

**Admixture between ancient lineages, selection, and the formation of sympatric stickleback species-pairs**

Dean, Laura; Magalhaes, Isabel; Foote, Andrew; D'Agostino, Daniele; McGowan, Suzanne; MacColl, Andrew

Molecular Biology and Evolution

DOI:

[10.1093/molbev/msz161](https://doi.org/10.1093/molbev/msz161)

Published: 01/11/2019

Peer reviewed version

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

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

Dean, L., Magalhaes, I., Foote, A., D'Agostino, D., McGowan, S., & MacColl, A. (2019). Admixture between ancient lineages, selection, and the formation of sympatric stickleback species-pairs. *Molecular Biology and Evolution*, 36(11), 2481-2497. <https://doi.org/10.1093/molbev/msz161>

Hawliau Cyffredinol / General rights

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

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

Take down policy

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

Admixture between ancient lineages, selection, and the formation of sympatric stickleback species-pairs

Laura L. Dean¹, Isabel S. Magalhaes^{1,2}, Andrew Foote³, Daniele D'Agostino¹, Suzanne McGowan⁴ and Andrew D. C. MacColl¹

¹School of Life Sciences, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK

²Department of Life Sciences, Whitelands College, University of Roehampton, London SW15 4JD, UK

³Molecular Ecology and Fisheries Genetics Laboratory, Bangor University, Bangor, Gwynedd, LL57 2DG, UK

⁴School of Geography, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Corresponding author Laura L. Dean: laura.dean@nottingham.ac.uk

Divergence in stickleback species-pairs

Key words: *Gasterosteus aculeatus*, reproductive isolation, three-spined stickleback, admixture, adaptive radiation, speciation

1 **Abstract (250 words)**

2 Ecological speciation has become a popular model for the development and maintenance of
3 reproductive isolation in closely related sympatric pairs of species or ecotypes. An implicit
4 assumption has been that such pairs originate (possibly with gene flow) from a recent,
5 genetically homogeneous ancestor. However, recent genomic data has revealed that currently
6 sympatric taxa are often a result of secondary contact between ancestrally allopatric lineages.
7 This has sparked an interest in the importance of initial hybridization upon secondary contact,
8 with genomic re-analysis of classic examples of ecological speciation often implicating
9 admixture in speciation. We describe a novel occurrence of unusually well-developed
10 reproductive isolation in a model system for ecological speciation: the three-spined stickleback
11 (*Gasterosteus aculeatus*), breeding sympatrically in multiple lagoons on the Scottish island of
12 North Uist. Using morphological data, targeted genotyping and genome-wide single nucleotide
13 polymorphism (SNP) data we show that lagoon resident and anadromous ecotypes are strongly
14 reproductively isolated with an estimated hybridization rate of only ~1%. We use
15 palaeoecological and genetic data to test three hypotheses to explain the existence of these
16 species-pairs. Our results suggest that recent, purely ecological speciation from a genetically
17 homogeneous ancestor is probably not solely responsible for the evolution of species-pairs.
18 Instead we reveal a complex colonisation history with multiple ancestral lineages contributing
19 to the genetic composition of species-pairs, alongside strong disruptive selection. Our results
20 imply a role for admixture upon secondary contact and are consistent with the recent suggestion
21 that the genomic underpinning of ecological speciation often has an older, allopatric origin.

22 **Introduction**

23 The sympatric co-existence of closely related but reproductively isolated ‘species-pairs’
24 represents a biological conundrum because it is difficult to explain how divergence is
25 maintained in the absence of obvious barriers to gene flow. Ecological speciation, in which
26 ecologically-dependent natural selection drives the evolution of variation in adaptive traits that
27 also influence reproductive isolation between ecotypes, has become a popular model to account
28 for this problem (Schluter 1996; Rundle and Nosil 2005; Schluter 2009; Nosil 2012). However,
29 it is becoming increasingly apparent that ecologically derived selection rarely results in
30 complete speciation, but rather tends to cause only partial divergence, either with reduced (but
31 still some) genome-wide gene flow between, compared to within ecotypes, or only certain
32 regions of the genome being unaffected by gene flow (Hendry 2009; Hendry et al. 2009).

33 Increasing resolution of genomic data have made it possible to untangle how the demographic
34 history of species-pairs varies across the genome, bringing potential for a much clearer picture
35 of how historic periods of secondary contact and admixture events have shaped the evolutionary
36 trajectory of species (Sousa and Hey 2013). Genomic data are beginning to suggest that stronger
37 recent divergence is often underlain by more ancient genetic incompatibilities, or adaptive
38 variation, that evolved in allopatry (Seehausen et al. 2014; Marques et al. 2019). As such, many
39 currently sympatric taxa, which appear to have diverged *in situ*, have later been found to be the
40 result of secondary contact following a long period of allopatry (Bernatchez and Dodson 1990;
41 Feder et al. 2003; Kuehne et al. 2007; Foote and Morin 2015; Lucek et al. 2018). These ancient
42 alleles contributing to speciation can be a result of standing genetic variation that is re-used to
43 seed new ecological divergences and reproductive barriers (Feder et al. 2005; Jones et al.
44 2012b) or admixture / introgression that provides genetic material for both adaptation and
45 reproductive isolation in the face of gene flow (Seehausen 2004; Keller et al. 2013; Marques et
46 al. 2019). The importance of genetic admixture upon secondary contact for speciation has long
47 been known in plants (Grant 1971; Rieseberg et al. 2003; Soltis and Soltis 2009), but in the
48 light of the realisation that purely ecological speciation may be rare (Hendry 2009; Wang et al.
49 2013), there has been a recent, rising awareness that hybridisation and admixture events are
50 often also involved in animal speciation (Mallet 2007; Abbott et al. 2013; Feder et al. 2013).
51 Many classic examples of ecological speciation, such as Darwin’s finches (Grant and Grant
52 2009) and adaptive radiations such as the cichlids of Lake Victoria (Kocher 2004; Terai et al.

53 2006) and the *Anopheles* species-complex (Simard et al. 2009), are turning out either to be
54 cryptic cases of homoploid hybrid speciation (Lamichhaney et al. 2018) or to have involved
55 genetic admixture events (Fontaine et al. 2015; Meier et al. 2017; Meier et al. 2018; Marques
56 et al. 2019). In this new genomic era, it is therefore necessary to reassess the mechanisms
57 responsible for speciation, particularly in model systems of adaptive radiation.

58 Divergence among populations of three-spined stickleback (*Gasterosteus aculeatus* L, hereafter
59 'stickleback') is common throughout their Holarctic range (Bell et al. 2004). Truly marine, sea-
60 spawning stickleback have repeatedly colonised coastal brackish and freshwater habitats
61 following the Pleistocene glacial retreat, giving rise to anadromous (migratory fish which spend
62 most of their lives in the sea but migrate into fresh or brackish water to spawn), lagoon resident
63 (fish that live year-round in shallow brackish coastal lagoons) and freshwater resident (fish that
64 live year round in freshwater lakes and streams) ecotypes (Taylor and McPhail 2000;
65 McKinnon et al. 2004). Very little is known about truly marine stickleback (Ahnelt 2018), but
66 studies of anadromous and resident (resident in either fresh or brackish water) populations show
67 that differences in ecological selection pressures acting between oceanic and enclosed waters
68 (e.g. lakes, streams and lagoons) have shaped the replicated heritable changes in morphology
69 and behaviour that are associated with these parallel transitions (Schluter and McPhail 1992;
70 McKinnon and Rundle 2002; McKinnon et al. 2004; Schluter et al. 2004; Jones et al. 2012b).
71 Whilst there is substantial evidence of divergent natural selection and genetic differentiation
72 between anadromous and resident ecotypes (Hagen 1967; McKinnon and Rundle 2002; Von
73 Hippel and Weigner 2004; Jones et al. 2006), extensive hybridisation is apparent across most
74 parapatric contact zones (Heuts 1947; Hagen 1967; Hay and McPhail 2000; Jones et al. 2006).
75 There are only a handful of cases in which admixed, morphologically intermediate individuals
76 are completely absent (Ziuganov 1995; Karve et al. 2008) and fewer still with direct genetic
77 evidence for the absence of admixture (Drevecky et al. 2013). Therefore, as with many other
78 species, it has been concluded that ecologically mediated selection alone is not sufficient in this
79 system for speciation to reach completion in the face of gene flow (Hendry 2009; Schluter and
80 Conte 2009).

81 There are a few, rare cases in which speciation has progressed further along the continuum of
82 reproductive isolation in stickleback. The most unambiguous example of speciation to
83 completion is between two marine forms: Pacific Ocean and Japan Sea stickleback (Higuchi

84 and Goto 1996). This is a clear example of intrinsic genetic speciation that is allopatric and
85 ancient in origin (Kitano et al. 2007; Kitano et al. 2009) and has occurred in the face of ongoing
86 gene flow (Ravinet et al. 2018). A second example is freshwater benthic-limnetic species-pairs
87 in multiple lakes around the Strait of Georgia, BC, Canada, which show strong reproductive
88 isolation with unusually low (~5%) hybridisation (McPhail 1992; Gow et al. 2006; Gow et al.
89 2008). The origins of benthic-limnetic pairs are much less clear. Ecologically-mediated
90 selection is important for maintaining distinct benthic and limnetic ecotypes (Schluter and
91 McPhail 1992; Schluter 1996; McKinnon and Rundle 2002), but is unlikely to be wholly
92 accountable for speciation (Hendry 2009). It was initially thought that benthic-limnetic pairs
93 evolved independently in each lake from a single, homogeneous ‘stock’ marine population and
94 the effects of ecological divergence were exaggerated because post-glacial fluctuations in
95 relative sea-level (RSL) caused a ‘double-invasion’ of marine fish, with an intermediate period
96 of spatial isolation in which gene-flow was prevented (McPhail 1993; Taylor and McPhail
97 2000; Rundle and Schluter 2004). This is the ‘classic’ model proposed to explain cases of
98 unusually strong reproductive isolation in many post-glacial fish (Ferguson and Taggart 1991;
99 Schluter 1996; Volpe and Ferguson 1996; Nesbo et al. 1999), but it is rarely, if ever, empirically
100 tested. Furthermore, RSL reconstructions for the Strait of Georgia are not consistent with a
101 double-invasion for benthic-limnetic pairs (Friele and Hutchinson 1993; Josenhans et al. 1997;
102 Hutchinson et al. 2004) and Jones et al. (2012a) showed that some of the adaptive genetic
103 variation in each lake arises from shared adaptive variants, rather than from novel mutations
104 that have arisen separately in each lake, suggesting a more important role for allopatric adaptive
105 divergence and re-use of standing genetic variation in the evolution of benthic-limnetic species-
106 pairs. Some prior genetic differentiation is therefore probably necessary for speciation to
107 progress beyond low level reproductive isolation under conditions of ecologically derived
108 selection in stickleback.

109 The island of North Uist (hereafter ‘Uist’), Scottish Western Isles, like most of the rest of
110 northern Europe, was likely colonised by marine stickleback following the melting of ice sheets
111 ~16,000 YBP (Colosimo et al. 2005; Jones et al. 2012b). The island comprises a series of
112 complex isolated or interconnected freshwater lakes and brackish coastal lagoons, which cover
113 almost one third of the land surface of the island, making it ideal for studying the oceanic-
114 resident radiation of stickleback. Here we use high resolution, genome-wide single nucleotide
115 polymorphism (SNP) data alongside targeted genotyping and morphological analysis to

116 identify apparently stable, strongly isolated, sympatrically breeding anadromous-lagoon
117 resident stickleback species-pairs in multiple brackish coastal lagoons on Uist (see Figure 1a-c
118 for phenotypic examples of species-pair parental and intermediate phenotypes). We then
119 combine genetic and palaeoecological data to test three possible hypotheses, which are not
120 mutually exclusive, to explain the origin of these previously unexplored species-pairs: (1) the
121 ‘classic’ stickleback model, in which multiple colonisations from a single, homogeneous
122 marine population occurred as a result of a double-invasion facilitated by changes in RSL, and
123 speciation occurred purely as a result of recent ecological divergent selection that occurred *in*
124 *situ* during the Holocene. This possibility is supported by some previous evidence that suggests
125 a spike (‘high-stand’) in RSL immediately after deglaciation in the Hebrides, followed by RSL
126 receding until ~10,000-14,000 YBP, before rising again to the present day (Jordan et al. 2010).
127 However, it is also clear that local patterns of sea-level change can be very variable as a result
128 of differences in glaciation and solid geology. This hypothesis makes four predictions. First,
129 there is evidence for strong, ecologically-based selection between divergent ecotypes; second,
130 the species-pairs originated from multiple colonisations from a genetically homogeneous
131 ‘stock’ marine population, with no prior genetic or behavioural isolation (Rundle and Schluter
132 2004). Third, the species-pairs are post-glacial in age as divergence can only have occurred
133 since the deglaciation of Uist ~16,000 YBP (Ballantyne 2010), and fourth, the lagoons in
134 question experienced a double peak in RSL, leading to two periods of marine inundation of
135 some lagoons, depending on altitude, during that time.

136 (2) Alternatively, genetic divergence evolved in allopatry, prior to colonisation, and Uist
137 experienced a double-invasion that was brought about by differential arrival times of different
138 lineages/ecotypes (and therefore not necessarily related to RSL change), leading to secondary
139 contact between two already distinct lineages. There is evidence that the west coast of Scotland
140 may be a rare contact zone for ancient mitochondrial stickleback lineages that persisted in
141 glacial refugia on either side of the Atlantic during the Pleistocene (Makinen and Merila 2008),
142 making this a distinct possibility. In this case it is possible that reproductive isolation may
143 already have been well developed upon secondary contact, and species-pairs are directly
144 descendent from two ancient allopatric marine lineages. This hypothesis makes four clear
145 predictions: that Uist was colonised by multiple pre-diverged lineages; there is a differential
146 genetic origin of ecotypes e.g. as in Bernatchez and Dodson (1990); this prior genetic

147 divergence significantly pre-dates the last glacial retreat and there is no evidence for recent
148 admixture between ecotypes.

149 (3) Finally, prior allopatric divergence may have existed, as in (2), but rather than reproductive
150 isolation having reached completion prior to secondary contact, species pairs could have formed
151 as a direct result of admixture between lineages upon secondary contact. This hypothesis
152 predicts that Uist was colonised by multiple, pre-diverged stickleback populations; that there is
153 evidence for recent admixture between those populations and one or both of the species-pair
154 ecotypes is genetically admixed in relation to putative parental populations. To test these three
155 hypotheses we reconstruct the Holocene marine inundation history of a series of coastal lagoons
156 on Uist and use targeted mitochondrial sequencing alongside genome-wide SNP data to
157 investigate the demographic history and genetic relationships of species-pairs inhabiting those
158 lagoons. By consideration of putative mechanisms for these three hypotheses we shed light on
159 the underlying mechanisms of speciation in post-glacial fish.

160 **Results**

161 **Geographical evidence for a ‘double invasion’**

162 Assessing the role of spatial isolation in speciation is notoriously problematic because inferring
163 spatial distributions of historic populations is difficult over long time-scales (Kozak et al. 2008).
164 However, estimates of past lake-sea connectivity can be reconstructed using sediment elemental
165 composition (Ziegler et al. 2008; Chague-Goff et al. 2016) and diatom assemblages (Fritz et al.
166 1991) from sediment deposited on the lakebed. This allows periods of potential colonisation
167 and spatial isolation for coastal aquatic species to be reconstructed and precisely dated, and is
168 a dramatically underutilised resource in speciation research. There are no records of RSL
169 change over the Holocene directly for North Uist, and therefore, we tested our hypothesis that
170 a double-invasion could have been caused by changes in RSL by reconstructing Holocene
171 lagoon-sea connectivity for a series of three lagoons, of varying elevation, containing
172 stickleback species-pairs (Obse, Faik and Strm, which had an elevation of 1.63m, 1.16m and
173 0.92m above datum respectively, see locations highlighted with a thick black border in Figure
174 1d for geographical locations and Table S1 for detailed sample site information).

175 The elemental composition of surface sediments from five locations (Iala, Crei, Bhoi, Dheo and
176 Port), deposited under known salinity conditions, was used as a calibration series to create a

177 discriminant function to predict the salinity under which older sediment was deposited. Surface
178 (upper 20cm) freshwater and marine (absolute water conductivity: <250 $\mu\text{S}/\text{cm}$ and >20,000
179 $\mu\text{S}/\text{cm}$ respectively) sediment deposits clustered distinctly and separately along a single linear
180 discriminant axis (LD1, Figure 2a) and jack-knifed validation indicated that the linear
181 discriminant analysis (LDA) was 100% accurate in classifying known sediment samples into
182 the correct group ($n = 43$). The LDA predicted the salinity of all elemental samples from Obse,
183 Faik and Strm long cores with posterior probabilities of > 0.99. LDA predictions indicated that
184 all three currently brackish lagoons transitioned from freshwater conditions during the time
185 spanned by our sediment sequences (Figure 2b-e), but with no previous evidence of a marine
186 phase. Diatom species counts from Obse validated LDA predictions, indicating an identical
187 freshwater to saline transition as predicted by the LDA, with freshwater diatom species
188 throughout the majority of the core and brackish species appearing only in the top 20cm of
189 sediment (Figure 2b, for a full list of identified diatom species see Table S2).

190 Radiocarbon dating of macrofossil material indicated that the Faik long core spanned the
191 Holocene period with basal sediment approximately 13,097-13,289 cal. YBP (Table S3). The
192 sedimentation rate in Faik was 24.26cm/kyr and almost completely linear (Figure S1).
193 Estimates from transposing the sedimentation rate in Faik suggested that basal sediment in Strm
194 was deposited approximately 14,500 cal. YBP and in Obse 6000-7000 cal. YBP. Based on
195 depth-age relationships, the transition to saline conditions in Obse probably occurred within the
196 last 1000 years with no indication of an older saline period (Figure 2b-c). The earliest indication
197 of saline influx in Faik occurred between 3,294-3,364 cal. YBP and consistently saline
198 conditions were reached by approximately 2,793-3,244 cal. YBP (Figure 2d, Table S3), again
199 with no indication of a second, older saline period. Strm contained the deepest initial saline
200 section, transitioning from freshwater approximately 10,750 cal. YBP (Figure 2e) based on the
201 same age-depth model and sedimentation rates as Faik. The stratified depth and estimated
202 timing of these saline transitions across lagoons is consistent with elevation data from outlet
203 sills, which revealed that Obse was the highest lagoon, followed by Faik, then Strm (1.63m,
204 1.16m and 0.92m above datum respectively).

205 **Divergence in species-pairs**

206 ***Morphological differentiation***

207 Lagoon resident and anadromous stickleback differed consistently in body armour, shape and
208 size (Figure 3) across six species-pair lagoons (Faik, Obse, Duin, Trum Strm and Dheo, see
209 sites marked with red borders in Figure 1d). Morphological differences in species-pairs were
210 largely consistent across sample sites (Figure S3) and therefore results from all sites combined
211 are discussed below. Firstly, anadromous ecotypes had considerably more lateral bony armour
212 plates than lagoon residents ($t = 280.94$, $df = 230.65$, $p < 0.0001$, Figure 3a), with the two
213 ecotypes possessing 31-34 plates and three to seven plates respectively (hereafter referred to as
214 ‘completely-plated’ and ‘low-plated’ morphs). Initial sampling of 239 individuals, euthanized
215 and stained for detailed morphological characterisation, identified only one partially-plated
216 individual (defined, for these purposes, as a fish with between eight and 30 plates), which had
217 25 lateral plates. This partially-plated individual will hereafter be termed ‘intermediate’. We
218 visually inspected a further 1260 individuals for lateral plate morph in the field across the six
219 lagoons and found only 17 partially-plated individuals (1.4%), and the proportion of partially-
220 plated fish did not differ between lakes ($X^2 = 4.26$, $df = 5$, $p = 0.5128$). In the 239 euthanized
221 individuals we measured the sizes of a further six elements of body armour (Figure 1c) and
222 found them all to be largely correlated, with the first principal component axis (‘armour PC1’),
223 explaining 50% of variation in the size of body armour describing an increase in the size of all
224 six armour elements. Armour PC2 (explaining 22% of variation in the size of body armour)
225 described fish with a proportionally larger pelvis and shorter spines. The two ecotypes were
226 largely similar in terms of the relative size of their body armour, although anadromous fish
227 showed slightly increased variability, particularly along armour PC2 (Figure 3b). The
228 intermediate individual fell within the 95% confidence ellipses of both anadromous and lagoon
229 resident ecotypes (Figure 3b).

230 Secondly, anadromous ecotypes scored higher, on average, than lagoon residents along the first
231 principal component axis of a PCA on body shape landmarking coordinates (‘shape PC1’,
232 Figure 3c, See Figure 1a for landmark configurations), which accounted for 29% of total
233 variation in body shape. Anadromous fish exhibited a larger, more pronounced snout, bodies
234 with a deeper anterior and narrower middle section, a thicker caudal peduncle, longer anal fin,
235 more rearward positioning of the pectoral fin and forward positioning of the anal spine (Figure

236 3e). The second principal component (shape PC2) explained 17% of body shape variation but
237 was strongly associated with specimen bending (see Figure S2 for shape changes associated
238 with shape PC2), a common occurrence in stickleback morphometrics (Wund et al. 2008), and
239 so was not considered further in our analysis. Shape PC3 explained 9% of variation in body
240 shape and largely described increasing overall body depth (Figure 3f). Ecotypes were
241 predominantly similar along shape PC3, although anadromous fish were slightly more variable
242 along this axis (Figure 3c).

243 Thirdly, the two ecotypes also differed in overall body size (measured as centroid size: the
244 square root of the sum of the square distances of each landmark to the centre of all landmarks)
245 with anadromous fish being considerably larger than lagoon resident fish ($t = 34.51$, $df = 189.9$,
246 $p = 0.002 \times 10^{-13}$, Figure 3d). The intermediately plated individual was also intermediate in
247 body size (although closer to lagoon resident fish; Figure 3d).

248 ***Genetic differentiation***

249 Genetic analyses from targeted genotyping and genome-wide SNP data strongly suggest that
250 anadromous and lagoon resident ecotypes on Uist are strongly reproductively isolated and
251 maintain genetically distinct genomic regions, despite low levels of gene-flow. First, targeted
252 genotyping at the *Ectodysplasin A (Eda)* locus, which has two key alleles; Eda^L (low) and Eda^C
253 (complete), that generally give rise to low- and completely-plated phenotypes respectively
254 (Colosimo et al. 2005), demonstrated that *Eda* is strongly associated with plate morph in Uist
255 species-pairs from Faik, Obse, Duin, Trum and Dheo ($\chi^2 = 81.2$, $df = 4$, $p = 0.0022 \times 10^{-13}$),
256 accounting for 99% of variation in plate number ($n=55$). Lagoon resident fish ($n=33$) were fixed
257 for the Eda^L allele and anadromous fish were approaching fixation for the Eda^C allele with only
258 one out of the 21 genotyped individuals possessing a heterozygous, rather than homozygous
259 CC genotype (Table 1). The only partially plated, intermediate individual sampled across all
260 species-pair lagoons was also an *Eda* heterozygote (Table 1), suggesting that the low frequency
261 of partial plate morphs in species-pairs reflects a true low frequency of adult F1 hybrids.
262 Second, outlier analyses on the Obse species-pair suggest that multiple regions of the genome
263 are highly differentiated between anadromous and lagoon resident fish, consistent with
264 evolving under divergent selection. Pairwise F_{st} computations identified seven fixed
265 differences (SNPs with $F_{st} = 1$), four of which fell within a 203,000 bp region within a known
266 chromosomal inversion on chromosome I (Jones et al. 2012b) that contains genes including

267 *atp1a1* and *Igfbp2a*, and three within a 22,500 bp region of chromosome IV containing *Eda*
268 and *Vma21* genes, amongst others (Figure 4a). POPULATIONS identified 50 SNPs as outliers
269 under putative selection (SNPs with associated p-values <0.01, Figure 4b). BayeScan identified
270 19 SNPs as ‘decisively’ under selection (posterior probabilities >0.99, Figure 4c) and all 19 of
271 the outlier SNPs identified by BayeScan were also identified by POPULATIONS. Of these 50
272 total outlier SNPs, 33 fell within the coding regions of 23 genes (Table S4), including some that
273 have previously been identified as diverging between marine and freshwater populations
274 (Pellissier et al. 2018). Six of the total outlier SNPs fell within the chromosomal inversion on
275 Chr I, which is known to be involved in both marine-freshwater (Jones et al. 2012b) and lake-
276 stream divergence (Roesti et al. 2015). Third, co-ancestry estimates for the Obse species-pair
277 reveal considerably stronger co-ancestry within than between ecotypes, with some anadromous
278 individuals from Obse sharing more common ancestry with isolated freshwater resident fish
279 from Scad, an inland lake ~10km away (through water), and / or more common ancestry with
280 marine stickleback from Nyps, from the coast of Iceland ~1000km away, than with sympatric
281 saltwater resident individuals in Obse (Figure 5a).

282 *Differential genetic origins of species-pairs*

283 MtDNA sequencing identified 37 composite (*cyt b* + CR) haplotypes across the 76 Uist species-
284 pair individuals (from Faik, Obse, Duin, Trum and Strm) sequenced in this study, 19 of which
285 are, to the best of our knowledge, previously undescribed. *Cyt b* and CR sequences were
286 submitted separately to GenBank under accession numbers MG602878-MG602914 and
287 MG602915-MG602951 respectively. We extended our genetic analysis to include 126
288 individuals from allopatric freshwater populations using sequences taken from Rahn et al.
289 (2016), in an attempt to understand the wider genetic structure and colonisation history of Uist
290 stickleback (see Figure 1d and Table S1 for sampling locations). Bonferroni corrected pairwise
291 ϕ_{ST} permutation values indicated significant genetic differentiation in mitochondrial
292 haplotypes between all three ecotypes (anadromous, lagoon resident and freshwater resident),
293 with anadromous fish being particularly differentiated from the other two ecotypes (Table 2).

294 Bayesian phylogenetic inference on all 202 North Uist mt sequences demonstrated that species-
295 pairs are comprised of two anciently diverged (~119,000 Ybp, although see error margins,
296 Table S5) mitochondrial lineages, which are separated with a posterior probability of 1.00 and
297 correspond to Trans-Atlantic and European lineages identified by Makinen and Merila (2008),

298 Figure S4. We found that the two mitochondrial lineages were present in very different
299 proportions across ecotypes (Chi-squared test: $\chi^2 = 49.97$, $df = 2$, $p < 0.0001$), with the Trans-
300 Atlantic lineage comprising 47% of anadromous, but only 3% and 6% of lagoon resident and
301 freshwater resident populations respectively (Figure 5f). This is consistent with different
302 colonisation histories for anadromous and lagoon resident fish, followed by directional
303 maternal introgression from the lagoon resident/freshwater resident populations into the
304 anadromous population. This strongly implies that, rather than originating from a single
305 homogeneous marine population, Uist was colonised by two divergent maternal lineages: one
306 originating in Europe, and the other from further afield in the Atlantic.

307 *Admixture and introgression*

308 To test the hypothesis that species-pairs may have resulted from recent admixture between older
309 populations or lineages we compared the genetic structure of all four ecotypes that exist on, or
310 have putatively contributed to, stickleback populations on Uist using genome-wide SNP data.
311 This included a species-pair (anadromous and lagoon resident ecotypes) from Obse on Uist, a
312 nearby freshwater resident ecotype from Scad (Uist) and an Atlantic marine population from
313 Nyps in Iceland (as the best available proxy for the second colonising lineage based on the
314 probable colonisation history of Europe (Fang et al. 2018)). These four populations were used
315 in all admixture and introgression analyses below (see Figure 1d and Table S1 for sampling
316 locations). Multiple analyses of the ancestry of the anadromous-lagoon resident species-pair
317 suggest that anadromous stickleback are genetically admixed between lagoon resident and
318 Atlantic marine stickleback (Figure 5). Firstly, co-ancestry estimates reveal that anadromous
319 stickleback are genetically more similar to either lagoon resident or marine populations than
320 lagoon resident and marine populations are to one another (Figure 5a), indicating that either the
321 anadromous ecotype in Uist species-pairs is of admixed origin, or represents a distinct lineage
322 but with substantial gene flow from both lagoon resident and Atlantic marine populations.

323 Secondly, with three inferred clusters, Bayesian estimates of population structure and admixture
324 identified the lagoon resident, marine and freshwater resident populations as unique and
325 genetically distinct, with anadromous fish being admixed and of ~25% lagoon resident and
326 ~75% marine ancestry (Figure 5b). With four inferred clusters (which was the optimal model
327 with a likelihood of -366,677.5), anadromous fish were treated as a distinct population but with
328 introgression from either marine and/or lagoon resident populations in most individuals (Figure

329 5b). Further increases in the number of clusters resulted almost exclusively in the anadromous
330 population gaining genetic input from alternative unsampled populations (Figure S5).

331 Third, principal coordinate analysis (PCoA) of allele frequencies based on all (12,171) SNPs
332 reveal that all four populations form distinct and separate genetic clusters (Figure 5c). The
333 primary axis of genetic differentiation (PCo1, explaining 12.0% of allele frequency variation)
334 separates the freshwater from the three saltwater ecotypes. The second axis of genetic variation
335 (PCo2, explaining 5.7% of allele frequency variation) separates the saltwater ecotypes, with
336 marine and lagoon resident ecotypes at either extreme and anadromous individuals falling
337 intermediately, but closer to marine fish, consistent with genetic admixture in the anadromous
338 population. PCoA including only putatively neutral (12,121) SNPs (Figure S6) was almost
339 identical to the PCoA with all SNPs (Figure 5c). PCoA including only the 50 SNPs under
340 putative selection between lagoon resident and anadromous ecotypes (identified by outlier
341 analyses, Figure 4b and c) lead to a primary axis (PCo1, explaining 79% of allele frequency
342 variation) of differentiation between anadromous and lagoon resident fish (Figure 5d).
343 Freshwater resident fish fell intermediately along PCo1 and marine fish fell on top of
344 anadromous individuals, suggesting marine and anadromous populations experience largely
345 similar selection pressures, while freshwater fish do not experience similar selection pressures
346 to the other ecotypes. PCo2 of the PCoA on selected SNPs separated freshwater from the three
347 saltwater ecotypes, further indicating different selection pressures for freshwater individuals.

348 Fourth, consistent with admixture during the evolution of Uist anadromous fish, inbreeding
349 coefficients suggested that anadromous fish had the highest mean heterozygosity, followed by
350 marine fish, lagoon resident fish and then freshwater resident fish (Figure 5e). Heterozygosity
351 estimates were significantly different between ecotypes (LM: $F_{3,64} = 33.842$, $p < 0.0001$), but
352 post-hoc pairwise comparisons indicated that this was driven by reduced heterozygosity in
353 freshwater resident fish, compared to the other 3 ecotypes (Table 4, Figure 5e).

354 Fifth, jackknifed ABBA BABA tests, with populations arranged to test for introgression from
355 the marine population into either lagoon resident or anadromous populations (based on 11,929
356 SNPs), indicated a significantly positive value of D , both with biallelic sites removed ($D \pm SD$
357 $= 0.1290 \pm 0.0027$, $p < 0.0001$) and with random substitution of biallelic sites ($D \pm SD = 0. \pm$
358 0.0027 , $p < 0.0001$). This is consistent with an introgression event from marine fish into the
359 anadromous population.

360 Sixth, Bayesian coalescent-based estimates of the species-tree indicated a largely consistent
361 consensus tree topology (present in 95.88% of 2000 sampled trees) with anadromous fish
362 diverging recently from lagoon residents (15,000-20,000 YBP, Figure 5g), and freshwater
363 resident fish on the island being more anciently derived (~21,000 YBP). Two alternate tree
364 topologies were identified in the SNAPP runs, indicating low-levels of incomplete lineage
365 sorting, but each was present in just 2.06% of all trees (Figure 5). Furthermore, the consensus
366 tree topology is identified by TreeMix, a software designed to construct the maximum
367 likelihood tree of a set of populations and then infer gene-flow based on residual genetic
368 variation not explained by the tree (Pickrell and Pritchard 2012). TreeMix identifies migration
369 from the marine into the anadromous population in the best fitting demographic model, which
370 further supports mixed ancestry of anadromous fish inferred from the analyses above (Figure
371 5h). TreeMix assumes gene flow occurs via discrete migration events (Pickrell and Pritchard
372 2012), but gene flow may have been long-term and may even be on-going. This is not tested
373 here.

374 **Discussion**

375 We present compelling morphological and genetic evidence that unusually strong reproductive
376 isolation has developed between lagoon resident and anadromous stickleback in numerous
377 coastal lagoons on the Scottish Island of North Uist, in the face of ongoing low level gene-flow.
378 Furthermore, we show that the ‘classic’ purely ecologically-based speciation model with a
379 recent double-invasion from a single genetically homogeneous marine founder population,
380 driven by RSL change, is unlikely to be responsible for the strong reproductive isolation in Uist
381 stickleback species-pairs. Instead, our results suggest that Uist was colonised by at least two,
382 genetically differentiated stickleback lineages, and are consistent with a role for admixture upon
383 secondary contact providing the genetic substrate needed for phenotypic divergence, alongside
384 strong selection in driving and maintaining phenotypic and genetic differentiation.

385 **Evidence for species-pairs**

386 Sympatrically breeding resident and anadromous stickleback ecotypes are common in coastal
387 regions across much of the Holarctic range of the species, and some morphological and genetic
388 divergence is almost ubiquitous (McKinnon and Rundle 2002). Despite this, persistent
389 admixture occurs to some extent across nearly all documented contact zones (Rafinski et al.

1989; McPhail 1994; Higuchi et al. 1996; Hendry et al. 2009) and reproductive isolation as complete as that which exists on Uist is extremely rare (Bell et al. 2010). The occurrence of low-plated resident and completely-plated anadromous morphs in the absence of intermediate partially-plated fish occurs in only two known locations: one in Alaska (Karve et al. 2008; Drevecky et al. 2013) and one in Russia (Ziuganov 1995). We identified six locations on Uist in which morphologically intermediate individuals are extremely rare (~1.4%), and are aware of several others. *Eda* genotyping revealed that *Eda* explains 99% of the variation in plate number in Uist species-pairs, confirming that the rarity of partially-plated individuals likely reflects a true absence of adult F1 hybrids rather than being the result of dominance of the *Eda^C* allele, which has been recorded in a handful of stickleback populations (Cresko et al. 2004; Lucek et al. 2012). Morphological data suggest that species-pairs are largely similar across individual lagoons on North Uist and the close proximity of the lagoons to one another (and their connection by sea) makes it likely that ecotypes across lagoons have a single origin.

Analyses of the species-pair for which SNP data were available (Obse) also implied substantial genetic segregation between sympatric anadromous and lagoon resident ecotypes. All of the methods we used to assess population structure identified the two ecotypes as distinct and separate genetic groups, with very little indication of ongoing admixture. Furthermore, our co-ancestry estimates showed that some anadromous individuals from the species-pair shared more common ancestry with both an isolated inland freshwater population on Uist that is ~10 km away through water and an allopatric marine population ~1000 km away in Iceland, than they did with sympatric lagoon resident fish, strongly suggesting that the two sympatric ecotypes form two largely independently evolving lineages. Outlier analyses also identified numerous regions of the genome that are under putative selection between lagoon resident and anadromous fish. When mapped back to the annotated genome, many of those regions fall within genes that are of known functional and ecological importance (Colosimo et al. 2005; Jones et al. 2012b; Escudero-Esparza et al. 2013), and some fall within a known inversion on Chr I, which is involved in multiple ecological divergence events in stickleback (Jones et al. 2012b; Roesti et al. 2015), suggesting that genomic divergence between ecotypes is likely to have an ecologically adaptive basis. The results of our Structure analysis indicated that there has been some (low-level) recent introgression from the lagoon resident into the anadromous population, suggesting that this phenotypic and genetic differentiation is maintained despite low levels of gene flow. Bayesian coalescence based estimates indicated that lagoon resident

422 and anadromous individuals in Obse diverged from one another approximately 15,000-20,000
423 YBP and our optimal TreeMix model of migration implied that there have been no substantial
424 migration events between these two ecotypes since that initial divergence. Taken together, our
425 findings suggest that reproductive isolation in this species-pair is considerably stronger than in
426 most other stickleback examples, and the two ecotypes are co-existing sympatrically with very
427 little gene flow. The evolutionary processes that lead to the formation of species-pairs such as
428 these are complex and still not well understood (Schluter 2009; Richardson et al. 2014), but
429 Uist stickleback populations provide excellent opportunities to shed light on these.

430 **Origins of species-pairs**

431 *'Classic' purely ecological speciation model*

432 We found little evidence to support the hypothesis that the evolution of species-pairs on Uist is
433 solely explained by the 'classic' double-invasion model of post-glacial speciation in fish. This
434 hypothesis states that two colonisations from the same, genetically homogeneous marine
435 population, made as a results of changes in spatial isolation driven by changes in relative sea-
436 level (RSL), coupled with ecologically based divergent selection in allopatry, are exclusively
437 responsible for the formation of species-pairs. Applied to Uist species-pairs, this theory made
438 four testable predictions. First that there is evidence for strong, ecologically based selection
439 between ecotypes; second, the species-pairs originated from multiple colonisations from a
440 genetically homogeneous 'stock' marine population; third that they are post-glacial in age; and
441 fourth, that there was a double-peak in RSL since the retreat of the Pleistocene glaciers allowing
442 a double-invasion with spatial isolation.

443 Our outlier analysis identified multiple regions of the genome in which strong divergence
444 between lagoon resident and anadromous ecotypes is maintained despite evidence for low-level
445 gene-flow. This suggests these genomic regions are putatively under strong selection between
446 ecotypes. Some of these outlier regions mapped to a chromosomal inversion on Chr I, which is
447 known to be involved in multiple ecological divergence events in stickleback (Jones et al.
448 2012b; Roesti et al. 2015). Other outliers mapped to genes such as *Eda*, which is associated
449 with lateral-plate morph in stickleback (Colosimo et al. 2005), a known anti-predator trait (Bell
450 2001; MacColl and Aucott 2014) and Igf family genes which are known to be under selection
451 between marine and freshwater environments (Pellissier et al. 2018), suggesting the selection

452 in Uist species-pairs has an ecological basis. This is consistent with ‘classic’ ecological
453 speciation. However, we showed that the species-pairs on Uist are comprised of two anciently
454 diverged mitochondrial lineages, revealing that there were likely multiple colonisation events,
455 but they were not made by a single homogeneous lineage of marine fish. While this prior genetic
456 differentiation could be completely unrelated to the evolution of species-pairs, it adds an extra
457 level of variation on which selection can act, and makes it difficult to attribute the evolution of
458 species-pairs entirely to ecological selection for adaptations to the different environments that
459 would have been experienced during a double-invasion.

460 Our Bayesian divergence time estimates indicated that the species-pairs probably are post-
461 glacial in origin, with a divergence time of 15,000-20,000 YBP, which could be consistent with
462 a double-invasion that was driven by changes in RSL. However, this divergence time does not
463 correspond to the timings of any saline influxes in any of the cored species-pair lagoons (which
464 were all considerably more recent), but rather is approximately the same as estimates of the
465 timing of deglaciation of the island (Ballantyne 2010). This is also consistent with speciation
466 being a direct result of secondary contact between older lineages independently colonising Uist
467 immediately following the glacial retreat. Moreover, our relative sea-level change
468 reconstructions detected an increase in relative sea-level to the present day, which is already
469 well documented for this part of the world (Jordan et al. 2010), but did not detect an earlier
470 period of saline flooding in any of the three species-pair lakes for which reconstructions were
471 made. Whilst it is possible that the basal sediment in our cores (~13,000 YBP for the
472 radiocarbon dated core) did not extend to the beginning of the interglacial period on the island
473 and we missed an earlier saline ‘high-stand’ period, or that the speciation event occurred in
474 another lake that was of a different elevation and did experience an earlier saline influx with
475 subsequent migration of the divergent ecotypes into the lakes that we sampled, our alternative
476 explanations are more parsimonious.

477 *Colonisation by pre-diverged lineages*

478 We also hypothesised that Uist may have been colonised by lineages that were already
479 reproductively isolated prior to colonisation and that secondary contact, with or without
480 admixture could be responsible for the evolution of species-pairs on North Uist. These
481 hypotheses predicted that Uist was colonised by multiple pre-diverged lineages, there was a
482 differential genetic origin of ecotypes, the prior genetic divergence existed before the

483 colonisation of Uist (and thus must pre-date the retreat of the Pleistocene glaciers on Uist), and
484 either there was no evidence for recent genetic admixture, or one or both of the species-pair
485 ecotypes is genetically admixed in relation to putative parental populations.

486 Our analyses confirmed that Uist is a meeting place for two predominantly allopatric, ancient
487 mitochondrial lineages: the Trans-Atlantic and European lineages (Makinen and Merila 2008),
488 which diverged ~119,000 YBP, long before the most recent glaciers on Uist would have melted
489 (Ballantyne 2010). We also showed that these lineages occur in very different proportions in
490 different ecotypes, with resident stickleback being almost entirely of European origin, but
491 anadromous fish being an approximately equal mix of the two. There are at least two
492 explanations for this pattern. First, it could suggest that resident and anadromous ecotypes were
493 independently founded by the European and trans-Atlantic lineages respectively and
494 experienced some (mostly unidirectional) introgression upon initial secondary contact,
495 implicating admixture in speciation. In this case introgression must have been almost
496 ubiquitously between lagoon resident females (with Eu mtDNA) and anadromous males, a
497 pairing which seems much more likely than anadromous females (with TA mtDNA) mating
498 with lagoon resident males, given that lagoon resident fish are considerably smaller than
499 anadromous fish. Alternatively, Uist may have initially been colonised by the European lineage,
500 which gave rise to both resident and anadromous populations, and fish of trans-Atlantic origin
501 arrived later, and failed to introgress into resident populations. In the latter case it is possible
502 that barriers to gene flow between the two ecotypes already existed, preventing trans-Atlantic
503 mitochondrial haplotypes from entering resident populations.

504 For a number of reasons, however, it is more likely that the primary reason for the lack of trans-
505 Atlantic haplotypes in resident populations is that anadromous fish carrying trans-Atlantic
506 mitochondrial haplotypes are less well adapted to a resident lifestyle. Firstly, Anadromous fish
507 of trans-Atlantic origin in Europe are descendant from fish which must have crossed the
508 Atlantic, and therefore probably possess a greater suite of adaptations to an oceanic rather than
509 a resident existence. Fish in Europe with European mitochondrial haplotypes, on the other hand,
510 have probably spent much more of their recent evolutionary history as resident populations,
511 making them likely to be better adapted to a resident lifestyle. Second, there are locations on
512 Uist in which freshwater resident and anadromous stickleback do hybridise (MacColl *et al.*,
513 unpublished data), which should result in trans-Atlantic mitochondrial haplotypes infiltrating

514 freshwater resident populations, and yet this largely appears not to be the case. Interestingly,
515 our outlier analysis identified six SNPs (by far the most within any one gene) within the
516 vacuolar H⁺-ATPase (*Vma21*) gene as being under strong selection between lagoon resident
517 and anadromous fish. The function of *Vma21* is likely related to ATP synthesis (Finbow and
518 Harrison 1997), a pathway which also involves many mitochondrially encoded proteins, and
519 thus perhaps mitonuclear conflict/incompatibilities have played a part in the evolution of
520 species-pairs. Further investigations would be necessary to draw conclusions about this exciting
521 possibility. Regardless of the mechanism, genetic differences between lineages could thus have
522 been involved in the formation of species-pairs.

523 *Admixture*

524 MtDNA sequencing revealed that the anadromous population on Uist is comprised of two
525 ancient maternal lineages, which occur in approximately equal proportions, suggesting that
526 admixture has been particularly important in the evolution of the anadromous ecotype on Uist.
527 MtDNA, however, only relays information about the maternal line, and thus we also compared
528 the autosomal DNA of a species-pair with that of local freshwater resident and Icelandic marine
529 (as a proxy for an Atlantic marine founder population) stickleback to investigate the genetic
530 relationships between ecotypes. Our analyses of the autosomal genome-wide SNP set indicated
531 that the anadromous population on Uist is genetically admixed, with genomic input from lagoon
532 resident (the other half of the species-pair) and Atlantic marine populations. Genetic admixture
533 can provide novel combinations of genes on which selection can act, and is most likely to be
534 involved in speciation when recombinant phenotypes are better adapted to a given niche than
535 either parental species, allowing admixed individuals to exploit environments that are
536 unavailable to either parent species (Schumer et al. 2014). We therefore hypothesise that an
537 initial marine colonisation event may have given rise to freshwater and lagoon resident ecotypes
538 on Uist as in many other parts of the world (McKinnon and Rundle 2002). Then the proximity
539 of Uist to the Atlantic lead to subsequent admixture between these derived ecotypes and the
540 fully marine stickleback population in the Atlantic, about which very little is known (Ahnelt
541 2018). Admixture would likely have produced some individuals with a combination of
542 adaptations to an oceanic lifestyle, but also a propensity to spawn in the safety of coastal
543 regions, a combination not found in either parental population, that could allow them to
544 simultaneously exploit both environments as anadromous fish. The habit of migrating to sea in

545 anadromous fish could be enough by itself to cause strong disruptive selection from lagoon
546 resident fish. Hybrids with a tendency to migrate, but without the full genetic physiological or
547 antipredator ‘toolkit’ to live in the sea, would fall in a valley of very low fitness and be unlikely
548 to reach adulthood.

549 In a phylogeny of Uist stickleback populations (with Icelandic marine fish as the outgroup),
550 Uist anadromous fish do not approximate the marine founders of the island, as would be
551 expected by traditional models of stickleback dispersal (Colosimo et al. 2005; Schluter and
552 Conte 2009). Rather, freshwater resident stickleback fall as the outgroup to other Uist
553 populations, and the anadromous population evolved more recently from lagoon resident fish.
554 By modelling historic migration events we were able to show that anadromous fish received
555 genetic input from the Icelandic marine population (our proxy for Atlantic marine stickleback)
556 during their divergence from lagoon resident fish. Whether or not this could be defined as
557 ‘hybrid speciation’ depends on how the term is defined (Schumer et al. 2018), but our findings
558 add to the growing body of evidence suggesting that speciation is not a linear, bifurcating
559 process, but is in fact far more reticulate than was once widely thought (Martin-Bravo et al.
560 2010; Frantz et al. 2013; Alexander et al. 2015), with admixture events often playing a key role
561 in the process (Mallet 2007; Comeault and Matute 2018; Marques et al. 2019).

562 **Conclusions**

563 We have identified unusually strong reproductive isolation between sympatric anadromous and
564 lagoon resident stickleback ecotypes, in multiple lagoons on the Scottish Hebridean island of
565 North Uist. We tested three hypotheses, which were not mutually exclusive, to explain how
566 such strong reproductive isolation has evolved. While we cannot completely rule out
567 contributions from any of our three models, our results indicate that the ‘classic’ explanation
568 for more pronounced reproductive isolation in stickleback, an ecological speciation model
569 driven by post-glacial changes in RSL, is unlikely to be responsible for speciation in the present
570 case. Instead, our results suggest that the most parsimonious explanation probably involves
571 genetic admixture upon secondary contact between multiple colonising lineages / ecotypes that
572 provided the basis for strong (at least partially ecologically based) divergent selection. These
573 findings are in line with much recent research that is beginning to suggest that cases of
574 seemingly recent, purely ecologically-based speciation are actually cryptic examples of
575 speciation that has a much older genetic basis, that developed allopatrically (Bernatchez and

576 Dodson 1990; Feder et al. 2003; Kuehne et al. 2007; Foote and Morin 2015; Foote 2018;
577 Marques et al. 2019). This study demonstrates that proper inter-disciplinary investigations of
578 localised geographical changes should be made before those changes can be assumed to have
579 driven speciation via changes in habitat connectivity and population range shifts, particularly
580 with regards to geographical events that can be highly variable across small spatial scales, such
581 as relative sea-level change. Our study also highlights how genetic data can be used to test
582 historic demographic hypotheses and demonstrates that an inter-disciplinary approach,
583 combining genetic, morphological and geographical data is likely to give the most complete
584 picture of historic speciation events. Finally, we have identified a new system which provides
585 an exciting future opportunity to investigate parallelism across species-pairs in multiple
586 lagoons.

587 **Materials and Methods**

588 **Sampling design**

589 We collected sediment core sequences for relative sea-level (RSL) change reconstructions from
590 species-pairs lagoons during two field trips to Uist in 2013 and 2015 (Section 1, Supplementary
591 information). We collected stickleback from species-pair lagoons during spring of 2015 for
592 morphological and genetic analyses (Section 2, Supplementary information). We also obtained
593 an additional RAD-seq SNP dataset including 70 individuals from 3 populations on Uist and
594 one in Iceland from Magalhaes et al. (2016) and mitochondrial sequences from a further 126
595 Uist stickleback from Rahn et al. (2016) to extend our understanding of where Uist species-
596 pairs fit within the wider radiation of stickleback, and to identify their origins.

597 **Relative sea-level (RSL) change reconstructions**

598 Briefly, to reconstruct changes in RSL on Uist over the Holocene period we precisely mapped
599 the elevation of three species-pair lagoons, collected long sediment sequences from them, and
600 predicted past changes in salinity using a discriminant function trained with modern Uist
601 sediment samples. We then radiocarbon dated ancient sediment samples to date marine –
602 freshwater / freshwater – marine transitions. For details, see Section 1 of the Supplementary
603 material.

604 **Morphological analyses**

605 To quantify morphological differentiation, stickleback were sampled from six species-pair
606 lakes on Uist: Faik, Obse, Duin, Trun, Strm and Dheo (see Figure 1d for lake locations and
607 Table S1 for detailed sampling information). We measured morphological differences in
608 species-pairs by quantifying differences in three key aspects of morphology: body size, body
609 shape and external body armour. Briefly, individuals were stained to highlight external skeletal
610 structures before measurements of various aspects of body armour were taken. We then used a
611 geometric morphometric landmarking approach to measure differences in body shape and body
612 size. For details, see Section 3 of the Supplementary material.

613 **Genetic analyses**

614 To ascertain whether morphological differences in lateral plate morph reflected underlying
615 genetic segregation in species-pairs we genotyped a subset of individuals (from Dheo, Duin,
616 Faik, Obse and Trun, see Figure 1d for lake locations and Table S1 for detailed sampling
617 information) at the *Eda* locus, which is involved in determining lateral plate phenotype
618 (Colosimo et al. 2005), and made genotype – phenotype comparisons (see Section 4a of the
619 Supplementary information for details). We constructed various SNP datasets from those
620 published in Magalhaes et al. (2016) and SNPs from an Icelandic population (Nyps), which was
621 sequenced and processed at the same time. Our SNP datasets were constructed to include
622 different individuals, populations and filtering for different analyses (for a detailed description
623 of SNP datasets see Table 3). To identify regions of the genome under putative selection in a
624 species-pair (Obse) we used POPULATIONS in the Stacks pipeline (Catchen et al. 2013) and
625 BayeScan version 2.1 (Foll and Gaggiotti 2008), which were run on SNP dataset 1 (see Table
626 3 for details). For further details about SNP analyses see section 4b, Supplementary material.

627 To investigate whether North Uist is a meeting place for multiple ancient mitochondrial
628 lineages, we sequenced 76 species-pair individuals collected from five lagoons (Faik, Obse,
629 Duin, Trun and Strm, See Table S1 for detailed sampling information) for two mitochondrial
630 regions: the cytochrome *b* (*cyt b*) gene and a partial fragment of the D-loop control region (CR),
631 which are known to resolve ancient mitochondrial lineages present in the Atlantic and the seas
632 around Europe (Makinen and Merila 2008). We obtained a further 126 concatenated sequences
633 from Uist stickleback from Rahn et al. (2016) and aligned them with our own, resulting in a

634 final 1380bp alignment of 202 Uist individuals. To determine whether the genetic structure in
635 Uist mtDNA sequences corresponded to the ancient mitochondrial lineages identified by
636 Makinen and Merila (2008), we collapsed our individual sequence data into haplotypes and
637 constructed a Bayesian phylogeny including the haplotype sequences published in Makinen and
638 Merila (2008), downloaded from Genbank, using MrBayes version 3.2.2 (Ronquist and
639 Huelsenbeck 2003). We then estimated divergence times between the two lineages that were
640 identified in Uist stickleback using coalescence based MCMC simulations implemented in
641 IMA2 (Hey and Nielsen 2004) to ensure that our divergence times were approximately similar
642 to those in Makinen and Merila (2008). For further details of all mitochondrial analyses see
643 section 4c, Supplementary material.

644 To test the hypothesis that admixture may have been important in the evolution of species-pairs
645 we conducted a variety of analyses using the genome-wide SNP data (Table 3). We attempted
646 to compare all current ecotypes present on Uist alongside those which probably resemble the
647 islands marine colonisers. To that end, we compared lagoon resident and anadromous
648 individuals from a species-pair (Obse) with those from a nearby (~10km through water), but
649 isolated freshwater population on Uist (Scad) and individuals from a marine population in
650 Iceland (Nyps, ~1000km away). The Icelandic marine population was used as the best available
651 proxy for oceanic stickleback in the Atlantic since marine stickleback in Iceland likely
652 approximate the ancestral Atlantic colonisers of Uist (Fang et al. 2018). First, to investigate the
653 extent of shared co-ancestry between populations we constructed a co-ancestry matrix using
654 dataset 2 (see Table 3 for details) in fineRADstructure version 0.3.1 (Malinsky et al. 2018).
655 Second, we used dataset 3 (Table 3) to estimate genetic structure and the optimal number of
656 genetic clusters across populations (models with one to six clusters were tested) in Structure
657 version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003). Third, we conducted a principal
658 coordinate analysis (PCoA) using the *adegenet* (Jombart 2008) package in R version 3.4.4
659 (R.Core.Team 2017) on dataset 4 (including all SNPs), dataset 4a (including only SNPs
660 determined to be under selection in outlier analyses above) and dataset 4b (including only SNPs
661 determined to be evolving neutrally in outlier analyses above, Table 3) to assess the relative
662 positions of populations in multidimensional genetic space, both overall, and in terms of shared
663 or different selection pressures. Fourth, to investigate heterozygosity we used dataset 4 to
664 estimate inbreeding coefficients (F) on a per-individual basis in VCFtools version 0.1.16
665 (Danecek et al. 2011). Fifth, to identify introgression from Nyps into the species-pair, we used

666 ABBA BABA tests on dataset 2 to estimate jackknifed D statistics, both with biallelic sites
667 removed and with random substitutions of biallelic sites using custom R scripts (see section 4b,
668 supplementary material for access to R scripts). Sixth, to investigate the colonisation history of
669 Uist we estimated multi-locus phylogenetic trees and population divergence times using
670 datasets 5a and 5b in SNAPP analyses (Bryant et al. 2012), implemented in Beast version 2.5.1
671 (Drummond and Rambaut 2007). Finally, we constructed a maximum likelihood tree for the
672 same populations and modelled historic migration events using dataset 5 in TreeMix version
673 1.13 (Pickrell and Pritchard 2012). For further details of all SNP analyses see section 4b of the
674 supplementary material.

675 **Acknowledgements**

676 Thanks to all members of the MacColl lab for invaluable advice and discussions during all
677 aspects of this project. Special thanks to Paul Hohenlohe for providing expertise for RAD
678 sequencing, Nick Wallerstein, Mark Stevenson, Xu Chen, Talib Chitheer, Abdul Rahman and
679 Ann Lowe for assistance in collecting sediment cores, to Teresa Needham for assistance in
680 processing ED-XRF samples, and to Gareth Lee for gathering and analysing digital terrain
681 modelling data to identify rock sill heights. This work was supported by the Natural
682 Environment Research Council (grant numbers NE/L002604/1 to L.L.D., NE/J02239X/1 to
683 A.D.C.M.) and the Natural Environment Research Council Radiocarbon Facility NRCF010001
684 (allocation number: 1942.1015). Cyt *b* and CR mitochondrial sequences were submitted
685 separately to gene bank under accession numbers MG602878-MG602914 and MG602915-
686 MG602951 respectively. Publication codes for radiocarbon dates are given in Table S1.

687 **Figure Legends**

688 Figure 1

689 Examples of lagoon resident (a), intermediate (b) and anadromous (c) phenotypes alongside a
690 map showing sampling locations (d). Images show stickleback that have been stained with
691 Alizarin red to highlight external skeletal structures. (a) Shows the positions of 27 landmarks
692 used in the geometric morphometric analysis of body shape, (b) shows how lateral plate counts
693 and measurements of standard length were taken and (c) how measurements of all continuous
694 body armour variables were taken. (d) Shows the locations of all sample sites from which data
695 were used, with marine (absolute conductivity $>35,000 \mu\text{S}/\text{cm}$) locations labelled in dark blue,
696 brackish (absolute conductivity $20,000\text{-}35,000 \mu\text{S}/\text{cm}$) locations labelled in mid-blue and
697 freshwater (absolute conductivity $<500 \mu\text{S}/\text{cm}$) labelled light blue. A thick black border
698 represents sites for which marine inundation history was reconstructed and a thick red border
699 indicate sites containing species-pairs examined in this manuscript (see Table S1 for a detailed
700 description of sampling sites and which analyses each site was used for).

701 Figure 2

702 Salinity reconstructions for currently brackish North Uist lagoons. (a) Separation of freshwater
703 (red bars, $<250 \mu\text{S}/\text{cm}$) and saline (blue bars, $>20,000 \mu\text{S}/\text{cm}$) waterbodies along linear
704 discriminant one (LD1) of a linear discriminant analyses (LDA) based on sediment elemental
705 composition used to classify long core sediment in (b). (b) Models of past salinity for Obse,
706 Faik and Strm (elevations: 1.63m, 1.16m and 0.92m above datum respectively) based on the
707 percentage of brackish diatom species (% Br diatoms) and the predictions of a linear
708 discriminant analysis of lake sediment elemental composition (LDA salinity predictions). For
709 the LDA, red circles and 'F' correspond to 'freshwater' conditions and blue circles and 'S' to
710 'saline' conditions. For diatom count blue circles indicate $> 5\%$ brackish diatom species and
711 red circles $< 5\%$. For salinity reconstructions in (b) age is estimated for Faik using a Bayesian
712 age-depth model based on six radiocarbon dates (Table S3), implemented using the R package
713 Bacon, and for all other cores by transposing the linear mean sedimentation rate in Faik (Figure
714 S1).

715 Figure 3

716 Morphological variation in North Uist stickleback species-pairs. (a) histogram showing lateral
717 plate counts. (b) Distribution of phenotypes and their associated 95% confidence ellipses in a
718 principal components analysis (PCA) of all size standardised, continuous body armour
719 variables (see Figure 1c for continuous armour measurements). Principal component 1 (Armour
720 PC1, explaining 50% of variation in the data) described an increase in the size of all armour
721 variables and Armour PC2 (explaining 22% of variation in the data) described a relative
722 increase in the size of the pelvis and decrease in spine length. (c) Distribution of phenotypes
723 and their associated 95% confidence ellipses in a PCA of 56 body shape variables (derived from
724 27 landmarks, see Figure 1a for landmark positions). Shape PC1 (explaining 29% of variation
725 in the data) described an increase in posterior body depth, mouth size, and a more rearward

726 positioning of the pectoral fin. Shape PC3 (explaining 9% of variation in the data) largely
727 described an increase in anterior body depth and shortening of the caudal peduncle. Shape
728 changes for Shape PC1 and 3 are shown in warped outline drawings (e) and (f) respectively,
729 with 1.5% scaling. (d) Box-plots showing centroid size, with error bars representing the
730 standard error of the mean (*SEM*). (a) – (f) are based on analyses of 239 individuals from 6
731 lakes containing species-pairs (Figure 1d, Table S1).

732 Figure 4

733 Analyses of outlier SNPs in the Obse species-pair. Manhattan plots showing (a) genome-wide
734 *Fst* estimates for 12,575 SNPs calculated using the POPULATIONS program in the Stacks
735 pipeline, and (b) the negative logarithm at base 10 of the *p*-values ($-\log_{10}(P)$) for the SNPs in
736 (a). Horizontal lines in (b) represent the 0.05 (dashed line) and 0.01 (solid line) significance
737 thresholds for SNPs under selection. (a) and (b) show the location of SNPs across the genome,
738 excluding the sex chromosomes. U describes SNPs that mapped to unassigned scaffolds. (c) *Fst*
739 and *q*-values (\log_{10}) for 12,575 SNPs, estimated using the Bayescan software. Vertical lines
740 mark ‘strong’ (dotted line), ‘very strong’ (dashed line) and ‘decisive’ (solid line) boundaries on
741 Jeffreys’ scale of interpretation, corresponding to posterior probabilities of loci being under
742 selection of 0.91, 0.97 and 0.99 respectively. SNPs identified as decisively under selection in
743 the BayeScan analysis are indicated by red triangles in (a) and (b). Outlier analyses were based
744 on 34 individuals from one species-pair lake (Obse, see Figure 1d for location and Table S1 for
745 lake details).

746 Figure 5

747 Genomic analyses of a North Uist lagoon resident-anadromous species-pair (Obse) alongside a
748 Uist freshwater resident (Scad) and Icelandic marine population (Nyps) based on genome-wide
749 SNP data. (a) Co-ancestry matrix constructed in fineRADstructure. (b) Population structure
750 plots with two (*K*=2), three (*K*=3) and four (*K*=4) inferred clusters output from Structure.
751 Principal coordinate analysis (PCoA) of allele frequencies computed using adegenet for (c) all
752 SNPs (d) SNPs under putative selection and (e) putatively neutral SNPs. (f) Proportion of
753 ancient European (Eu) and Trans-Atlantic (TA) mitochondrial lineages across anadromous
754 (Anad), lagoon resident (LR) and freshwater resident (FR) stickleback from North Uist. (g) A
755 tree cloud produced using Densitree to visualise the range of alternate topologies of a Bayesian
756 phylogeny produced from a SNAPP analysis in Beast. Divergence time estimates are shown in
757 kya and were calculated using the Icelandic marine population as an outgroup, with an
758 estimated divergence time of 21,100 YBP taken from Fang et al. (2018). Trees shown in black
759 made up 95.8% of consensus tree topologies and trees shown in red and blue made up 2.06%
760 of consensus tree topologies each. Red and blue tree topologies in (g) are intensified two-fold.
761 (h) Maximum likelihood tree estimated in TreeMix, with the arrow representing a single
762 migration event that was identified by the optimal TreeMix demographic model. Lag res:
763 lagoon resident, Anad: anadromous, Fw res: freshwater resident.

764 **Tables**

765 **Table 1**

766 **Plate morph phenotype vs. genotype.** Lateral plate morph and *Eda* genotype of 55 genotyped
 767 individuals from species-pair lagoons on North Uist. CC indicates two copies of the *Eda*^C allele,
 768 LL, two copies of the *Eda*^L allele and CL, one copy of *Eda*^C and one of *Eda*^L.

Genotype	Plate morph		
	Low	Partial	Complete
CC	0	0	20
CL	0	1	1
LL	33	0	0

769

770

771

772 **Table 2**

773 **Pairwise ϕ_{ST} and associated *p*-values for composite Cyt *b* + CR mtDNA haplotypes.** *P*-
 774 values shown are Bonferroni corrected for multiple comparisons. Anad: anadromous, lag res:
 775 lagoon resident, fw res: freshwater resident.

Comparison	ϕ_{ST}	<i>p</i> -value
anad vs. fw res	0.13621	0.00300
anad vs. lag res	0.15065	0.00300
fw res vs. lag res	0.02682	0.02398

776

777 **Table 3**

778 **SNP datasets used in population genetic analyses.** N ind: number of individuals, N SNPs:
 779 number of SNPs, n: sample size, LD thinned: whether or not SNPs were thinned to 2000+bp
 780 apart to account for linkage disequilibrium, Sibs removed: whether or not siblings were
 781 removed, neutral SNPs: whether or not neutral SNPs were included, selected SNPs: whether or
 782 not selected SNPs were included.

Dataset	N ind	N SNPs	Lake / ecotype / n	LD thinned	Sibs removed	neutral SNPs	selected SNPs
dataset 1	34	12,575	Obse: anad (16) Obse: lag res (18)	✗	✗	✓	✓
dataset 2	68	11,930	Obse: anad (14) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✗	✓	✓	✓
dataset 3	68	9464	Obse: anad (14) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✓	✓	✓	✓
dataset 4	70	12,171	Obse: anad (16) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✗	✗	✓	✓
dataset 4a	70	50	Obse: anad (16) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✗	✗	✗	✓
dataset 4b	70	12,121	Obse: anad (16) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✗	✗	✓	✗
dataset 5	70	9,464	Obse: anad (16) Obse: lag res (18) Nyps: marine (19) Scad: fw res (17)	✓	✗	✓	✓
dataset 5a	23	1000	Obse: anad (3) Obse: lag res (9) Nyps: marine (4) Scad: fw res (7)	✓	✗	✓	✓
dataset 5b	23	1000	Obse: anad (4) Obse: lag res (10) Nyps: marine (5) Scad: fw res (4)	✓	✗	✓	✓

784 **Table 4**

785 **Post-hoc pairwise *t*-tests for population differences in heterozygosity.** *T* - statistics are given
786 above the diagonal and *p*-values (adjusted for multiple testing using the fdr method) below the
787 diagonal.

	Marine	Lag res	Anad	Fw res
Marine	-	0.13	0.00	8.29
Lag res	0.90	-	0.00	8.06
Anad	0.65	0.65	-	8.27
Fw res	<0.0001	<0.0001	<0.0001	-

788

789 **References**

- 790 Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman JW, Brelsford A,
791 Buerkle CA, Buggs R et al. 2013. Hybridization and speciation. *Journal of Evolutionary*
792 *Biology* 26:229-246.
- 793 Ahnelt H. 2018. Imprecise naming: the anadromous and the sea spawning threespine
794 stickleback should be discriminated by names. *Biologia* 73:389-392.
- 795 Alexander PJ, Windham MD, Beck JB, Al-Shehbaz IA, Allphin L, Bailey CD. 2015. Weaving
796 a Tangled Web: Divergent and Reticulate Speciation in *Boechera fendleri* sensu lato
797 (Brassicaceae: Boechereae). *Systematic Botany* 40:572-596.
- 798 Ballantyne CK. 2010. Extent and deglacial chronology of the last British-Irish ice sheet:
799 implications of exposure dating using cosmogenic isotopes. *Journal of Quaternary Science*
800 25:515-534.
- 801 Bell MA. 2001. Lateral plate evolution in the threespine stickleback: getting nowhere fast.
802 *Genetica* 112:445-461.
- 803 Bell MA, Aguirre WE, Buck NJ. 2004. Twelve years of contemporary armor evolution in a
804 threespine stickleback population. *Evolution* 58:814-824.
- 805 Bell MA, Gangavalli AK, Bewick A, Aguirre WE. 2010. Frequency of Ectodysplasin alleles
806 and limited introgression between sympatric threespine stickleback populations. *Environmental*
807 *Biology of Fishes* 89:189-198.
- 808 Bernatchez L, Dodson JJ. 1990. Allopatric origin of sympatric populations of lake whitefish
809 (*Coregonus clupeformis*) as revealed by mitochondrial-DNA restriction analysis. *Evolution*
810 44:1263-1271.
- 811 Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring
812 Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full
813 Coalescent Analysis. *Molecular Biology and Evolution* 29:1917-1932.
- 814 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool
815 set for population genomics. *Molecular Ecology* 22:3124-3140.
- 816 Chague-Goff C, Chan JCH, Goff J, Gadd P. 2016. Late Holocene record of environmental
817 changes, cyclones and tsunamis in a coastal lake, Mangaia, Cook Islands. *Isl. Arc.* 25:333-349.
- 818 Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz
819 J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks
820 by repeated fixation of ectodysplasin alleles. *Science* 307:1928-1933.
- 821 Comeault AA, Matute DR. 2018. Genetic divergence and the number of hybridizing species
822 affect the path to homoploid hybrid speciation. *Proceedings of the National Academy of*
823 *Sciences of the United States of America* 115:9761-9766.

- 824 Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB,
825 Postlethwait JH. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan
826 threespine stickleback populations. *Proceedings of the National Academy of Sciences of the*
827 *United States of America* 101:6050-6055.
- 828 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter
829 G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics*
830 27:2156-2158.
- 831 Drevecky CJ, Falco R, Aguirre WE. 2013. Genetic divergence of a sympatric lake-resident-
832 anadromous three-spined stickleback *Gasterosteus aculeatus* species pair. *Journal of Fish*
833 *Biology* 83:111-132.
- 834 Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.
835 *Bmc Evolutionary Biology* 7:8.
- 836 Escudero-Esparza A, Kalchishkova N, Kurbasic E, Jiang WG, Blom AM. 2013. The novel
837 complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes
838 factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex
839 assembly. *Faseb J.* 27:5083-5093.
- 840 Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus
841 genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.
- 842 Fang BH, Merila J, Ribeiro F, Alexandre CM, Momigliano P. 2018. Worldwide phylogeny of
843 three-spined sticklebacks. *Molecular Phylogenetics and Evolution* 127:613-625.
- 844 Feder JL, Berlocher SH, Roethele JB, Dambroski H, Smith JJ, Perry WL, Gavrilovic V, Filchak
845 KE, Rull J, Aluja M. 2003. Allopatric genetic origins for sympatric host-plant shifts and race
846 formation in *Rhagoletis*. *Proceedings of the National Academy of Sciences of the United States*
847 *of America* 100:10314-10319.
- 848 Feder JL, Flaxman SM, Egan SP, Nosil P. 2013. Hybridization and the build-up of genomic
849 divergence during speciation. *Journal of Evolutionary Biology* 26:261-266.
- 850 Feder JL, Xie XF, Rull J, Velez S, Forbes A, Leung B, Dambroski H, Filchak KE, Aluja M.
851 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*.
852 *Proceedings of the National Academy of Sciences of the United States of America* 102:6573-
853 6580.
- 854 Ferguson A, Taggart JB. 1991. Genetic differentiation among the sympatric brown trout (*Salmo*
855 *trutta*) populations of Lough Melvin, Ireland. *Biological Journal of the Linnean Society* 43:221-
856 237.
- 857 Finbow ME, Harrison MA. 1997. The vacuolar H⁺-ATPase: A universal proton pump of
858 eukaryotes. *Biochem. J.* 324:697-712.
- 859 Foll M, Gaggiotti O. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for
860 Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180:977-993.

861 Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall
862 AB, Catteruccia F, Kakani E et al. 2015. Extensive introgression in a malaria vector species
863 complex revealed by phylogenomics. *Science* 347.

864 Foote AD. 2018. Sympatric speciation in the genomic era. *Trends in Ecology & Evolution*
865 33:85-95.

866 Foote AD, Morin PA. 2015. Sympatric speciation in killer whales? *Heredity* 114:537-538.

867 Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, Semiadi G, Meijaard
868 E, Li N, Crooijmans RPMA et al. 2013. Genome sequencing reveals fine scale diversification
869 and reticulation history during speciation in *Sus*. *Genome Biology* 14.

870 Friele PA, Hutchinson I. 1993. Holocene sea-level change in the central west-coast of
871 Vancouver-Island, British-Columbia. *Canadian Journal of Earth Sciences* 30:832-840.

872 Fritz SC, Juggins S, Battarbee RW, Engstrom DR. 1991. Reconstruction of past changes in
873 salinity and climate using a diatom-based transfer-function. *Nature* 352:706-708.

874 Gow JL, Peichel CL, Taylor EB. 2006. Contrasting hybridization rates between sympatric
875 three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily
876 young species. *Molecular Ecology* 15:739-752.

877 Gow JL, Rogers SM, Jackson M, Schluter D. 2008. Ecological predictions lead to the discovery
878 of a benthic-limnetic sympatric species pair of threespine stickleback in Little Quarry Lake,
879 British Columbia. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 86:564-571.

880 Grant PR, Grant BR. 2009. The secondary contact phase of allopatric speciation in Darwin's
881 finches. *Proceedings of the National Academy of Sciences of the United States of America*
882 106:20141-20148.

883 Grant V. 1971. Plant Speciation. *Plant Speciation*:435.

884 Hagen DW. 1967. Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *Journal of*
885 *the Fisheries Board of Canada* 24:1637-1692.

886 Hay DE, McPhail JD. 2000. Courtship behaviour of male threespine sticklebacks (*Gasterosteus*
887 *aculeatus*) from old and new hybrid zones. *Behaviour* 137:1047-1063.

888 Hendry AP. 2009. Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries*
889 *and Aquatic Sciences* 66:1383-1398.

890 Hendry AP, Bolnick DI, Berner D, Peichel CL. 2009. Along the speciation continuum in
891 sticklebacks. *Journal of Fish Biology* 75:2000-2036.

892 Heuts MJ. 1947. Experimental studies on adaptive evolution in *Gasterosteus aculeatus* L.
893 *Evolution* 1:89-102.

- 894 Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates
895 and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D-*
896 *persimilis*. *Genetics* 167:747-760.
- 897 Higuchi M, Goto A. 1996. Genetic evidence supporting the existence of two distinct species in
898 the genus *Gasterosteus* around Japan. *Environmental Biology of Fishes* 47:1-16.
- 899 Higuchi M, Goto A, Yamazaki F. 1996. Genetic structure of threespine stickleback,
900 *Gasterosteus aculeatus*, in Lake Harutori, Japan, with reference to coexisting anadromous and
901 freshwater forms. *Ichthyological Research* 43:349-358.
- 902 Hutchinson I, James T, Clague J, Barrie JV, Conway K. 2004. Reconstruction of late
903 Quaternary sea-level change in southwestern British Columbia from sediments in isolation
904 basins. *Boreas* 33:183-194.
- 905 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
906 *Bioinformatics* 24:1403-1405.
- 907 Jones FC, Brown C, Pemberton JM, Braithwaite VA. 2006. Reproductive isolation in a
908 threespine stickleback hybrid zone. *Journal of Evolutionary Biology* 19:1531-1544.
- 909 Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Absher DM, Myers
910 RM, Reimchen TE, Deagle BE et al. 2012a. A genome-wide SNP genotyping array reveals
911 patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* 22:83-
912 90.
- 913 Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M,
914 Zody MC, White S et al. 2012b. The genomic basis of adaptive evolution in threespine
915 sticklebacks. *Nature* 484:55-61.
- 916 Jordan JT, Smith DE, Dawson S, Dawson AG. 2010. Holocene relative sea-level changes in
917 Harris, Outer Hebrides, Scotland, UK. *Journal of Quaternary Science* 25:115-134.
- 918 Josenhans H, Fedje D, Pienitz R, Southon J. 1997. Early humans and rapidly changing
919 Holocene sea levels in the Queen Charlotte Islands Hecate Strait, British Columbia, Canada.
920 *Science* 277:71-74.
- 921 Karve AD, von Hippel FA, Bell MA. 2008. Isolation between sympatric anadromous and
922 resident threespine stickleback species in Mud Lake, Alaska. *Environmental Biology of Fishes*
923 81:287-296.
- 924 Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, Wittwer S, Seehausen O.
925 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation
926 in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* 22:2848-2863.
- 927 Kitano J, Mori S, Peichel CL. 2007. Phenotypic divergence and reproductive isolation between
928 sympatric forms of Japanese threespine sticklebacks. *Biological Journal of the Linnean Society*
929 91:671-685.

- 930 Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz
931 J, Myers RM et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature*
932 461:1079-1083.
- 933 Kocher TD. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature*
934 *Reviews Genetics* 5:288-298.
- 935 Kozak KH, Graham CH, Wiens JJ. 2008. Integrating GIS-based environmental data into
936 evolutionary biology. *Trends in Ecology & Evolution* 23:141-148.
- 937 Kuehne HA, Murphy HA, Francis CA, Sniegowski PD. 2007. Allopatric divergence, secondary
938 contact and genetic isolation in wild yeast populations. *Current Biology* 17:407-411.
- 939 Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid
940 speciation in Darwin's finches. *Science* 359:224-227.
- 941 Lucek K, Haesler MP, Sivasundar A. 2012. When phenotypes do not match genotypes-
942 unexpected phenotypic diversity and potential environmental constraints in Icelandic
943 stickleback. *Journal of Heredity* 103:579-584.
- 944 Lucek K, Keller I, Nolte AW, Seehausen O. 2018. Distinct colonization waves underlie the
945 diversification of the freshwater sculpin (*Cottus gobio*) in the Central European Alpine region.
946 *Journal of Evolutionary Biology* 31:1254-1267.
- 947 MacColl ADC, Aucott B. 2014. Inappropriate analysis does not reveal the ecological causes of
948 evolution of stickleback armour: a critique of Spence et al. 2013. *Ecology and Evolution*
949 4:3509-3513.
- 950 Magalhaes IS, Agostino DD, Hohenlohe PA, Maccoll ADC. 2016. The ecology of an adaptive
951 radiation of three-spined stickleback from North Uist, Scotland. *Molecular Ecology* 25:4319-
952 4336.
- 953 Makinen HS, Merila J. 2008. Mitochondrial DNA phylogeography of the three-spined
954 stickleback (*Gasterosteus aculeatus*) in Europe - Evidence for multiple glacial refugia.
955 *Molecular Phylogenetics and Evolution* 46:167-182.
- 956 Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure
957 Population Inference from RADseq Data. *Molecular Biology and Evolution* 35:1284-1290.
- 958 Mallet J. 2007. Hybrid speciation. *Nature* 446:279-283.
- 959 Marques DA, Meier JI, Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive
960 Radiation. *Trends in ecology & evolution*.
- 961 Martin-Bravo S, Valcarcel V, Vargas P, Luceno M. 2010. Geographical speciation related to
962 Pleistocene range shifts in the western Mediterranean mountains (Reseda sect. *Glaucoreseda*,
963 *Resedaceae*). *Taxon* 59:466-482.
- 964 McKinnon JS, Mori S, Blackman BK, David L, Kingsley DM, Jamieson L, Chou J, Schluter
965 D. 2004. Evidence for ecology's role in speciation. *Nature* 429:294-298.

- 966 McKinnon JS, Rundle HD. 2002. Speciation in nature: the threespine stickleback model
967 systems. *Trends in Ecology & Evolution* 17:480-488.
- 968 McPhail JD. 1992. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*) - evidence
969 for a species-pair in Paxton Lake, Texada island, British Columbia. *Canadian Journal of*
970 *Zoology-Revue Canadienne De Zoologie* 70:361-369.
- 971 McPhail JD. 1993. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*) - origin of
972 the species pairs. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 71:515-523.
- 973 McPhail JD. 1994. Speciation and the evolution of reproductive isolation in the sticklebacks
974 (*Gasterosteus*) of south-western British Columbia. In: Bell MA, Foster SA, editors. *The*
975 *evolutionary biology of the threespine stickleback*: Oxford University Press.
- 976 Meier JJ, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of Parallel
977 Ecological Speciation in Lake Victoria Cichlids *Molecular Biology and Evolution* 35:2594-
978 2594.
- 979 Meier JJ, Sousa VC, Marques DA, Selz OM, Wagner CE, Excoffier L, Seehausen O. 2017.
980 Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia
981 cichlid species after hybridization. *Molecular Ecology* 26:123-141.
- 982 Nesbo CL, Fossheim T, Vollestad LA, Jakobsen KS. 1999. Genetic divergence and
983 phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect
984 glacial refugia and postglacial colonization. *Molecular Ecology* 8:1387-1404.
- 985 Nosil P. 2012. *Ecological speciation*. Oxford: Oxford University Press.
- 986 Pellissier T, Al Nafea H, Good SV. 2018. Divergence of insulin superfamily ligands, receptors
987 and Igf binding proteins in marine versus freshwater stickleback: Evidence of selection in
988 known and novel genes. *Comparative Biochemistry and Physiology D-Genomics & Proteomics*
989 25:53-61.
- 990 Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-
991 Wide Allele Frequency Data. *Plos Genetics* 8:17.
- 992 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus
993 genotype data. *Genetics* 155:945-959.
- 994 R.Core.Team. 2017. R: a language and environment for statistical computing. Version Version
995 3.4.1. Vienna, Austria: R Foundation for Statistical Computing.
- 996 Rafinski J, Banbura J, Przybylski M. 1989. Genetic differentiation of fresh-water and marine
997 sticklebacks, (*Gasterosteus aculeatus*) of eastern-Europe. *Zeitschrift Fur Zoologische*
998 *Systematik Und Evolutionsforschung* 27:33-43.
- 999 Rahn AK, Krassmann J, Tsobanidis K, MacColl ADC, Bakker TCM. 2016. Strong neutral
1000 genetic differentiation in a host, but not in its parasite. *Infection Genetics and Evolution* 44:261-
1001 271.

- 1002 Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018. The genomic
1003 landscape at a late stage of stickleback speciation: High genomic divergence interspersed by
1004 small localized regions of introgression. *Plos Genetics* 14.
- 1005 Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the
1006 spatial scale of evolution. *Trends in Ecology & Evolution* 29:165-176.
- 1007 Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL,
1008 Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers
1009 facilitated by hybridization. *Science* 301:1211-1216.
- 1010 Roesti M, Kueng B, Moser D, Berner D. 2015. The genomics of ecological vicariance in
1011 threespine stickleback fish. *Nature Communications* 6.
- 1012 Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under
1013 mixed models *Bioinformatics* 19:1572-1574.
- 1014 Rundle HD, Nosil P. 2005. Ecological speciation. *Ecology Letters* 8:336-352.
- 1015 Rundle HD, Schluter D. 2004. Natural selection and ecological speciation in sticklebacks. In:
1016 Dieckmann U, Doebeli M, Metz JAJ, Tautz D, editors. *Adaptive Speciation* Cambridge:
1017 Cambridge University Press.
- 1018 Schluter D. 1996. Ecological speciation in postglacial fishes. *Philosophical Transactions of the*
1019 *Royal Society of London Series B-Biological Sciences* 351:807-814.
- 1020 Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323:737-741.
- 1021 Schluter D, Clifford EA, Nemethy M, McKinnon JS. 2004. Parallel evolution and inheritance
1022 of quantitative traits. *American Naturalist* 163:809-822.
- 1023 Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proceedings of the National*
1024 *Academy of Sciences of the United States of America* 106:9955-9962.
- 1025 Schluter D, McPhail JD. 1992. Ecological character displacement and speciation in
1026 sticklebacks. *American Naturalist* 140:85-108.
- 1027 Schumer M, Rosenthal GG, Andolfatto P. 2014. HOW COMMON IS HOMOPLOID HYBRID
1028 SPECIATION? *Evolution* 68:1553-1560.
- 1029 Schumer M, Rosenthal GG, Andolfatto P. 2018. What do we mean when we talk about hybrid
1030 speciation? *Heredity* 120:379-382.
- 1031 Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution*
1032 19:198-207.
- 1033 Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL,
1034 Saetre G-P, Bank C, Braennstroem A et al. 2014. Genomics and the origin of species. *Nature*
1035 *Reviews Genetics* 15:176-192.

- 1036 Simard F, Ayala D, Kamdem GC, Pombi M, Etouna J, Ose K, Fotsing J-M, Fontenille D,
1037 Besansky NJ, Costantini C. 2009. Ecological niche partitioning between *Anopheles gambiae*
1038 molecular forms in Cameroon: the ecological side of speciation. *BMC Ecology* 9:17-Article
1039 No.: 17.
- 1040 Soltis PS, Soltis DE. 2009. The Role of Hybridization in Plant Speciation. *Annual Review of*
1041 *Plant Biology* 60:561-588.
- 1042 Sousa V, Hey J. 2013. Understanding the origin of species with genome-scale data: modelling
1043 gene flow. *Nature Reviews Genetics* 14:404-414.
- 1044 Taylor EB, McPhail JD. 2000. Historical contingency and ecological determinism interact to
1045 prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society B-Biological*
1046 *Sciences* 267:2375-2384.
- 1047 Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M,
1048 Konijnendijk N, Mrosso HDJ et al. 2006. Divergent selection on opsins drives incipient
1049 speciation in Lake Victoria cichlids. *Plos Biology* 4:2244-2251.
- 1050 Volpe JP, Ferguson MM. 1996. Molecular genetic examination of the polymorphic arctic charr
1051 *Salvelinus alpinus* of Thingvallavatn, Iceland. *Molecular Ecology* 5:763-772.
- 1052 Von Hippel FA, Weigner H. 2004. Sympatric anadromous-resident pairs of threespine
1053 stickleback species in young lakes and streams at Bering Glacier, Alaska. *Behaviour* 141:1441-
1054 1464.
- 1055 Wang IJ, Glor RE, Losos JB. 2013. Quantifying the roles of ecology and geography in spatial
1056 genetic divergence. *Ecology Letters* 16:175-182.
- 1057 Wund MA, Baker JA, Clancy B, Golub JL, Fosterk SA. 2008. A test of the "Flexible stem"
1058 model of evolution: Ancestral plasticity, genetic accommodation, and morphological
1059 divergence in the threespine stickleback radiation. *American Naturalist* 172:449-462.
- 1060 Ziegler M, Jilbert T, de Lange GJ, Lourens LJ, Reichart GJ. 2008. Bromine counts from XRF
1061 scanning as an estimate of the marine organic carbon content of sediment cores. *Geochem.*
1062 *Geophys. Geosyst.* 9.
- 1063 Ziuganov VV. 1995. Reproductive isolation among lateral plate phenotypes (low, partial,
1064 complete) of the threespine stickleback, *Gasterosteus aculeatus*, from the White Sea basin and
1065 the Kamchatka Peninsula, Russia. *Behaviour* 132:1173-1181.

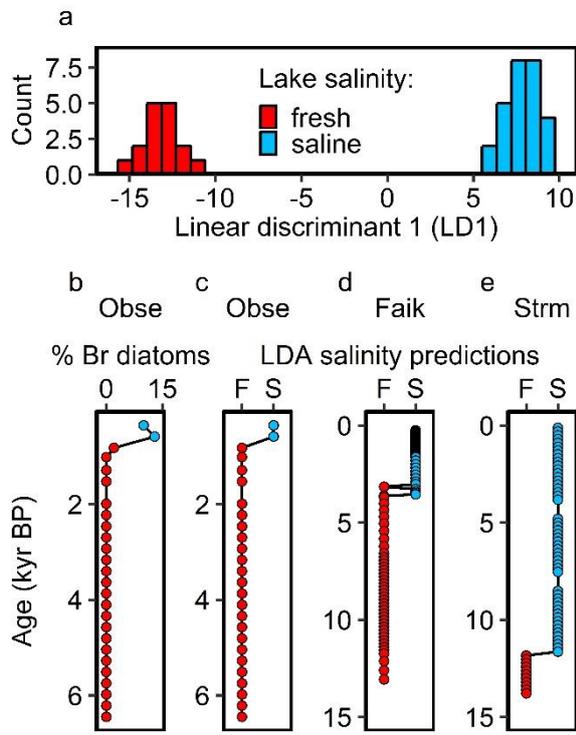
1066 **Data accessibility**

- 1067 Data from this manuscript is provided as supplementary material. Mitochondrial sequences are
1068 available on GeneBank under accession numbers: MG602878-MG602951.

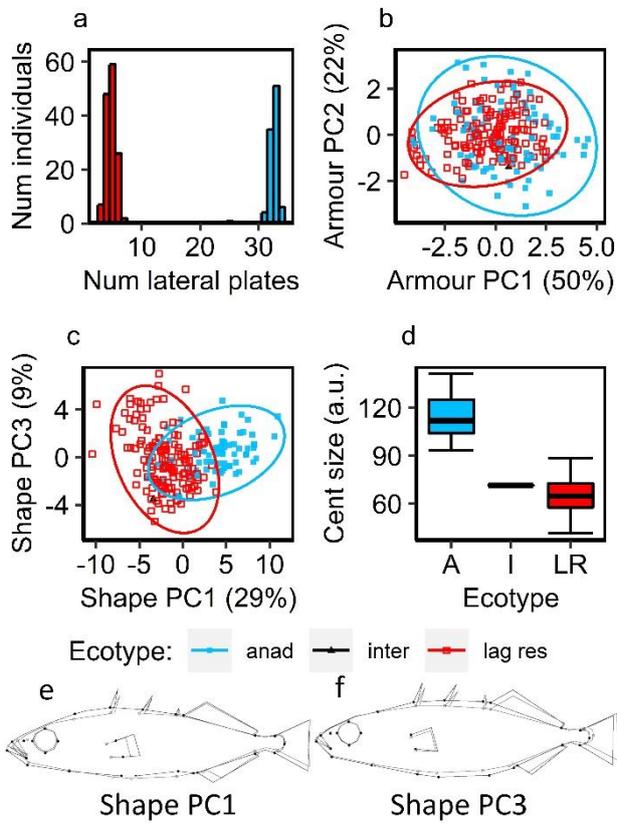
1069 **Author contributions**

1070 L.L.D. wrote the manuscript, collected sediment cores and stickleback samples, conducted lab
1071 work and data analyses. I.S.M. collected and conducted initial processing and analyses of RAD-
1072 seq SNP data. A.F. assisted in genetic analyses and drafting the manuscript. D.D. collected
1073 samples for RAD-seq SNP data. S.M. collected sediment core samples and assisted in their
1074 analyses. A.D.C.M. collected stickleback samples, assisted in data analyses and helped draft
1075 the manuscript. A.D.C.M and L.L.D conceived the study.

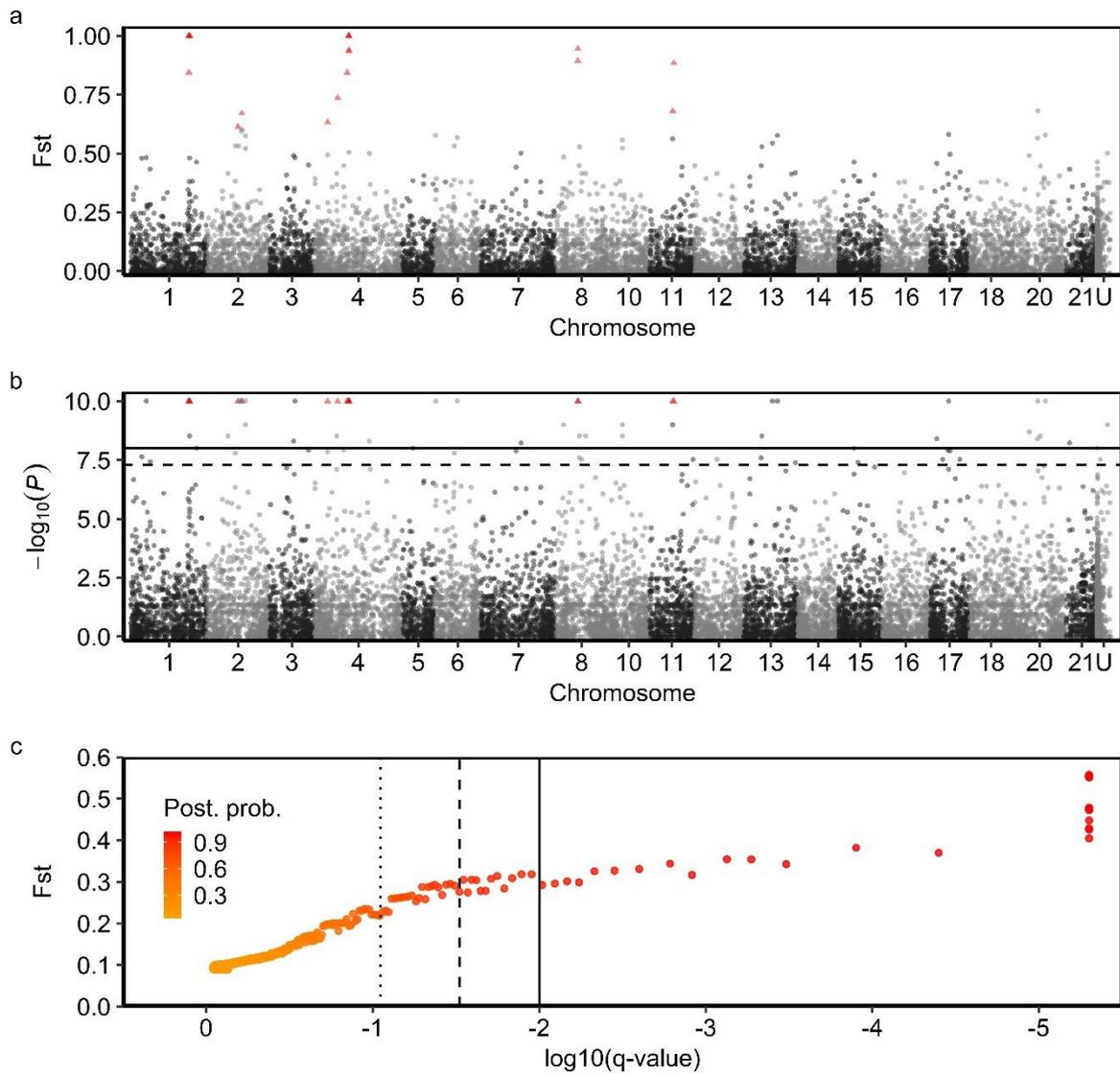
1080 **Figure 2**



1081



1084 **Figure 4**



1085

