2 and assessed for convergence within Tracer v.1.7.1 (Rambaut, et al. 2018). All parameters
7 695 converged with ESS values greater than 6,000. A maximum clade credibility tree was then generated
8 696 using TreeAnnotator v.2.5.2 (part of the BEAST software distribution) and visualized in FigTree
9 697 v.1.4.3.

15 698 **Selection Analyses**

16
17 699 Phylogenetic inference was carried out on phased sequence data for each TLR locus using RAxML-
18 700 NG using a GTR substitution matrix (Kozlov, et al. 2019). To detect selection, maximum likelihood
19 701 analysis of ratios of non-synonymous to synonymous nucleotide substitutions (dN/dS; ω) was
20 702 performed with the *codeml* package of programs in PAML v. 4.9 (Yang 1997; Yang 2007). Various
21 703 models were fitted to the multiple alignments: M1a (neutral model; two site classes: $0 < \omega_0 < 1$ and ω_1
22 704 = 1); M2a (positive selection; three site classes: $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 > 1$); M7 (neutral model;
23 705 values of ω fit to a beta distribution where $\omega > 1$ disallowed); M8 (positive selection; similar to M7
24 706 but with an additional codon class of $\omega > 1$); M8a (neutral model; similar to M8 but with a fixed
25 707 codon class at $\omega = 1$). Likelihood ratio tests were performed on pairs of models to assess whether
26 708 models allowing positively selected codons gave a significantly better fit to the data than neutral
27 709 models (model comparisons were M1a vs. M2a, M7 vs. M8, and M8a vs. M8). In situations where the
28 710 null hypothesis of neutral codon evolution could be rejected ($p < 0.05$), the posterior probability of
29 711 codons under selection in M2a and M8 were inferred using the Bayes Empirical Bayes algorithm
30 712 (Yang, et al. 2005).

39 713 **In Silico Prediction of Polymorphism Functional Consequences**

40 714 Physiochemical distances between amino acid variants were assessed using distance matrices
41 715 provided by several authors (Sneath, 1966; Epstein, 1967; Grantham, 1974; Miyata, 1979; Urbina, et
42 716 al. 2006; **Supplementary Table S7**). Predicted functional consequences of amino acid substitutions
43 717 were assessed using the homology-based tools SIFT (Ng and Henikoff, 2003) and PolyPhen-2
44 718 (Adzhubei, et al. 2013), using online servers (<https://sift.bii.a-star.edu.sg/> and
45 719 <http://genetics.bwh.harvard.edu/pph2/>; both accessed December 2019). Transmembrane domain
46 720 positions were predicted using Phobius (Käll, et al. 2004; <http://phobius.sbc.su.se/>; accessed
47 721 December 2019).

54 722 **Functional Analysis of TLR5 Genotype Expression in CRISPR-Cas9 edited HEK-Blue Cells**

55 723 In order to functionally assess positively selected *TLR5* polymorphisms *in vitro*, two full-length *TLR5*
56 724 sequences were synthesised including the two signalling domain polymorphisms that had the

1
2
3 725 strongest signature of selection (GBlocks, IDT). Synthetic genes were cloned using the Gibson
4 726 assembly method (Gibson, et al. 2009) into the p3XFLAG-CMVTM-14 expression vector (Sigma)
5 727 which incorporates a 3x-FLAG sequence on the C-terminus of the expressed construct. Insert-
6 728 containing vector was purified using the ZymoPURE II plasmid Maxiprep with the optional
7 729 endotoxin-removal step (Zymo). Both constructs were transiently expressed using TransIT[®]-2020
8 730 (Mirus Bio) in custom HEK-BlueTM Null1 cells (InvivoGen) that had undergone genome editing using
9 731 the CRISPR-Cas9 technique to disrupt endogenous human *TLR5*. Cells expressing Gentoo *TLR5*
10 732 constructs were challenged with *Salmonella* Typhimurium-derived flagellin (FLA-ST; InvivoGen) at
11 733 100 ng/mL and incubated for 24 h. Cell supernatants were harvested and NF- κ B activity was assessed
12 734 by measuring the absorbance at 405 nm on a FLUOstar[®] Omega microplate reader (BMG Labtech)
13 735 following the addition of p-nitrophenyl phosphate substrate, according to the manufacturer's
14 736 instructions (SIGMAFASTTM, Sigma). Expression levels were monitored by subjecting cell lysates to
15 737 a direct anti-FLAG ELISA. Cell lysates were harvested in ice-cold RIPA buffer (ThermoFisher) and
16 738 proteins were immobilised on high-bind ELISA plates (VWR) using coating buffer (BioLegend)
17 739 overnight at 4 °C. Wells were blocked using StartingBlockTM (PBS) blocking buffer (ThermoFisher)
18 740 for 1 h and then incubated with mouse monoclonal anti-FLAG M2 antibody (Sigma, 1:1000) at 37 °C
19 741 for 1 h, followed by incubation with goat anti-mouse IgG-HRP (ThermoFisher, 1:10000) for 1 h at 37
20 742 °C. Reactive protein amount was then assessed by the addition of 3,3',5,5'-tetramethylbenzidine
21 743 substrate and measurement of absorbance at 650 nm. Expression data were then used to normalise
22 744 signalling data. Statistical differences between means were determined by a two-tailed Student's t-
23 745 test, and statistical significance was considered to be $p < 0.05$. Transfections were carried out in three
24 746 independent wells per condition, and the experiment was conducted on at least three independent
25 747 occasions.

40 748 **Acknowledgments**

41
42
43 749 Financial support for this study was provided by an Oxford Clarendon Fund scholarship for HL and
44 750 from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number
45 751 BB/M011224/1] for SRF. Sample collection was funded in part by CONICYT PIA ACT172065 GAB
46 752 and Spanish Ministry of Science projects CGL2007-60369 and CTM2015-64720. Logistic and field
47 753 costs for sampling at the Crozet and Kerguelen archipelagos were supported by the Institut Polaire
48 754 Français Paul-Emile Victor (IPEV: Programme 137 – C.L.B. and Programme 354 – F.B.,
49 755 respectively) with additional support from the Laboratoire International Associé 647 'BioSensib'
50 756 (CSM/CNRS-University of Strasbourg). We appreciate the hospitality of all of our sampling logistics
51 757 providers, including W. Trivelpiece and the U.S. Antarctic Marine Living Resource Program, the
52 758 Argentinean Antarctic Station Carlini and the transportation by the Spanish Polar ship Las Palmas.
53 759 We also thank A. D. Rogers for providing laboratory resources and guiding the early phases of this
54 760 study.

1
2
3 761 **Author Contributions**
4

5 762 HL and SRF conducted all analyses and wrote the paper, guided by TH and ALS who designed the
6
7 763 study. HL, JAV, DN, GVC, MJP, CAB, RAP, SC, GDM, PP, FB, CLB, AAB, PT, ARR, and TH
8
9 764 performed fieldwork and contributed samples. HL, GVC, and DN extracted all DNA. HL, SRF, and
10
11 765 JKHS performed all PCR amplifications. JAV and DN prepared and provided whole genome
12
13 766 sequence data. SRF and JKHS performed all cell culture, cloning, and expression assays. JAV, DN
14
15 767 and LAFF provided analytical support with phylogenetic analyses. All authors discussed the study
16
17 768 results and implications and contributed to editing the manuscript.

17 769 **References**
18
19

- 20 770 Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations
21 771 using PolyPhen-2. *Current Protocols in Human Genetics* 76(1):7.20.1-7.20.41.
22 772 Akira S, Takeda K. 2004. Toll-like receptor signalling. *Nature Reviews Immunology* 4:499-511.
23 773 Akira S, Takeda K, Kaisho T. 2001. Toll-like receptors: critical proteins linking innate and acquired
24 774 immunity. *Nature Immunology* 2:675-680.
25 775 Alcaide M, Edwards SV. 2011. Molecular Evolution of the Toll-Like Receptor Multigene Family in
26 776 Birds. *Molecular Biology and Evolution* 28:1703-1715.
27 777 Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for
28 778 PCR-based techniques. *Nucleic Acids Research* 25:4692-4693.
29 779 Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, Aderem A. 2005.
30 780 Evasion of Toll-like receptor 5 by flagellated bacteria. *Proceedings of the National Academy of*
31 781 *Sciences of the United States of America* 102:9247-9252.
32 782 Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies.
33 783 *Molecular Biology and Evolution* 16:37-48.
34 784 Barbosa A, Benzal J, Vidal V, D'Amico V, Coria N, Diaz J, Motas M, Palacios MJ, Cuervo JJ, Ortiz
35 785 J, et al. 2011. Seabird ticks (*Ixodes uriae*) distribution along the Antarctic Peninsula. *Polar Biology*
36 786 34:1621-1624.
37 787 Barbosa A, De Mas E, Benzal J, Diaz JI, Motas M, Jerez S, Pertierra L, Benayas J, Justel A,
38 788 Lauzurica P, et al. 2013. Pollution and physiological variability in gentoo penguins at two rookeries
39 789 with different levels of human visitation. *Antarctic Science* 25:329-338.
40 790 Barbosa A, Palacios MJ. 2009. Health of Antarctic birds: a review of their parasites, pathogens and
41 791 diseases. *Polar Biology* 32:1095-1115.
42 792 Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O,
43 793 Gicquel B, et al. 2009. Evolutionary Dynamics of Human Toll-Like Receptors and Their Different
44 794 Contributions to Host Defense. *Plos Genetics* 5.
45 795 Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR, Segal DM. 2003. Leucine-rich repeats and
46 796 pathogen recognition in Toll-like receptors. *Trends in Immunology* 24:528-533.
47 797 Bollmer JL, Vargas FH, Parker PG. 2007. Low MHC variation in the endangered Galapagos penguin
48 798 (*Spheniscus mendiculus*). *Immunogenetics* 59:593-602.
49 799 Bonnedahl J. 2011. Antibiotic Resistance in Enterobacteriaceae Isolated from Wild Birds. In. Uppsala
50 800 University.
51 801 Botos I, Segal DM, Davies DR. 2011. The Structural Biology of Toll-like Receptors. *Structure*
52 802 19:447-459.
53 803 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J,
54 804 Jones G, Kuhnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for
55 805 Bayesian evolutionary analysis. *Plos Computational Biology* 15.
56 806 Boyd A, Philbin VJ, Smith AL. 2007. Conserved and distinct aspects of the avian Toll-like receptor
57 807 (TLR) system: implications for transmission and control of bird-borne zoonoses. *Biochemical Society*
58 808 *Transactions* 35:1504-1507.

- 1
2
3 809 Brownlie R, Allan B. 2011. Avian toll-like receptors. *Cell and Tissue Research* 343:121-130.
- 4 810 Carpenter-Kling T, Handley JM, Connan M, Crawford RJM, Makhado AB, Dyer BM, Froneman W,
5 811 Lamont T, Wolfaardt AC, Landman M, et al. 2019. Gentoo penguins as sentinels of climate change at
6 812 the sub-Antarctic Prince Edward Archipelago, Southern Ocean. *Ecological Indicators* 101:163-172.
- 7 813 Carvajal-Rodriguez A. 2018. Myriads: P-value-based multiple testing correction. *Bioinformatics*
8 814 34:1043-1045.
- 9 815 Carvajal-Rodriguez A, de Uña-Alvarez J. 2011. Assessing Significance in High-Throughput
10 816 Experiments by Sequential Goodness of Fit and q-Value Estimation. *PLoS One* 6:e24700.
- 11 817 Cerda-Cuellar M, More E, Ayats T, Aguilera M, Munoz-Gonzalez S, Antilles N, Ryan PG, Gonzalez-
12 818 Solis J. 2019. Do humans spread zoonotic enteric bacteria in Antarctic? *Science of the Total*
13 819 *Environment* 654:190-196.
- 14 820 Chow JC, Young DW, Golenbock DT, Christ WJ, Gusovsky F. 1999. Toll-like receptor-4 mediates
15 821 lipopolysaccharide-induced signal transduction. *Journal of Biological Chemistry* 274.
- 16 822 Chown S, Clarke A, Fraser C, Cary SC, Moon KL, McGeoch MA. 2015. The changing form of
17 823 Antarctic biodiversity. *Nature* 522(7557):431-438.
- 18 824 Clarke J, Kerry K. 2000. Diseases and parasites of penguins. *Penguin Conservation* 13:5-24.
- 19 825 Clausen A, Putz K. 2003. Winter diet and foraging range of gentoo penguins (*Pygoscelis papua*) from
20 826 Kidney Cove, Falkland Islands. *Polar Biology* 26:32-40.
- 21 827 Clucas GV, Dunn MJ, Dyke G, Emslie SD, Levy H, Naveen R, Polito MJ, Pybus OG, Rogers AD,
22 828 Hart T. 2014. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula
23 829 penguins. *Scientific Reports* 4.
- 24 830 Clucas GV, Younger JL, Kao D, Emmerson L, Southwell C, Wienecke B, Rogers AD, Bost C-A,
25 831 Miller GD, Polito MJ, et al. 2018. Comparative population genomics reveals key barriers to dispersal
26 832 in Southern Ocean penguins. *Molecular Ecology* 27:4680-4697.
- 27 833 Crawford RJM, Cooper J, Du Toit M, Greyling MD, Hanise B, Holness CL, Keith DG, Nel JL,
28 834 Petersen SL, Spencer K, et al. 2003. Population and breeding of the gentoo penguin *Pygoscelis papua*
29 835 at Marion Island, 1994/95-2002/03. *African Journal of Marine Science* 25:463-474.
- 30 836 Crawford RJM, Dyer BM, Upfold L, Makhado AB. 2014. Congruent, decreasing trends of gentoo
31 837 penguins and Crozet shags at sub-Antarctic Marion Island suggest food limitation through common
32 838 environmental forcing. *African Journal of Marine Science* 36:225-231.
- 33 839 Crawford RJM, Whittington PA, Upfold L, Ryan PG, Petersen SL, Dyer BM, Cooper J. 2009. Recent
34 840 trends in numbers of four species of penguins at the Prince Edward Islands. *African Journal of Marine*
35 841 *Science* 31:419-426.
- 36 842 Dalton DL, Vermaak E, Roelofse M, Kotze A. 2016a. Diversity in the Toll-Like Receptor Genes of
37 843 the African Penguin (*Spheniscus demersus*). *Plos One* 11.
- 38 844 Dalton DL, Vermaak E, Smit-Robinson HA, Kotze A. 2016b. Lack of diversity at innate immunity
39 845 Toll-like receptor genes in the Critically Endangered White-winged Flufftail (*Sarothrura ayresi*).
40 846 *Scientific Reports* 6.
- 41 847 Darfour-Oduro KA, Megens H-J, Roca A, Groenen MAM, Schook LB. 2016. Evidence for adaptation
42 848 of porcine Toll-like receptors. *Immunogenetics* 68:179-189.
- 43 849 Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and
44 850 parallel computing. *Nature Methods* 9:772-772.
- 45 851 de Dinechin M, Dobson FS, Zehtindjiev P, Metcheva R, Couchoux C, Martin A, Quillfeldt P,
46 852 Jouventin P. 2012. The biogeography of Gentoo Penguins (*Pygoscelis papua*). *Canadian Journal of*
47 853 *Zoology-Revue Canadienne De Zoologie* 90:352-360.
- 48 854 Dewar ML, Arnould JPY, Dann P, Trathan P, Groscolas R, Smith S. 2013. Interspecific variations in
49 855 the gastrointestinal microbiota in penguins. *Microbiologyopen* 2:195-204.
- 50 856 Dewar ML, Arnould JPY, Krause L, Dann P, Smith SC. 2014. Interspecific variations in the faecal
51 857 microbiota of Procellariiform seabirds. *Fems Microbiology Ecology* 89:47-55.
- 52 858 Dimitrov K, Metcheva R, Kenarova A. 2009. Salmonella presence - an indicator of direct and indirect
53 859 human impact on Gentoo in Antarctica. *Biotechnology & Biotechnological Equipment* 23:246-249.
- 54 860 Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L. 2007. Clinal variation in MHC diversity
55 861 with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic
56 862 salmon. *Evolution* 61:2154-2164.
- 57
58
59
60

- 1
2
3 863 Epstein CJ. 1967. Non-randomness of amino-acid changes in the evolution of homologous proteins.
4 864 Nature 215(5099):355-359.
- 5 865 Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform
6 866 population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.
- 7 867 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric
8 868 distances among DNA haplotypes - application to human mitochondrial-DNA restriction data.
9 869 Genetics 131:479-491.
- 10 870 Faber E, Tedin K, Speidel Y, Brinkmann MM, Josenhans C. 2018. Functional expression of TLR5 of
11 871 different vertebrate species and diversification in intestinal pathogen recognition. Scientific Reports 8.
12 872 Fan J, Li L, Han J, Ming H, Li J, Na G, Chen J. 2013. Diversity and structure of bacterial
13 873 communities in Fildes Peninsula, King George Island. Polar Biology 36:1385-1399.
- 14 874 Fix AS, Waterhouse C, Greiner EC, Stoskopf MK. 1988. Plasmodium relictum as a cause of avian
15 875 malaria in wild-caught magellanic penguins (*Spheniscus magellanicus*). Journal of Wildlife Diseases
16 876 24:610-619.
- 17 877 Flach EJ, Stevenson MF, Henderson GM. 1990. Aspergillosis in Gentoo penguins (*Pygoscelis papua*)
18 878 at Edinburgh Zoo, 1964 to 1988. Veterinary Record 126:81-85.
- 19 879 Freeman NM, Lovenduski NS, Gent PR. 2016. Temporal variability in the Antarctic Polar Front
20 880 (2002-2014). Journal of Geophysical Research-Oceans 121:7263-7276.
- 21 881 Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and
22 882 background selection. Genetics 147:915-925.
- 23 883 García-Peña FJ, Llorente MT, Serrano T, Ruano MJ, Belliure J, Benzal J, Herrera-Leon S, Vidal V,
24 884 D'Amico V, Perez-Boto D, et al. 2017. Isolation of *Campylobacter* spp. from Three Species of
25 885 Antarctic Penguins in Different Geographic Locations. Ecohealth 14:78-87.
- 26 886 Gewirtz AT, Navas TA, Lyons S, Godowski PJ, Madara JL. 2001. Cutting edge: Bacterial flagellin
27 887 activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. Journal
28 888 of Immunology 167:1882-1885.
- 29 889 Ghys MI, Raya Rey A, Schiavini A. 2008. Population trend and breeding biology of Gentoo penguin
30 890 in Martillo Island, Tierra Del Fuego, Argentina. Waterbirds 31(4):625-631.
- 31 891 Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, III, Smith HO. 2009. Enzymatic
32 892 assembly of DNA molecules up to several hundred kilobases. Nature Methods 6:343-U341.
- 33 893 Gilroy DL, van Oosterhout C, Komdeur J, Richardson, DS. 2017. Toll-like receptor variation in the
34 894 bottlenecked population of the endangered Seychelles warbler. Anim Conserv 20:235-250.
- 35 895 González-Acuna D, Hernandez J, Moreno L, Herrmann B, Palma R, Latorre A, Medina-Vogel G,
36 896 Kinsella MJ, Martin N, Araya K, et al. 2013. Health evaluation of wild gentoo penguins (*Pygoscelis*
37 897 *papua*) in the Antarctic Peninsula. Polar Biology 36:1749-1760.
- 38 898 Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science
39 899 185(4154):862-864)
- 40 900 González-Quevedo C, Spurgin LG, Carlos Illera J, Richardson DS. 2015. Drift, not selection, shapes
41 901 toll-like receptor variation among oceanic island populations. Molecular Ecology 24:5852-5863.
- 42 902 Goudet J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of
43 903 Heredity 86:485-486.
- 44 904 Grilo ML, Vanstreels RET, Wallace R, Garcia-Parraga D, Braga EM, Chitty J, Catao-Dias JL,
45 905 Madeira de Carvalho LM. 2016. Malaria in penguins - current perceptions. Avian Pathology 45:393-
46 906 407.
- 47 907 Grimaldi WW, Seddon PJ, Lyver POB, Nakagawa S, Tompkins DM. 2015. Infectious diseases of
48 908 Antarctic penguins: current status and future threats. Polar Biology 38:591-606.
- 49 909 Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic Positive Selection in the Evolution of Avian
50 910 Toll-Like Receptor Innate Immunity Genes. Plos One 9.
- 51 911 Grueber CE, Wallis GP, Jamieson IG. 2013. Genetic drift outweighs natural selection at toll-like
52 912 receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular
53 913 Ecology 22:4470-4482.
- 54 914 Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at Innate Immunity Toll-Like
55 915 Receptor Genes in a Bottlenecked Population of a New Zealand Robin. Plos One 7.
- 56
57
58
59
60

- 1
2
3 916 Guégan JF, Prugnolle F, Thomas F. 2007. Global spatial patterns of infectious diseases and human
4 917 evolution. In: Stearns SC, Koella JC, editors. *Evolution in health and disease*. Oxford Scholarship
5 918 Online. DOI: 10.1093/acprof:oso/9780199207466.001.0001 .
6 919 Guernier V, Hochberg ME, Guégan JF. 2004. Ecology drives the worldwide distribution of human
7 920 diseases. *PLoS Biol* 2:e141.
8 921 Hartmann SA, Schaefer HM, Segelbacher G. 2014. Genetic depletion at adaptive but not neutral loci
9 922 in an endangered bird species. *Molecular Ecology* 23:5712-5725.
10 923 Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human-ape splitting by a molecular clock of
11 924 mitochondrial-DNA. *Journal of Molecular Evolution* 22:160-174.
12 925 Hinke JT, Cossio AM, Goebel ME, Reiss CS, Trivelpiece WZ, Watters GM. 2017. Identifying Risk:
13 926 Concurrent Overlap of the Antarctic Krill Fishery with Krill-Dependent Predators in the Scotia Sea.
14 927 *Plos One* 12.
15 928 Jaeger A, Lebarbenchon C, Bourret V, Bastien M, Lagadec E, Thiebot J-B, Boulinier T, Delord K,
16 929 Barbraud C, Marteau C, et al. 2018. Avian cholera outbreaks threaten seabird species on Amsterdam
17 930 Island. *Plos One* 13.
18 931 Käll L, Krogh A, Sonnhammer ELL. 2004. A Combined Transmembrane Topology and Signal
19 932 Peptide Prediction Method. *Journal of Molecular Biolog*, 338(5):1027-1036.
20 933 Kane OJ, Uhart MM, Rago V, Pereda AJ, Smith JR, Van Buren A, Clark JA, Boersma PD. 2012.
21 934 Avian Pox in Magellanic Penguins (*Spheniscus magellanicus*). *Journal of Wildlife Diseases* 48:790-
22 935 794.
23 936 Kawai T, Akira S. 2006. TLR signaling. *Cell Death and Differentiation* 13:816-825.
24 937 Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through
25 938 comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.
26 939 Kloch A, Wenzel MA, Laetsch DR, Michalski O, Welc-Faleciak R, Piertney SB. 2018. Signatures of
27 940 balancing selection in Toll-like receptor (TLRs) genes novel insights from a free-living rodent.
28 941 *Scientific Reports* 8.
29 942 Knafler GJ, Clark JA, Boersma PD, Bouzat JL. 2012. MHC Diversity and Mate Choice in the
30 943 Magellanic Penguin, *Spheniscus magellanicus*. *Journal of Heredity* 103:759-768.
31 944 Korbie DJ, Mattick JS. 2008. Touchdown PCR for increased specificity and sensitivity in PCR
32 945 amplification. *Nature Protocols* 3:1452-1456.
33 946 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable, and
34 947 user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics (Oxford, England)*.
35 948 Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction.
36 949 *Methods in Ecology and Evolution* 6:1110-1116.
37 950 Lemaho Y, Karmann H, Briot D, Handrich Y, Robin JP, Mioskowski E, Cherel Y, Farni J. 1992.
38 951 Stress in birds due to routine handling and a technique to avoid it. *American Journal of Physiology*
39 952 263:R775-R781.
40 953 Lepage D, Vaidya G, Guralnick R. 2014. Avibase - a database system for managing and organizing
41 954 taxonomic concepts. *Zookeys*:117-135.
42 955 Lescroel A, Bost CA. 2006. Recent decrease in gentoo penguin populations at Iles Kerguelen.
43 956 *Antarctic Science* 18:171-174.
44 957 Levin II, Zwiers P, Deem SL, Geest EA, Higashiguchi JM, Iezhova TA, Jimenez-Uzategui G, Kim
45 958 DH, Morton JP, Perlut NG, et al. 2013. Multiple Lineages of Avian Malaria Parasites (*Plasmodium*)
46 959 in the Galapagos Islands and Evidence for Arrival via Migratory Birds. *Conservation Biology*
47 960 27:1366-1377.
48 961 Levy H, Clucas GV, Rogers AD, Leache AD, Ciborowski KL, Polito MJ, Lynch HJ, Dunn MJ, Hart
49 962 T. 2016. Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across
50 963 the Scotia Arc. *Ecology and Evolution* 6:1834-1853.
51 964 Lund JM, Alexopoulou L, Sato A, Karow M, Adams NC, Gale NW, Iwasaki A, Flavell RA. 2004.
52 965 Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the National*
53 966 *Academy of Sciences of the United States of America* 101:5598-5603.
54 967 Lynch HJ, Naveen R, Trathan PN, Fagan WF. 2012. Spatially integrated assessment reveals
55 968 widespread changes in penguin populations on the Antarctic Peninsula. *Ecology* 93:1367-1377.
56
57
58
59
60

- 1
2
3 969 Ma D, Zhu R, Ding W, Shen C, Chu H, Lin X. 2013. Ex-situ enzyme activity and bacterial
4 970 community diversity through soil depth profiles in penguin and seal colonies on Vestfold Hills, East
5 971 Antarctica. *Polar Biology* 36:1347-1361.
- 6 972 Mac Donald JW, Conroy JWH. 1971. Virus disease resembling puffinosis in the gentoo penguin
7 973 (*Pygoscelis papua*) on signy island south orkney islands. *British Antarctic Survey Bulletin*:80-82.
- 8 974 Mayewski P. 2012. State of the Antarctic climate system. Excerpts from SASOCS (Mayewski et al.
9 975 2009). *Anales del Instituto de la Patagonia* 40:25-30.
- 10 976 Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N. 2012. Molecular evolution of
11 977 vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their
12 978 TIR domains. *Gene* 503:235-243.
- 13 979 Miyata T. 1979. Two types of amino acid substitution in protein evolution. *Journal of Molecular*
14 980 *Evolution* 12(3):219-236
- 15 981 Moore JK, Abbott MR, Richman JG. 1999. Location and dynamics of the Antarctic Polar Front from
16 982 satellite sea surface temperature data. *Journal of Geophysical Research-Oceans* 104:3059-3073.
- 17 983 Morgan IR, Westbury HA. 1988. Studies of viruses in penguins in the Vestfold Hills. *Hydrobiologia*
18 984 165:263-269.
- 19 985 Mukherjee S, Ganguli D, Majumder PP. 2014. Global Footprints of Purifying Selection on Toll-Like
20 986 Receptor Genes Primarily Associated with Response to Bacterial Infections in Humans. *Genome*
21 987 *Biology and Evolution* 6:551-558.
- 22 988 Nahori MA, Fournie-Amazouz E, Que-Gewirth NS, Balloy V, Chignard M, Raetz CRH, Saint Girons
23 989 I, Werts C. 2005. Differential TLR recognition of leptospiral lipid A and lipopolysaccharide in muine
24 990 and human cells. *Journal of Immunology* 175:6022-6031.
- 25 991 Neira V, Tapia R, Verdugo C, Barriga G, Mor S, Ng TFF, García V, Del Río J, Rodrigues P, Briceño
26 992 C, et al. 2017. Novel Avulaviruses in Penguins, Antarctica. *Emerg Infect Dis* 23:1212-1214.
- 27 993 Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic*
28 994 *Acids Research* 31(13):3812-3814.
- 29 995 Novak K, Bjelka M, Samake K, Valcikova T. 2019. Potential of TLR-gene diversity in Czech
30 996 indigenous cattle for resistance breeding as revealed by hybrid sequencing. *Archives Animal Breeding*
31 997 62:477-490.
- 32 998 Olivares F, Tapia R, Gálvez C, Meza F, Barriga GP, Borrás-Chavez R, Mena-Vasquez J, Medina RA,
33 999 Neira V. 2019. Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the
34 1000 Antarctic Peninsula. *Transbound Emerg Dis*.
- 35 1001 Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. 2009. The structural basis of
36 1002 lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458:1191-U1130.
- 37 1003 Peña M F, Poulin E, Dantas GPM, González-Acuña D, Petry MV, Vianna JA. 2014. Have Historical
38 1004 Climate Changes Affected Gentoo Penguin (*Pygoscelis papua*) Populations in Antarctica? *Plos One* 9.
39 1005 Pertierra L SN, Martínez P, Barbosa A, Raya Rey A, Pistorius P, Polanowski A, Bonadonna F, Le
40 1006 Bohec C, Bi K, et al. in review. Integrated phylogenomic and niche analyses reveal cryptic speciation
41 1007 in Gentoo penguins driven by local adaptation. In. *Diversity and Distribution*.
- 42 1008 Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*
43 1009 14:817-818.
- 44 1010 Qureshi ST, Lariviere L, Leveque G, Clermont S, Moore KJ, Gros P, Malo D. 1999. Endotoxin-
45 1011 tolerant mice have mutations in toll-like receptor 4 (Tlr4). *Journal of Experimental Medicine* 189.
- 46 1012 Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in
47 1013 Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology* 67:901-904.
- 48 1014 Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The
49 1015 evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci U S A* 102:9577-9582.
- 50 1016 Roeder AD, Ritchie PA, Lambert DM. 2002. New DNA markers for penguins. *Conservation Genetics*
51 1017 3:341-344.
- 52 1018 Rohde K, Heap M. 1998. Latitudinal differences in species and community richness and in
53 1019 community structure of metazoan endo- and ectoparasites of marine teleost fish. *Int J Parasitol*
54 1020 28:461-474.
- 55 1021 Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE,
56 1022 Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.
57 1023 *Molecular Biology and Evolution* 34:3299-3302.

- 1
2
3 1024 Sallaberry-Pincheira N, González-Acuña D, Herrera-Tello Y, Dantas GP, Luna-Jorquera G, Frere E,
4 1025 Valdés-Velasquez A, Simeone A, Vianna JA. 2015. Molecular Epidemiology of Avian Malaria in
5 1026 Wild Breeding Colonies of Humboldt and Magellanic Penguins in South America. *Ecohealth* 12:267-
6 1027 277.
7 1028 Sallaberry-Pincheira N, González-Acuña D, Padilla P, Dantas GPM, Luna-Jorquera G, Frere E,
8 1029 Valdés-Velasquez A, Vianna JA. 2016. Contrasting patterns of selection between MHC I and II across
9 1030 populations of Humboldt and Magellanic penguins. *Ecol Evol* 6:7498-7510.
10 1031 Sneath PHA. 1966. Relations between chemical structure and biological activity in peptides. *Journal*
11 1032 *of Theoretical Biology* 12(2):157-195.
12 1033 Song WS, Jeon YJ, Namgung B, Hong M, Yoon SI. 2017. A conserved TLR5 binding and activation
13 1034 hot spot on flagellin. *Scientific Reports* 7: 40878
14 1035 Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from
15 1036 population data. *American Journal of Human Genetics* 68:978-989.
16 1037 Stonehouse B. 1970. Geographic Variation in Gentoo Penguins *Pygoscelis*-Papua. *Ibis* 112:52-57.
17 1038 Subramanian S, Beans-Picon G, Swaminathan SK, Millar CD, Lambert DM. 2013. Evidence for a
18 1039 recent origin of penguins. *Biology Letters* 9.
19 1040 Świderská Z, Smidova A, Buchtova L, Bryjova A, Fabianova A, Munclinger P, Vinkler M. 2018.
20 1041 Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic
21 1042 chicken breeds. *Scientific Reports* 8.
22 1043 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
23 1044 *Genetics* 123:585-595.
24 1045 Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of
25 1046 mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512-526.
26 1047 Trathan PN, Forcada J, Murphy EJ. 2007. Environmental forcing and Southern Ocean marine predator
27 1048 populations: effects of climate change and variability. *Philosophical Transactions of the Royal Society*
28 1049 *B-Biological Sciences* 362:2351-2365.
29 1050 Trivelpiece WZ, Trivelpiece SG, Volkman NJ. 1987. Ecological segregation of Adelie, Gentoo, and
30 1051 Chinstrap penguins at King George Island, Antarctica. *Ecology* 68:351-361.
31 1052 Tsuda TT, Tsuda M, Naruse T, Kawata H, Ando A, Shiina T, Fukuda M, Kurita M, LeMaho I, Kulski
32 1053 JK, et al. 2001. Phylogenetic analysis of penguin (*Spheniscidae*) species based on sequence variation
33 1054 in MHC class II genes. *Immunogenetics* 53:712-716.
34 1055 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-
35 1056 new capabilities and interfaces. *Nucleic Acids Research* 40.
36 1057 Urbina D, Tang B, Higgs PG. 2006. The response of amino acid frequencies to directional mutation
37 1058 pressure in mitochondrial genome sequences is related to the physical properties of the amino acids
38 1059 and to the structure of the genetic code. *Journal of Molecular Evolution* 62(3):340-361.
39 1060 Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-Like Receptor Evolution in Birds:
40 1061 Gene Duplication, Pseudogenization, and Diversifying Selection. *Molecular Biology and Evolution*
41 1062 35:2170-2184.
42 1063 Vianna JA, Noll D, Dantas GPM, Petry MV, Barbosa A, Gonzalez-Acuña D, Le Bohec C, Bonadonna
43 1064 F, Poulin E. 2017. Marked phylogeographic structure of Gentoo penguin reveals an ongoing
44 1065 diversification process along the Southern Ocean. *Molecular Phylogenetics and Evolution* 107:486-
45 1066 498.
46 1067 Vinkler, M., Bainová, H., Bryjová, A. Tomášek, O, Albrecht, T, Bryja, J. 2015. Characterisation of
47 1068 Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. *Genetica* 143(1):101-
48 1069 112.
49 1070 Wallensten A, Munster VJ, Osterhaus A, Waldenstrom J, Bonnedahl J, Broman T, Fouchier RAM,
50 1071 Olsen B. 2006. Mounting evidence for the presence of influenza A virus in the avifauna of the
51 1072 Antarctic region. *Antarctic Science* 18:353-356.
52 1073 Walsh SA, Suarez ME. 2006. New penguin remains from the Pliocene of northern Chile. *Historical*
53 1074 *Biology* 18:119-130.
54 1075 Wang J, Zhang Z, Liu J, Zhao J, Yin D. 2016. Ectodomain Architecture Affects Sequence and
55 1076 Functional Evolution of Vertebrate Toll-like Receptors. *Scientific Reports* 6.
56 1077 Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure.
57 1078 *Evolution* 38:1358-1370.

- 1
2
3 1079 Wille M, Aban M, Wang J, Moore N, Shan S, Marshall J, González-Acuña D, Vijaykrishna D, Butler
4 1080 J, Hall RJ, et al. 2019. Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *J Virol*
5 1081 93.
6 1082 Wilson RP, Alvarez B, Latorre L, Adelung D, Culik B, Bannasch R. 1998. The movements of gentoo
7 1083 penguins *Pygoscelis papua* from Ardley Island, Antarctica. *Polar Biology* 19:407-413.
8 1084 Wlasiuk G, Nachman MW. 2010. Adaptation and Constraint at Toll-Like Receptors in Primates.
9 1085 *Molecular Biology and Evolution* 27:2172-2186.
10 1086 Woods R, Jones HI, Watts J, Miller GD, Shellam GR. 2009. Diseases of Antarctic seabirds. *Health of*
11 1087 *Antarctic wildlife: a challenge for science and policy.*:35-55.
12 1088 Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-
13 1089 1591.
14 1090 Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood.
15 1091 *Computer Applications in the Biosciences* 13:555-556.
16 1092 Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under
17 1093 positive selection. *Mol Biol Evol* 22:1107-1118.
18 1094 Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA. 2012. Structural
19 1095 basis of TLR5-flagellin recognition and signaling. *Science* 335(6070):859-864.
20 1096 Younger J, Emmerson L, Southwell C, Lelliott P, Miller K. 2015. Proliferation of East Antarctic
21 1097 Adelie penguins in response to historical deglaciation. *Bmc Evolutionary Biology* 15.
22 1098 Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. 2015.
23 1099 Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the
24 1100 last glacial maximum. *Global Change Biology* 21:2215-2226.
25 1101 Zdanowski MK, Weglenski P, Golik P, Sasin JM, Borsuk P, Zmuda MJ, Stankovic A. 2004. Bacterial
26 1102 diversity in Adelie penguin, *Pygoscelis adeliae*, guano: molecular and morpho-physiological
27 1103 approaches. *Fems Microbiology Ecology* 50:163-173.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures

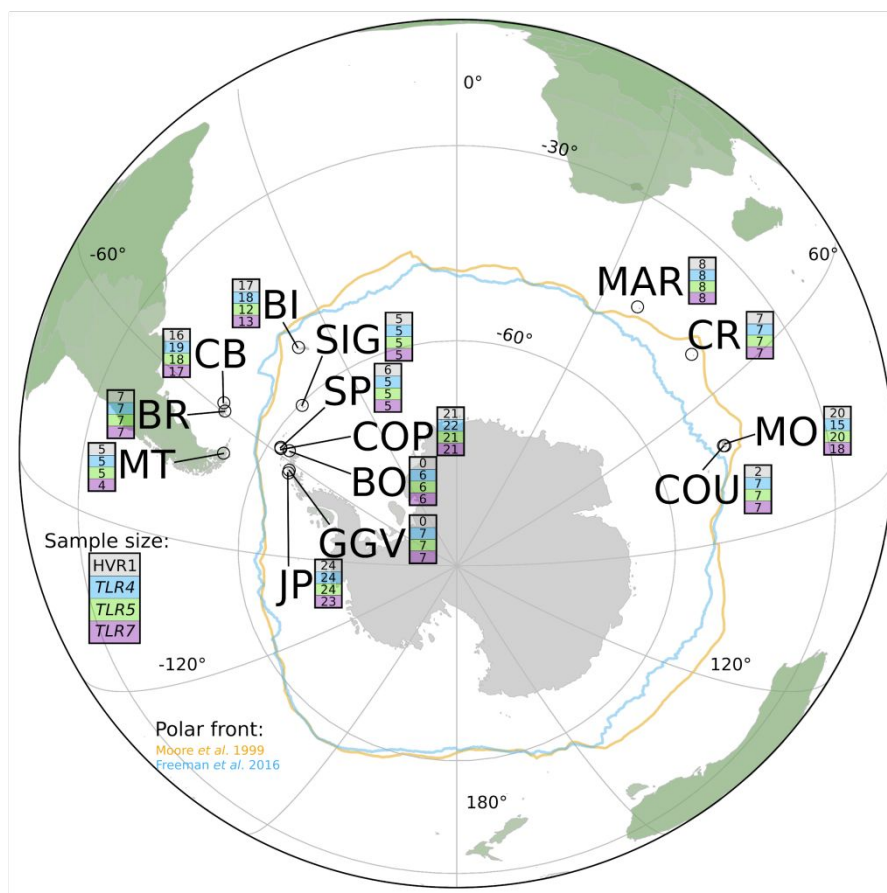
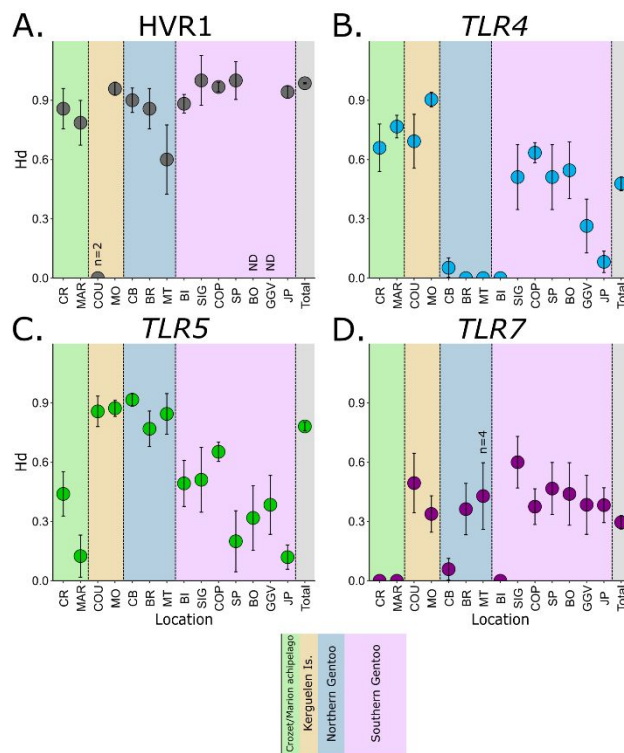


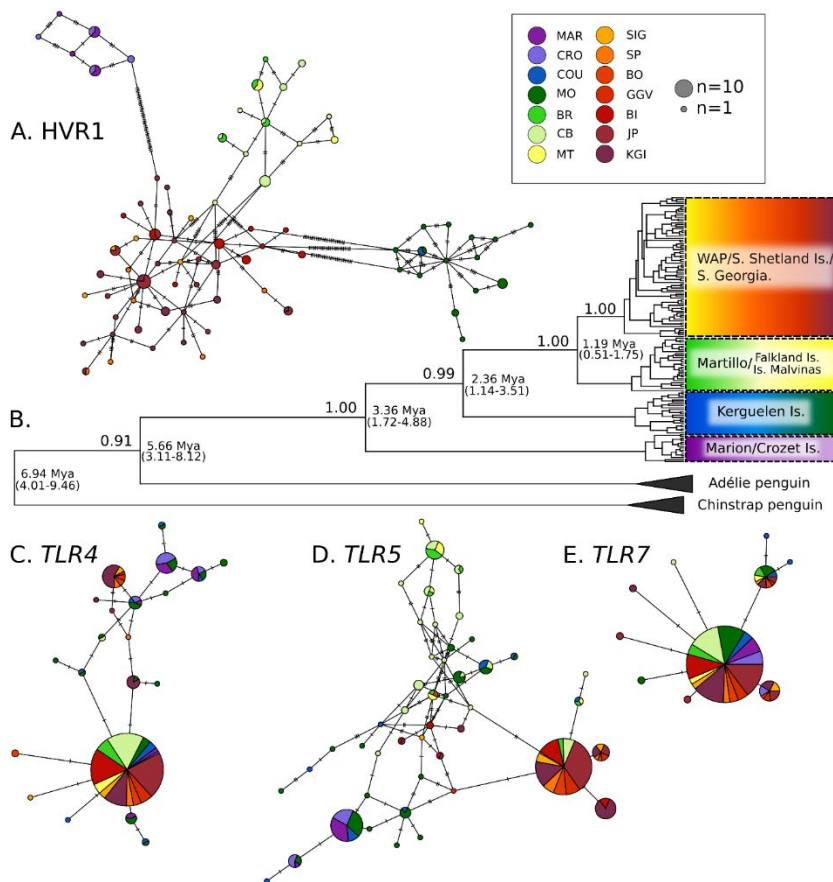
Figure 1. Locations and per-locus sample sizes for each sampled Gentoo penguin colony in the Southern Ocean and Antarctica. Solid colored lines represent the reported position of the Polar Front, based on the analyses of Freeman et al. (2016) and Moore et al. (1999). Depending on the analysis, the Kerguelen and Crozet Islands can lie just north or just south of the Polar Front (Moore, et al. 1999; Freeman, et al. 2016). (CR = Crozet Island; MAR = Marion Island; COU = Courbet Peninsula, Kerguelen; MO = Pointe du Morne, Kerguelen; CB = Cow Bay, Falkland/Malvinas Islands; BR = Bull Roads, Falkland/Malvinas Islands; BI = Bird Island, South Georgia; MT = Martillo Island, Tierra del Fuego; SIG = Signy Island, South Orkney Islands; COP = Copacabana (Admiralty Bay), King George Island, South Shetland Islands; SP = Stranger Point, King George Island, South Shetland Islands; BO = Bernardo O'Higgins Base, Western Antarctic Peninsula; GGV = Gabriel González Videla Base, Western Antarctic Peninsula; JP = Jougla Point, Western Antarctic Peninsula).

1117 *Figure 2.* Haplotype diversity (Hd) for the hypervariable
 1118 mitochondrial control region (HVR1; A) and three TLR
 1119 loci (B, C, D), for all sampling sites for which data was
 1120 available. All sites had $n \geq 5$ unless otherwise indicated
 1121 (ND = no data). Location abbreviations are the same as in
 1122 previous figure.

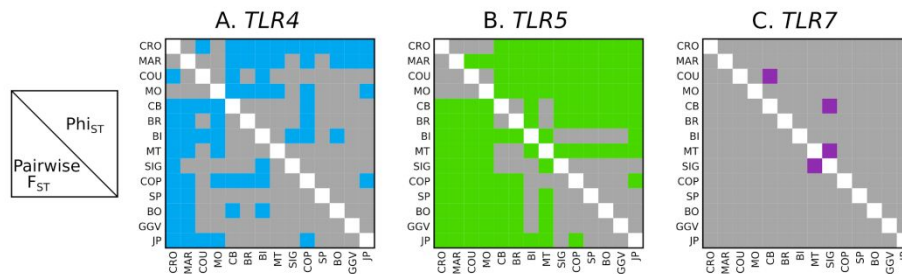
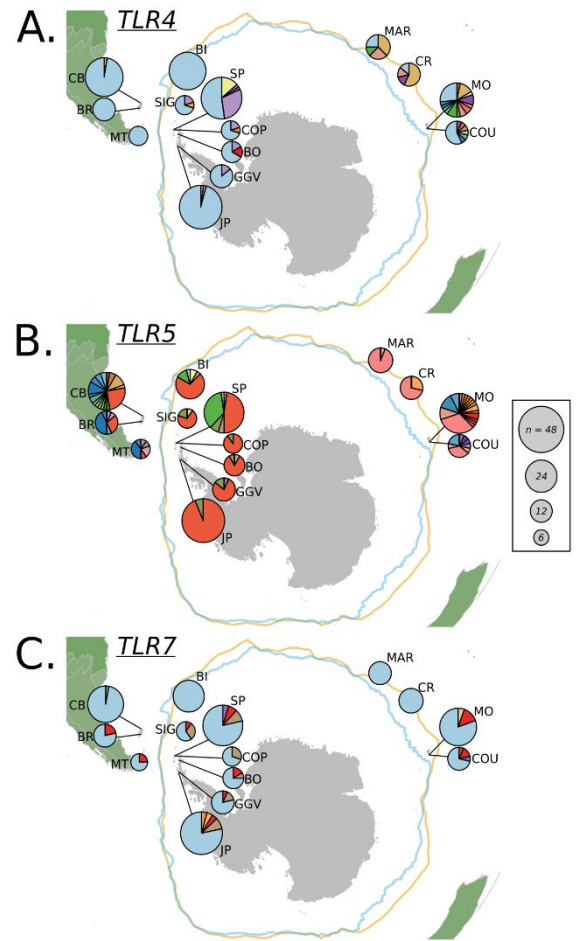


1123

1124 *Figure 3.* Minimum spanning haplotype
 1125 networks for mtDNA HVR1 (A), TLR4
 1126 (C), TLR5 (D), and TLR7 (E) for Gentoo
 1127 penguin colonies, along with mtDNA
 1128 HVR1 maximum clade credibility tree
 1129 (B), using congeneric penguin species as
 1130 outgroups. Location abbreviations are the
 1131 same as in previous figures. For minimum
 1132 spanning haplotype networks, pie charts
 1133 represent single haplotypes, while
 1134 segment size refers to the contribution of
 1135 individual sampled sites to the proportion
 1136 of overall haplotype frequency. Size of pie
 1137 charts reflects the number of individual
 1138 birds with the observed haplotype. Dashes
 1139 on connecting lines each denote one
 1140 nucleotide change.



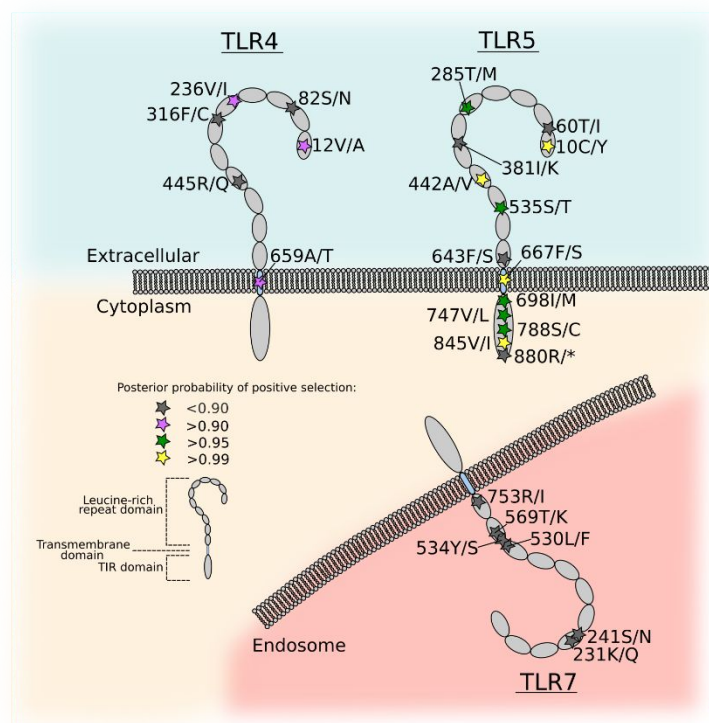
1
2
3 1141 *Figure 4.* Haplotype diversity across Gentoo penguin sample
4 1142 populations for (A) *TLR4*, (B) *TLR5*, and (C) *TLR7*. For each
5 1143 locus, different colors represent unique haplotypes and each
6 1144 segment size reflects the proportion of birds in each location
7 1145 with that haplotype. Overall size of the pie chart reflects the
8 1146 number of birds sampled in each location. Location
9 1147 abbreviations are the same as in previous figures.



11 1149 *Figure 5.* Visualization of pairings of Gentoo penguin breeding colonies, where pairwise F_{ST} values
12 1150 (below diagonal) or Φ_{ST} values (above diagonal) with significance $p < 0.01$ after correction for multiple
13 1151 tests using SGOF+ (Carvajal-Rodriguez and de Uña-Alvarez 2011) are shown in color for (A) *TLR4*, (B)
14 1152 *TLR5*, and (C) *TLR7*. Location abbreviations are the same as in previous figures.

15 1153

1
2
3 1155 *Figure 6.* Positions of polymorphic and positively
4 1156 selected sites in *P. papua* TLR4, 5 and 7. Schematic
5 1157 diagrams of TLRs in the extracellular (TLR4 and 5)
6 1158 and endosomal (TLR7) compartments show
7 1159 positions of amino acid variants resulting from non-
8 1160 synonymous nucleotide substitutions. Variant
9 1161 positions are marked with stars and are colored
10 1162 according to the likelihood of positive selection as
11 1163 determined by the M2a model in the *codeml*
12 1164 program in PAML.



13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29 1169 *Figure 7.* Distribution and functional
30 1170 differences of two selected amino
31 1171 acid residues in TLR5. (A) Mapped
32 1172 areas represent the four clades of *P.*
33 1173 *papua* as determined by HVR1
34 1174 analysis. Pie charts represent the
35 1175 allele proportions at the two selected
36 1176 residues in TLR5, and chart area is
37 1177 proportional to the number of
38 1178 samples. (B) Bar chart displays NF-
39 1179 kB response of the two genotypes
40 1180 following stimulation with FLA-ST.

