1   Contrasting hybridization rates between sympatric threespine sticklebacks highlight the

2   fragility of reproductive barriers between evolutionarily young species

3

4   JENNIFER L. GOW*, CATHERINE L. PEICHEL† AND ERIC B. TAYLOR*

5

6   *Department of Zoology, Biodiversity Research Centre and Native Fishes Research Group, University

7   of British Columbia, 6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada

8   † Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109,

9   USA

10

12

14

15   Corresponding author: J. L. Gow, School of Biological Sciences, University of Wales, Bangor,

16   Gwynedd, LL57 2UW, UK

17

18   Fax: (01248) 371644, E-mail: j.gow@bangor.ac.uk

19

20   Running head: Hybridization in stickleback species pairs

1

Abstract

Threespine sticklebacks (*Gasterosteus aculeatus*) are a powerful evolutionary model system due to the rapid and repeated phenotypic divergence of freshwater forms from a marine ancestor throughout the northern hemisphere. Many of these recently derived populations are found in overlapping habitats, yet are reproductively isolated from each other. This scenario provides excellent opportunities to investigate the mechanisms driving speciation in natural populations. Genetically distinguishing between such recently derived species, however, can create difficulties in exploring the ecological and genetic factors defining species boundaries, an essential component to our understanding of speciation. We overcame these limitations and increased the power of analyses by selecting highly discriminatory markers from the battery of genetic markers now available. Using species diagnostic molecular profiles, we quantified levels of hybridization and introgression within three sympatric species pairs of threespine stickleback. Sticklebacks within Priest and Paxton lakes exhibit a low level of natural hybridization and provide support for the role of reinforcement in maintaining distinct species in sympatry. In contrast, our study provides further evidence for a continued breakdown of the Enos Lake species pair into a hybrid swarm, with biased introgression of the "limnetic" species into that of the "benthic"; a situation that highlights the delicate balance between persistence and breakdown of reproductive barriers between young species. A similar strategy utilizing the stickleback microsatellite resource can also be applied to answer an array of biological questions in other species pair systems in this geographically widespread and phenotypically diverse model organism.

Introduction

Threespine sticklebacks (*Gasterosteus aculeatus*) are a powerful evolutionary model system due to the rapid phenotypic divergence of freshwater forms from an ancestral marine colonizer throughout the northern hemisphere since the end of the last ice age, ten to fifteen thousand years ago (Bell & Foster 1994). The presence of multiple, independently derived populations, which are now reproductively isolated from each other, provides excellent opportunities to investigate the mechanisms driving speciation in natural populations. For example, parallel evolution of size-assortative mating between anadromous and stream-resident forms throughout the range of the *G. aculeatus* species complex suggests that reproductive isolation may arise largely as a result of ecological differences and divergent selection on a few phenotypic traits (McKinnon *et al*. 2004). As a first step in understanding the genetics of such phenotypic evolution, parallel inheritance of body shape and lateral plate number has been shown in different freshwater lineages (Schluter *et al*. 2004). Further insight into the genetic architecture of parallel phenotypic evolution in natural populations is now being gleaned from genetic mapping studies made feasible by the development of a microsatellite linkage map for sticklebacks (Peichel *et al*. 2001). Parallel changes in lateral plate number (Colosimo *et al*. 2004; Cresko *et al*. 2004) and pelvic reduction (Cresko *et al*. 2004; Shapiro *et al*. 2004) appear to be caused by changes at the same genetic loci. Indeed, Colosimo *et al*. (2005) have shown that the Ectodysplasin gene is crucial to stickleback plate morph development and that parallel evolution of most low-plated freshwater phenotypes has occurred by repeated selection of ancestral marine alleles at this locus. This highlights the potential for adaptive divergence to arise repeatedly from relatively few genetic changes, a phenomenon that may help to explain the rapid evolution of natural stickleback populations (Bell *et al*. 2004).

The development of new genomic and genetic tools for threespine sticklebacks now holds promise for the identification of the actual genes and mutations responsible for evolutionary change in physiological, behavioural and morphological traits, as well as the processes that drive such change (Kingsley *et al*. 2004). These same genomic resources, however, may also be utilized to address challenging questions in population ecology. For example, understanding the population dynamics at stickleback species boundaries is an essential component of our understanding of the mechanisms driving speciation (Barton & Hewitt 1985). This has been hampered, however, by the difficulty in adequately differentiating genetically between such recently derived species. Whilst many studies have used species diagnostic markers, which greatly increase the power of analyses, to examine hybridization dynamics amongst older species of fish (e.g. Ostberg *et al*. 2004; Rubridge & Taylor 2004; Bettles *et al*. in press), identifying such markers amongst younger species that share a more recent common ancestry is challenging. Armed with the battery of stickleback genetic markers now available (Peichel *et al*. 2001), however, we are able to efficiently search for highly discriminatory markers, overcoming these limitations and enabling us to explore interactions at the interface of stickleback species pairs.

Species pairs of threespine stickleback

Several systems of stickleback species pairs have been recognized, where a pair of morphologically and ecologically divergent forms naturally coexists and exhibit varying degrees of reproductive isolation (reviewed in McKinnon & Rundle 2002). Those systems that include a freshwater form are

86   considered of recent origin, having diverged from marine ancestors since the colonization of fresh
87   water after the retreat of the Pleistocene glaciers, about ten to fifteen thousand years ago. This includes
88   many parapatric anadromous-freshwater and lake-stream pairs, as well as sympatric benthic-limnetic
89   systems. Evidence amongst the more recently derived forms implicates ecologically-based divergent
90   selection in the evolution of both pre- and postzygotic reproductive isolation (reviewed in McKinnon &
91   Rundle 2002), with the benthic-limnetic lake pairs having been studied the most extensively.
92
93   Extant benthic-limnetic pairs have evolved independently in three separate water drainages (Taylor &
94   McPhail 1999, 2000) in coastal British Columbia: Enos Lake on Vancouver Island (49^o17'N,
95   124^o10'W); Priest and Emily lakes in the Vananda Creek drainage (49^o45'N, 124^o34'W), and Paxton
96   Lake (49^o43'N, 124^o30'W) on Texada Island (McPhail 1984, 1992, 1994). Much of their divergence
97   from their marine ancestors may have occurred in sympatry after a brief period of allopatry (about 2000
98   years) between a first and second marine incursion that brought two waves of colonists into the lakes
99   (known as the double invasion hypothesis [McPhail 1993; Taylor & McPhail 2000]). Regardless of the
100   exact sequence of initial events bringing the fish into these lakes, it appears that ecological speciation
101   has been a major factor driving their divergence. Both comparative and experimental work on the
102   group has strongly implicated divergent selection caused by interspecific resource competition
103   (Bentzen & McPhail 1984; Schluter 1993, 1994, 1995, 2003; Schluter & McPhail 1992) and predation
104   (Rundle *et al*. 2003; Vamosi & Schluter 2004) in the origin and maintenance of the divergence of the
105   species pairs.
106
107   In each of these sympatric pairs, one member is known as the "benthic" stickleback, a bottom-dwelling
108   fish foraging mainly on large invertebrates from sediment or on plants in littoral habitats, and the other
109   member is known as the "limnetic" stickleback, a specialized pelagic zooplankton feeder. This
110   ecological divergence is associated with consistent morphological differences; limnetics tend to be
111   shorter and more slender, with a narrower mouth and more numerous, longer gill rakers (the
112   protruberances along the gill arches that sieve ingested prey and direct fluid movement within the
113   buccal cavity; Sanderson *et al.* 1991), compared with the longer, more robust, wider-mouthed benthics
114   (Schluter & McPhail 1992). During the breeding season, they share the littoral zone, where they
115   frequently encounter one another despite occupying different microhabitats (Bentzen *et al.* 1984).
116   Although strong assortative mating, predominantly based on size, prevents extensive admixture,
117   premating isolation between ecomorphs is not complete, and rare natural hybrids (about 1-2 % of
118   adults) have been reported based on their intermediate morphology (McPhail 1984, 1992). There is no
119   evidence of genetic incompatibilities in laboratory-bred hybrids (with the possible exception of benthic
120   backcrosses), which are fully viable and fertile (McPhail 1984, 1992; Hatfield & Schluter 1999).
121   However, given the role of divergent natural selection and strong size-based assortative mating, these
122   phenotypically intermediate fish can be expected to show a poorer fit to available niches and reduced
123   mating success relative to the parental species. Indeed, experimental evidence suggests that extrinsic,
124   rather than intrinsic, post-zygotic reproductive isolating mechanisms, including both ecologically based
125   post-zygotic isolation (Schluter 1995; Hatfield & Schluter 1999; Rundle 2002) and sexual selection
126   against hybrid males (Vamosi & Schluter 1999) are in force.
127
128   In order to better understand the role of these forms of reproductive isolation in the divergence of the
129   species pairs, as well as their influence on reinforcement (Rundle & Schluter 1998), the extent of gene
130   flow and rates of hybridization and introgression between the species in nature must be quantified.

131 Although estimates based on morphological assessment of wild populations suggest a low but
132 persistent level of hybridization between the two species (McPhail 1984, 1992, 1994), they may be
133 underestimates of the amount of intercrossing as the discriminant function analysis used probably does
134 not detect backcrosses (McPhail 1992). Furthermore, although the phenotypic differences between the
135 species have a genetic basis and lab-reared hybrids are morphologically intermediate (McPhail 1984,
136 1992; Hatfield 1997; Peichel *et al.* 2001), the traits exhibit plasticity in an adaptive direction (Day *et al.*
137 1994), creating wide error margins around estimates of natural hybridization rate based on morphology.
138
139 Here we seek unambiguous genetic assessment of hybridization rates between benthics and limnetics.
140 Although previous allozyme (McPhail 1984, 1992) and molecular work (Taylor & McPhail 1999,
141 2000) indicated that the species are genetically distinct within each lake, the resolution of the genetic
142 markers has been inadequate to accurately assess gene flow. We overcame this constraint by selecting
143 highly discriminatory markers from the large microsatellite resources generated by the creation of the
144 stickleback linkage map. The development of such species diagnostic molecular profiles for the three
145 extant sympatric species pairs of threespine sticklebacks has enabled us to provide the first quantitative
146 estimates of hybridization and introgression within the species pairs.
147
148 **Materials and methods**
149
150 *Selection of highly discriminatory microsatellite markers*
151
152   *Sample selection*
153
154 'Pure' representatives of each species from each of the three species pairs were required to identify
155 genetic markers that clearly differentiated the two species. Therefore, 48 benthic and 48 limnetic fish
156 were chosen from existing specimen collections from Priest and Paxton lakes (preserved in 95%
157 ethanol) using the following morphometric analysis. After tissue was removed for DNA extraction,
158 fish were soaked in 10% formalin for one week and stained with alizarin red as previously described
159 (Peichel *et al.* 2001). A suite of morphological traits associated with differences between benthics and
160 limnetics were measured: body length, body depth, pelvic spine length, pelvic girdle width, gape width,
161 snout length and number of gill rakers were assessed as described in Schluter and McPhail (1992); the
162 number of lateral plates and gill raker length was measured according to Lavin & McPhail (1985). For
163 each lake, a two-dimensional multivariate scatter plot produced using non-standardized Euclidean
164 distances in SPSS version 11 (SPSS Inc.) revealed two distinct clusters (Figure 1) that corresponded to
165 previous morphological differences used to distinguish benthics from limnetics, highlighting the
166 suitability of these samples for marker selection.
167
168 Morphological samples were more limited for Enos Lake. It is suspected that Enos Lake is
169 experiencing a species breakdown (Kraak *et al.* 2001; Taylor *et al.* in press) and ethanol-preserved
170 material prior to extensive introgression is scarce. In order to overcome this potentially confounding
171 factor, only the most morphologically and genetically extreme samples from the oldest DNA surveys of
172 this lake were included (benthics: 10 from 1994 and 15 from 1997; limnetics: 11 from 1994 and 15
173 from 1997), based on analysis from Taylor & McPhail (2000) and Taylor *et al.* (in press).
174
175   *Screening of microsatellite library*

176
177 In order to confidently identify hybrids within the lakes (Cornuet *et al*. 1999; Anderson & Thompson
178 2002), we searched for a suite of species-diagnostic markers for each lake pair. DNA was extracted
179 from the right pectoral fin of these individuals using a Qiagen DNeasy Tissue Kit. The DNA from
180 each of the 48 benthics and 48 limnetics from Priest and Paxton lakes were pooled at equal
181 concentration into their species and lake types. These four pools of DNAs were screened for
182 microsatellite variation at 288 *G. aculeatus*-derived loci using the polymerase chain reaction (PCR) and
183 genotyping procedures outlined in Peichel *et al*. (2001), with the exception that 0.5 ng of each
184 individual sample's DNA was present in the PCR. Those loci showing potentially non-overlapping
185 allele (NOA) ranges between benthics and limnetics from either lake were explored further by
186 genotyping eight individuals of each species, according to Peichel *et al*. (2001). Those loci still
187 showing a NOA range were verified by genotyping all 96 individuals from the entire lake's panel.
188
189 As the DNA samples from Enos Lake were restricted in number and amount available, and were
190 variable in quality, an altered strategy from the pooled DNA approach was used. Initially, four of the
191 most morphologically and genetically differentiated limnetics and benthics were screened for
192 microsatellite variation at 192 *G. aculeatus*-derived loci using the methods outlined in Peichel *et al*.
193 (2001). PCR products were visualized using ethidium bromide on a 2 % agarose gel and those loci
194 showing a potentially discriminating pattern between benthics and limnetics were shortlisted for
195 genotyping of the entire Enos sample panel of 51 individuals on an ABI 3700 sequencer according to
196 Peichel *et al*. (2001). Together, these procedures used to screen sticklebacks from Priest, Paxton and
197 Enos lakes identified nine loci that highly discriminated between benthics and limnetics in one or more
198 of the species pairs (see Results; Table 1).
199
200 *Assessment of gene flow and hybridization within species pairs*
201
202   *Sample collection and microsatellite genotyping*
203
204 We collected between 192 and 198 sub-adult fish from Priest, Paxton and Enos lakes during September
205 and October 2003. Thirty minnow traps distributed approximately evenly along the entire littoral zone
206 shoreline were used in conjunction with dip-netting throughout the same area to ensure lake-wide
207 samples in which the two species were represented approximately equally. No selection was made
208 against indeterminate forms i.e. fish with ambiguous morphology were not discarded. Fish were
209 sacrificed in MS-222 and preserved in 95 % ethanol before DNA extraction. In addition to the suite of
210 nine species diagnostic microsatellite loci, these samples were genotyped at five additional *G.*
211 *aculeatus*-derived microsatellite loci (*Gac*4, 7, 10 and 14: Taylor 1998; *Cir*51: Rico *et al*. 1993) that
212 have been useful in previous population genetic surveys of these populations (Taylor & McPhail 2000).
213 This not only increased the statistical power of analysis, but also retained consistency for population
214 monitoring purposes. All 14 loci were amplified using the PCR conditions described above except that
215 primer concentration varied from 25 to 500 nM, and different cycling conditions were used for the
216 latter five loci: initial denaturation at 95 $^{\circ}$C for 3 min was followed by 5 cycles of 94 $^{\circ}$C for 30 s, 60 $^{\circ}$C
217 for 1 min and 72 $^{\circ}$C for 1 min. A subsequent 25 cycles of 92 $^{\circ}$C for 30 s, 58 $^{\circ}$C for 30 s and 72 $^{\circ}$C for
218 30 s finished with a 7 min extension step at 72 $^{\circ}$C. PCR products from 2 to 3 different primer pairs
219 were then pooled and analyzed on a CEQ 8000 Genetic Analysis System (Beckman Coulter) with CEQ
220 DNA Size Standard Kit-400 used as internal size standard. Locus *Stn*387 consistently amplified only

221   Priest samples on the CEQ 8000 despite attempted re-optimization of PCR conditions. Therefore, this
222   locus was excluded from statistical analyses.
223
224   *Statistical analyses*
225
226   **Hybrid identification and assignment:** NewHybrids Version 1.1 (Anderson & Thompson 2002) was
227   used to categorize individual fish from each lake into benthics, limnetics or hybrids (F_1, F_2, limnetic or
228   benthic backcross). This program implements a model-based Bayesian method that employs Markov
229   chain Monte Carlo (MCMC) sampling to compute posterior probabilities that individuals in a sample
230   (known to consist of pure individuals and recent hybrids of two species) fall into parental or different
231   hybrid categories. Individuals were assigned to the category with the highest posterior probability. To
232   minimize the effect of the over-dispersed starting values during the Monte Carlo simulation, we
233   simulated 1000 sweeps of the Markov chain before data for the parameter estimation were collected
234   from another 10^6 iterations. Three independent runs of the Markov chain, each of least 10^6 updates,
235   were performed to assure convergence of the chain and homogeneity among runs. Differences in
236   individual posterior probabilities between different runs of the Markov chain for Priest and Paxton
237   lakes never exceeded 1 %. Individual Enos values varied up to 4 % but individual category assignment
238   never changed.
239
240   NewHybrids assumes that any linkage disequilibrium or deviations from Hardy-Weinberg equilibrium
241   (HWE) expectations are entirely the result of admixing of the 'parent' populations. The fulfillment of
242   these assumptions in our data set was tested using the 'parent' benthic and limnetic subsamples from
243   Priest and Paxton lakes, which excluded individuals identified as hybrids by NewHybrids. Enos Lake
244   samples were excluded from this test as no 'pure' limnetics were detected from this lake's sample (see
245   Results). Using a Markov chain method in GENEPOP version 3.3 (Raymond & Rousset 2001), the Fisher
246   exact test revealed only four out of 300 tests between locus pairs within populations to be in genotypic
247   linkage disequilibrium ($P < 0.05$, using the sequential Bonferroni procedure [Rice 1989]), involving
248   three locus pairs in three populations. Weir & Cockerham's (1984) estimator *f* of the inbreeding
249   coefficient, F_{IS}, was tested at each locus within each population using FSTAT version 2.9.3 (Goudet
250   2001). About half of the tests showed significant deviations from genotypic frequencies expected
251   under HWE ($P < 0.05$ in 27 out of 52 tests using the sequential Bonferroni procedure [Rice 1989]).
252   Populations had between four and nine significant single locus results. No locus was in Hardy-
253   Weinberg disequilibrium across all samples, although *Stn216* was the only one showing no incidence
254   of departure from HWE within any sample. These deviations from HWE are perhaps unsurprising,
255   given the biased method that was employed to select diagnostic markers. Indeed, two of these
256   microsatellites are linked to known morphological quantitative trait loci; *Stn216* is linked to a plate size
257   modifier (Colosimo *et al*. 2004) and *Stn*43 is on linkage group four, near the plate morph major locus
258   (Peichel *et al*. 2001; Colosimo *et al*. 2004).
259
260   The power of many new assignment-based methods has not been widely explored over different
261   evolutionary scenarios or with markers exhibiting various levels of variation. We, therefore, ran
262   simulations to test the ability of NewHybrids to detect hybrids given the level of polymorphism in our
263   data set. We generated artificial multilocus hybrid genotypes between benthics and limnetics from
264   Priest and Paxton lakes using HYBRIDLAB (see Nielsen *et al*. 2001), which draws alleles randomly from

265 the observed allele frequency distribution for each population. The 'pure' parental populations used for
266 this comprised individuals whose multilocus genotype was assigned to a parent population with a
267 posterior probability exceeding 0.99 in the NewHybrids analysis (n = 92, 91, 93 and 74 for benthics
268 and limnetics in Priest and Paxton lakes, respectively). Two different sets of hybrids were constructed
269 for simulation: firstly, we created ten F_1 hybrids alone for each species pair; secondly, we generated ten
270 of each of F_1, F_2, limnetic and benthic backcross using the parents for the F_1, as well as two simulated
271 groups of 100 F_1 each for the latter three categories. These groups of ten or forty artificial hybrids,
272 along with the 'parent' samples used to generate them, were subsequently analysed in NewHybrids, as
273 described above, and the accuracy of their assignment assessed. Each type of simulation was repeated
274 five times for each species pair. Simulations were not run using Enos Lake samples because no 'pure'
275 limnetics were detected from this lake from which to construct artificial hybrids.
276
277 Factorial Correspondence Analysis (FCA) in GENETIX version 4.03 (Belkhir et al. 2001) was used to
278 project individuals in microsatellite allele frequency space. By visualizing the relative similarity
279 among all the samples in this way, the relative distribution of the hybrids identified by NewHybrids
280 was presented. Ten of the 19 hybrids from Priest and Paxton lakes showed a smaller probability in
281 NewHybrids of belonging to one of the two pure parental forms, in addition to the higher probability of
282 belonging to their designated hybrid class. However, nine had a posterior probability of 1.00 for their
283 hybrid class or had posterior probabilities distributed only amongst hybrid classes. In order to define
284 two populations within each lake for the purpose of further analyses, we calculated which parental
285 species the hybrids were genetically more similar to using a likelihood-based Bayesian method of
286 assignment (Rannala & Mountain 1997) in GeneClass2 (Piry et al. 2004). This method employed a
287 Monte Carlo resampling algorithm (Paetkau et al. 2004) to compute the probability that each of the 19
288 hybrids belonged to either 'pure/parental' reference population. Based on 10 000 simulated
289 individuals, we assigned hybrids to the closest one. These Priest and Paxton groupings of benthics and
290 limnetics were used to analyze long term gene flow. Given that no 'pure' limnetics were detected in
291 Enos Lake, its fish were divided into a benthic and a hybrid group based on the NewHybrids results.
292
293 **Long term gene flow estimation:** To complement the information on recent gene flow derived from
294 hybridization rate estimates, we also sought comparative estimates of longer term gene flow (m). Such
295 estimators derived from the simple mathematical relationship between gene flow and genetic
296 differentiation (e.g. $F_{ST} = 1/[1 + 4N_e m]$, Wright [1931]) have been widely criticized for their unrealistic
297 ecological assumptions, such as constant population size, symmetrical migration and population
298 persistence to enable genetic equilibrium (Whitlock & McCauley 1999). Furthermore, this data set is
299 likely to overestimate F_{ST} (and so underestimate m) because of bias derived from using a set of
300 diagnostic markers that were chosen to accentuate species' differences relative to the average marker.
301
302 Coalescence theory, which follows ancestral genealogies of samples opposed to modeling changes of
303 gene frequencies in the entire population, has enabled the relaxation of many of these restrictive
304 assumptions. Accounting for unequal migration rates and population sizes, the maximum-likelihood
305 coalescent program MIGRATE version 2.0 (Beerli & Felsenstein 1999) measures long-term gene flow.
306 We used the microsatellite model to simultaneously estimate the effective population size (N_e) and the
307 proportion of migrants (m) for each population. The microsatellite threshold, which specifies the
308 window in which probabilities of change between allelic states are calculated, was set at 15 repeat units

309 to ensure that all allelic ranges were encompassed in the analysis. A mutation rate of 10^{-4} (Feldman
310 1999) was assumed, which seems reasonable given that the most extensive assessment of microsatellite
311 mutation rate in fish yielded an estimate of 1.5×10^{-4} (Shimoda *et al.* 1999). Summing the two
312 unidirectional estimates of N_e gives the total N_e for each species pair, whilst dividing the sum of the
313 two $N_e m$ estimates by total N_e calculates the total m. Approximate 95 % confidence intervals for
314 estimates of N_e and m were generated from a summary of profile likelihood percentiles of all
315 parameters.
316
317
318 **Results**
319
320 *Species diagnostic molecular profiles*
321
322 The screening of 288 microsatellites identified 84 and 103 loci with potentially NOA (non-overlapping
323 allele) ranges between benthic and limnetic pools of DNA from Paxton and Priest lakes, respectively.
324 Further screening of eight benthics and eight limnetics at 84 (Paxton) and 18 (Priest) of these loci,
325 revealed four and two markers, respectively, that showed NOA ranges between benthics and limnetics.
326 This pattern was verified by genotyping all individuals from the entire lake's panel, with only a few
327 exceptional instances of individuals carrying an allele that was from the range of the other species
328 (Table 1). These exceptions were assumed to be a signature of introgression, although the possibility
329 of incomplete assortment of ancestral polymorphism cannot be excluded. The alternative strategy
330 designed to conserve the historical Enos Lake DNA resources found 23 potentially discriminatory loci
331 from the 192 microsatellites screened. Of these, 5 distinguished between benthics and limnetics upon
332 further analysis (Table 1).
333
334 *Hybridization and long term gene flow within species pairs*
335
336 Out of 198 individuals from Priest Lake, NewHybrids categorized 92 as pure benthics ($P \geq 0.99$) and
337 97 as pure limnetics ($P = 0.62 - 1.00$; Figure 2). Ninety of these limnetics had a posterior probability
338 exceeding 0.99. The remaining seven with posterior probabilities < 0.99 also had a lower posterior
339 probability of belonging to a hybrid category ($P = 0.01 - 0.38$). Nine hybrids yield a hybridization rate
340 for Priest Lake species pair of 4.5 %. All of the hybrids were categorized as F_2 ($P = 0.67 - 1.00$),
341 although seven of them also had a lower posterior probability of belonging to another hybrid category
342 ($P = 0.02 - 0.33$). In contrast, only three had a posterior probability ($P = 0.04 - 0.24$) of belonging to
343 one or other of the parental populations, with six having a probability 0.00 of being either pure benthic
344 or limnetic. That is, two-thirds of the hybrids have probability 1.00 of being hybrids of some sort.
345 GeneClass2 assigned three hybrids to the limnetic reference population and six to the benthic one.
346
347 In Paxton Lake, 96 out of 192 individuals were categorized as pure benthics ($P = 0.79 - 1.00$) and 86 as
348 pure limnetics ($P = 0.74 - 1.00$; Figure 2). Ninety-three of these benthics and 74 of these limnetics had
349 a posterior probability exceeding 0.99. The remaining three benthics and 12 limnetics with posterior
350 probabilities < 0.99 also had a lower posterior probability of belonging to a hybrid category ($P = 0.01 -$
351 0.18). Ten hybrids yield a hybridization rate for Paxton Lake species pair of 5.2 %. Eight of them
352 categorized as F_2 ($P = 0.47 - 1.00$) and two as benthic backcrosses ($P = 0.72$ and 0.85), although nine
353 of them also had a lower posterior probability of belonging to another hybrid category (three with $P \leq$

354 0.10, six with $P > 0.10$). In contrast, six showed generally lower probabilities of belonging to one or
355 other of the parental populations (four with $P < 0.10$, two with $P > 0.10$), and four had a probability
356 0.00 of being either pure benthic or limnetic. That is, 80 % of the hybrids have probability greater than
357 0.90 of being hybrids of some sort. GeneClass2 assigned half of them to each parental reference
358 population.
359
360 Whilst 146 out of 192 individuals from Enos lake were categorized as pure benthics ($P = 0.49$ -1.00),
361 no pure limnetics were recognized (Figure 2). Furthermore, a smaller portion was assigned with high
362 posterior probabilities compared to Priest and Paxton lakes; only 56 % of those categorized as pure
363 benthics had a posterior probability greater than 0.99, compared to 100 % for Priest benthics and 97 %
364 for Paxton benthics. All 46 hybrids were categorized as F_2 ($P = 0.45$-1.00), although forty of these also
365 had a lower posterior probability of belonging to one (n = 35) or two (n = 5) other hybrid categories (P
366 $= \leq 0.29$): only one hybrid had a lower probability of being F_1; seven had a lower probability of being a
367 limnetic backcross; and 36 had a lower probability of being a benthic backcross. Fewer hybrids
368 showed any chance of belonging to one or other of the parental populations; 25 out of 46 hybrids had a
369 lower posterior probability of being pure benthic (11 with $P \leq 0.10$, 14 with $P \geq 0.10$) and one had of
370 being pure limnetic ($P = 0.04$). That is, 70 % of the hybrids have probability greater than 0.90 of being
371 hybrids of some sort.
372
373 These contrasting levels of hybridization detected within the three lakes are visualized by a two-
374 dimensional FCA (Figure 3). Although the two axes describe only 7.5 and 8.1 % of variability within
375 Priest and Paxton species pairs, respectively, the distinction of two clusters with few intermediate
376 (hybrid) fish is clear. This visualization agrees with the low level of hybridization detected within
377 these lakes, as well as the high posterior probabilities assigned to the vast majority of 'pure' samples by
378 NewHybrids. In contrast, Enos Lake individuals form one diffuse cluster, reflecting the high level of
379 hybridization and introgression of the limnetic form into that of the benthic within this lake. A greater
380 variability amongst the hybrids relative to the benthics is consistent with their having a wider array of
381 genotypic classes from multiple generations of hybridization and introgression.
382
383 The robustness of NewHybrids to deviations from ideal model conditions was tested by looking at
384 assignment patterns of known hybrids. Five repeats of simulations including ten F_1 hybrids along with
385 the 'parent' samples resulted in all samples being correctly assigned with very high posterior
386 probabilities ($P > 0.99$). The simulations that included ten hybrids of each of the hybrid categories F_1,
387 F_2, limnetic and benthic backcross, as well as the 'parent' samples, also resulted in correct assignment
388 of all F_1 (Table 2, $P > 0.77$ with some also showing a lower probability of belonging to other hybrid,
389 but not parental, categories). While there were very few instances of mis-assigned parents (< 1 % of
390 parents from each species pair were mis-assigned as a backcross), there were more mis-assignments of
391 the second generation hybrids: 30 % and 20 % of F_2 from Priest and Paxton lakes, respectively, were
392 wrongly assigned to another hybrid category while 38 % and 56 % of backcrosses from these lakes
393 were wrongly assigned to another, usually parental, category (Table 2). Only some backcrosses had a
394 posterior probability (n = 37 from both species pairs) of belonging to one or other of the parental
395 populations, with the vast majority of simulated hybrids (82 % for each species pair) having a
396 probability 0.00 of being either pure benthic or limnetic i.e. these hybrids have probability 1.00 of
397 being hybrids of some sort.
398

399   Using the populations of benthics and limnetics (or hybrids in the case of Enos Lake) within each lake
400   defined by NewHybrids and Genclass2, MIGRATE estimated N_e to be remarkably consistent across
401   species and lakes, at approximately 1000 individuals per population (Table 3). In contrast, long term
402   gene flow estimates (m) varied four-fold among lakes (Table 3): whilst migration was relatively
403   symmetrical within Priest and Paxton lakes, an overall estimate for the Priest Lake species pair was less
404   than half that for the Paxton Lake one; Enos Lake sticklebacks exhibited migration levels that were
405   more than four times greater than those found in Priest Lake, and nearly two times greater than those
406   found in Paxton Lake.
407
408   **Discussion**
409
410   *Development of species diagnostic molecular profiles from targeted exploration of genome linkage*
411   *map*
412
413   Two strategies that we explored proved effective at targeting diagnostic markers. Using selected
414   samples to screen a large collection of microsatellite loci, we minimized the time and resources
415   expended to identify a suite of microsatellites that can unambiguously distinguish between benthics and
416   limnetics in each of the extant sympatric species pairs of threespine sticklebacks in British Columbia.
417   This was a successful strategy even when baseline samples were limited in number, as was the case for
418   the Enos Lake species pair.
419
420   This approach selected markers that showed accentuated differences between species relative to the
421   average marker. These patterns of allele frequency and range differences between benthics and
422   limnetics within species pairs may simply be the outcome of genetic drift. Another possibility,
423   however, is that they may be linked to loci under selection, with hitchhiking on nearby loci subject to
424   selective sweeps having driven the current differences in allele ranges and frequencies (Maynard Smith
425   & Haigh 1974). Indeed, two of the microsatellites from the species diagnostic molecular profile (*Stn*43
426   and *Stn*216) are linked to known morphological quantitative trait loci (Peichel *et al.* 2001; Colosimo *et*
427   *al.* 2004). Both these and four other microsatellites from the species diagnostic molecular profile show
428   an extremely restricted allele range (one to three alleles) in one member of the species pair. It is
429   conceivable that this pattern is the product of selection via hitchhiking. While demographic
430   bottlenecks could also produce such a pattern, they would do so across the entire genome rather than be
431   restricted to specific regions (Schlötterer 2003). Two of these loci were identified from the screening
432   to have NOA ranges in two of the three species pairs (*Stn*387 and *Stn*254 for both Priest and Enos
433   lakes), which also argues against the sole role of genetic drift in producing these patterns. Genetic drift
434   is unlikely to produce repetitive shifts in the same direction under a specific environmental setting,
435   given the independent origin of each pair (Taylor & McPhail 1999, 2000). Instead, this pattern of a
436   locus being discriminatory in more than one species pair is suggestive of parallel evolution (Schluter &
437   Nagel 1995). This speculation, however, awaits a more rigorous statistical investigation as other
438   processes, such as genomic variation in recombination and mutation rate, could also generate the
439   observed patterns.
440
441   Regardless of the processes driving these differences in microsatellite allele ranges between benthics
442   and limnetics, biasing marker selection towards those that most clearly distinguished between them has
443   enabled us to overcome previous limitations of genetically distinguishing between such evolutionarily

444 young species. As co-dominant markers, these microsatellites can now be used not just as simple
445 diagnostic markers but can serve as powerful population genetic tools. Many pertinent questions about
446 the ecological and genetic basis of adaptive divergence and speciation can now be addressed in
447 sticklebacks, which serve as excellent models for these studies (e.g. Coyne and Orr 2004). The utility
448 of sticklebacks is demonstrated in this study by our exploration of the interactions between species
449 through a quantification of hybridization rates and levels of gene flow within each species pair.
450
451 *Hybridization rates and gene flow within each species pair*
452
453 The genetic assessment of hybridization rates within each lake revealed two contrasting scenarios.
454 Priest and Paxton lakes exhibited a remarkably similar, relatively low level of natural hybridization of
455 about 5 %, with no significant bias in the direction of introgression. In contrast, our study provides
456 further evidence for a continued breakdown of the Enos Lake species pair into a hybrid swarm (see
457 Taylor *et al*. in press), with pronounced biased introgression of limnetics into the benthic population.
458
459 The results for Priest and Paxton lakes support earlier estimates of low hybridization rates (1 to 2 %)
460 within species pairs based on morphological criteria (McPhail 1984, 1992). Although our estimates are
461 slightly higher than these earlier ones, McPhail (1992) recognized the potential for underestimating
462 hybridization rates based on morphology due to the poor ability to detect backcrosses. The
463 NewHybrids methodology enabled the detection of two generations of hybrids (F_1, F_2, and limnetic or
464 benthic backcrosses). Additionally, this study looked at the hybridization rate amongst sub-adult fish,
465 and a reduction in the number of hybrids found amongst sexually mature adults may be expected if
466 ecologically based post-zygotic isolation that has been detected experimentally (Schluter 1995;
467 Hatfield & Schluter 1999) is in force within the natural populations.
468
469 The accuracy of the NewHybrid analysis is supported, firstly, by the generally very high posterior
470 probabilities with which individuals were assigned within Priest and Paxton lakes. Levels of potential
471 mis-assignment between hybrid and parental groups were low, with fewer than four percent of
472 individuals from Priest Lake being assigned with less than 0.95 probability, and fewer than nine
473 percent in Paxton Lake. In addition, these potential mis-assignments were evenly distributed amongst
474 the parent and hybrid groups (three and five from parent versus hybrid groups in Priest Lake species
475 pair, and nine versus eight for the same groupings in Paxton Lake). Secondly, this program assigned
476 samples of known origin with generally very high probabilities during simulations, and was found to be
477 robust to violation of the assumption of HWE prior to hybridization (which was approximated by
478 testing for HWE in parental data subsets). There were very few instances of parents being mis-
479 assigned as hybrids (less than one percent) but a higher rate of the opposite scenario, with 33 second
480 generation hybrids being wrongly assigned to a parental category (8 % of all simulated hybrid
481 assignments). Therefore, any deviation of hybridization rate estimates from the true value is likely to
482 be an underestimate.
483
484 Generally, lower posterior probabilities were obtained for Enos Lake individuals compared to Paxton
485 or Priest lake fish. Although the lack of baseline samples (ethanol-preserved material prior to extensive
486 introgression) makes it difficult to verify, this is likely the outcome of introgression that has occurred
487 beyond a second generation of hybrids (Taylor *et al*. in press) reducing the assignment power of
488 NewHybrids. As the number of generations over which introgression has been occurring, the number

489 of possible genotype frequency classes to which an individual may belong increases exponentially and
490 distinguishing becomes increasingly difficult, with a prohibitive amount of data required (Boecklen &
491 Howard 1997; Anderson & Thompson 2002).
492
493 Our estimates of long term gene flow support this idea of gene exchange between benthics and
494 limnetics in each species pair. Given that the markers used accentuate differences between species
495 relative to the average neutral marker, these estimates may underestimate true levels of gene flow.
496 Nevertheless, the relative comparison between lakes gives insight into patterns of gene flow. In
497 agreement with the estimates of hybridization rates, gene flow was highest within Enos Lake, although
498 a long-term migration rate of 0.3 % is eighty times lower than the current hybridization rate estimate of
499 24 %. Similarly, long term gene flow estimates within Priest and Paxton lakes are over an order of
500 magnitude less (61 and 29 times lower, respectively) than current hybridization estimates. With the
501 lowest levels of gene flow recorded in Priest Lake, the Paxton Lake pair has a value intermediate
502 between the other two species pairs.
503
504 *Evolutionary implications*
505
506 The contrasting scenarios found between species pairs raise the question as to what processes are
507 controlling the rates of hybridization and gene flow. The findings within Priest and Paxton lake species
508 pairs support experimental evidence that strong assortative mating is playing a significant role in
509 limiting hybridization (Ridgway & McPhail 1984; Nagel & Schluter 1998). Our demonstration of a
510 low background level of hybridization, however, shows that this pre-mating reproductive isolation is
511 incomplete.
512
513 Despite this hybridization, however, there is no evidence of extensive introgression of these hybrids
514 within Priest and Paxton lakes. Indeed, there is over an order of magnitude of discrepancy between the
515 hybridization rate estimates, which reflect recent gene flow, and the lower long term estimates of gene
516 flow. The long term gene flow estimates may underestimate true levels to a certain degree because
517 biased markers that accentuated species differences were used. Nevertheless, congruence in the
518 magnitude of total $N_e m$ estimates between the MIGRATE results from this study (1.67, 7.337 and 3.618
519 for Priest, Paxton and Enos lake species pairs, respectively) and those calculated by applying Wright's
520 (1931) infinite island model ($F_{ST} = 1/[1 + 4N_e m]$) to previous F_{ST} estimates (Taylor & McPhail 2000;
521 1.892, 1.847 and 1.892 for Priest, Paxton and Enos lake species pairs, respectively) supports our
522 conclusion that levels of long term gene flow are much reduced compared to current hybridization
523 rates. The F_{ST} estimates derived from a population genetic survey of six (seemingly) neutral
524 microsatellites that conform to Hardy Weinberg expectations (Taylor & McPhail 2000) represent the
525 only other gene flow estimates for the species pairs to date.
526
527 This trend of lower longer term gene flow supports experimental work suggesting that selection is
528 acting against hybrids. Field enclosure experiments found that reduced F_1 hybrid growth rates in both
529 parental habitats is likely due to reduced foraging efficiency (Schluter 1995; Hatfield & Schluter 1999).
530 The ecological basis of this post-zygotic isolation was confirmed by a similar experiment using
531 backcrosses which controlled for intrinsic factors (Rundle 2002). Sexual selection against hybrid males
532 has also been implicated by the reduced mating success of F_1 hybrid males in their preferred nesting
533 habitat compared to limnetics, the parental species sharing the same nesting preference (Vamosi &

534 Schluter 1999). Collectively, this work supports the view that post-zygotic reproductive isolation via
535 reduced hybrid fitness is important in maintaining distinct gene pools in sympatry. Indeed, this has
536 been suggested by experimental work supporting the role of reinforcement in the divergence of the
537 species pairs (Rundle & Schluter 1998).
538
539 The detection of only F_2 hybrids and backcrosses within each species pair is unexpected, given that F_1
540 hybrids are, of course, essential to their production. It is feasible that F_1 hybrids may occur at lower
541 frequency relative to other, more abundant hybrid categories, and so have evaded detection in this
542 survey. Although there is no evidence to support intrinsic selection against any hybrid class (with the
543 possible exception of benthic backcrosses [McPhail 1984, 1992; Hatfield & Schluter 1999]), the pattern
544 of hybrid abundance observed here could reflect extrinsic selection directed primarily against F_1
545 hybrids. Indeed, if divergent selection against hybrids plays an important role in maintaining species
546 integrity, then phenotypically intermediate individuals (F_1) would be expected to fair worse than those
547 hybrids that are more parental-like (second generation hybrids; Hatfield 1997); the latter would be
548 better able to exploit parental niches. Indeed, reciprocal field enclosure experiments show just this,
549 with F_1 hybrids showing a significant growth disadvantage relative to the parent adapted to the
550 environment (Hatfield & Schluter 1999), whilst neither backcross differed significantly from the parent
551 from which it was mainly derived (Rundle 2002). Although mis-assignment of hybrid category cannot
552 be excluded, the NewHybrids simulations consistently assigned F_1 hybrids correctly with high
553 probabilities, corroborating this hybrid assignment pattern as a real biological phenomenon. This view
554 is also supported by a review of hybrid fitness which consistently found significant variation in the
555 relative fitness of hybrid classes (Arnold & Hodges 1995). Indeed, some experimental studies have
556 directly demonstrated lower F_1 hybrid viability relative to other hybrid classes (e.g. Reed & Site 1995).
557 Furthermore, indirect evidence of this comes from other empirical studies that have also observed a low
558 level of F_1 hybrids relative to post-F_1 hybrid categories in natural hybrid populations (e.g. Arnold 1994;
559 Redenbach and Taylor 2003; Ostberg *et al.* 2004).
560
561 The forces driving the demise of natural hybrids in Paxton and Priest Lakes are still unknown. The
562 species diagnostic molecular profile that we have described here can now be used to tackle this
563 question. For the first time, these genetic tools will enable us to assess hybridization rates temporally
564 across the various life-history stages of the stickleback, overcoming limitations of morphological
565 methods to distinguish between immature benthics and limnetics. A decrease in the relative number of
566 hybrids throughout the stickleback life-cycle would provide compelling evidence for ecological
567 selection against hybrids playing a significant role in post-zygotic isolation, while a consistent
568 proportion of sexually mature hybrids would support a role for sexual selection against hybrids in
569 reinforcement. Whilst not defining precise mechanisms of selection against hybrids, assessing their
570 existence and relative contributions within natural populations would give valuable insight into their
571 role in speciation, an area where empirical tests are lacking (Rundle & Nosil 2005).
572
573 Whatever the mechanism of selection, the prerequisite conditions for it have now been altered in Enos
574 Lake, such that postzygotic isolating mechanisms are no longer effective at maintaining divergence
575 between the species. The balance between hybridization and selection has tilted towards increased
576 levels of gene flow, resulting in a breakdown of the species pair into a hybrid swarm. The collapse of
577 reproductive isolating mechanisms is likely due to environmental change within the lake. An account
578 of the possibilities accompanies a full description of the demise of this species pair elsewhere (Taylor

579 *et al*. in press). This breakdown has been asymmetrical, with biased introgression of the limnetic form
580 into that of the benthic. The cause of this directionality awaits investigation i.e. do limnetic females
581 now mate preferentially with benthic or hybrid males, or are limnetic males managing to mate
582 successfully with benthic or hybrid females? Patterns of mtDNA inheritance in hybrid lines have
583 tested for directionality in other fish from this region (Redenbach and Taylor 2003; Ostberg *et al*. 2004;
584 Bettles *et al*. in press). Unfortunately, the utility of this marker for this purpose in the evolutionarily
585 young benthic-limnetic system is limited by a lack of clear distinction between the mtDNA of the two
586 forms within each lake (Taylor & McPhail 1999; JL Gow unpublished data), caused either by historical
587 introgression or incomplete lineage sorting. Circumstantial evidence suggesting that limnetic females
588 mating preferentially with benthic or hybrid males may be the predominant mode of hybridization
589 includes: the known role of visual cues in stickleback mate choice (Ridgway & McPhail 1984); female
590 perceptual sensitivity to red light diverging according to habitat differences in light environment, and
591 male nuptial colour being tuned to this female perceptual sensitivity (Boughman 2001); and suspected
592 (although unsubstantiated) increased turbidity in Enos Lake. Clarification of the mechanism behind the
593 observed directionality may be best addressed by mating trials conducted under the altered
594 environmental conditions.
595
596 The breakdown of the species pair within Enos Lake highlights the delicate balance between
597 persistence and breakdown of reproductive barriers between young species, where pre-zygotic isolation
598 and extrinsic post-zygotic isolation are typically thought to evolve before intrinsic post-zygotic
599 isolation (Coyne & Orr 2004). This collapse via elevated hybridization following anthropologically-
600 induced environmental change in lake conditions (Taylor *et al*. in press) highlights a serious threat to
601 freshwater fish faunas (e.g. Seehausen *et al*. 1997; Bettles *et al*. in press). The demise of the benthic-
602 limnetic species pair in Enos Lake is not the first recorded; the Hadley Lake species pair on Lasqueti
603 Island became extinct following human habitat disturbance. In this instance, the fish were
604 exterminated by an introduced non-native catfish, *Ameiurus nebulosus* (Hatfield 2001). Intriguingly,
605 there is also some evidence of historical introgression within Paxton Lake species pair: the lower level
606 of long-term gene flow but the same level of hybridization estimated in this species pair compared to
607 that of Priest Lake could be indicative of a recovery from historical introgression. Such a phenomenon
608 would correlate with a known history of human-induced, major environmental change. For over 20
609 years, from the late 1950's to the late 1970's, the lake level varied greatly due to annual draw down for
610 quarry-mining purposes (Larson 1976; McPhail 1992). Furthermore, five thousand coho salmon
611 (*Oncorhynchus kisutch*) were introduced to the lake during this period and became significant
612 stickleback predators before their extinction five years later (Larson 1976; McPhail 1992). The higher
613 proportion of intermediate morphological forms during, rather than after, this period of disturbance
614 (McPhail 1992) supports the hypothesis that such environmental change may have triggered increased
615 hybridization rates.
616
617 If this was indeed the case, then the subsequent recovery of the species pair after the termination of the
618 human disturbances should be heartening to conservation efforts for the Enos Lake species pair.
619 Indeed, given the sensitivity of these highly endemic young species to environmental change, as well as
620 their importance as scientific models for the study of adaptive divergence and speciation (see Coyne &
621 Orr 2004; Rundle & Nosil 2005), the benthic-limnetic species pairs are now listed as endangered. The
622 species diagnostic molecular profile developed here is now serving as a tool to monitor the population
623 status of these endangered species (COSEWIC 2004). They can also aid in more proactive

624 conservation management by helping to select the most "benthic-like" and "limnetic-like" Enos Lake
625 fish for use in captive breeding programs designed to artificially produce offspring for use in possible
626 re-introductions.
627
628 In conclusion, species diagnostic marker profiles can be applied to answer an array of biological
629 questions in populations of this geographically widespread and phenotypically diverse model organism.
630 Given that the approach we used successfully identified diagnostic markers for evolutionarily young
631 species (about 13 000 years old), a similar methodology would likely uncover discriminatory markers
632 useful to assessing gene flow in other species pairs of sticklebacks which have diverged over longer
633 periods, such as the sympatric Japan Sea system, or over a similar time frame, including parapatric
634 freshwater-anadromous and lake-stream systems (McKinnon & Rundle 2002; Hendry & Taylor 2004).
635 Such tools could also be beneficial to experimental approaches exploring ways in which selection, gene
636 flow and adaptive divergence interact in natural populations over multiple generations.
637

References

637   **References**

638

639   Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using

640   multilocus genetic data. *Genetics*, **160**, 1217-1229.

641

642   Arnold ML (1994) Natural hybridization and Louisiana Irises. *Biosciences*, **44**, 141-147.

643

644   Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in*

645   *Ecology and Evolution*, **10**, 67-71.

646

647   Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*,

648   **16**, 113-148.

649

650   Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective

651   population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763-773.

652

653   Belkhir K, Borsa P, Chikhi N, Raufaste N, Bonhomme F (2001) GENETIX 4.03, logiciel sous

654   Windows TM pour la génétique des populations. *in*. Laboratoire Génome, Populations, Interactions,

655   CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

656

657   Bell MA, Aguirre WE, Buck NJ (2004) Twelve years of contemporary armor evolution in a threespine

658   stickleback population. *Evolution*, **58**, 814-824.

659

660   Bell MA, Foster SA (1994) *The evolutionary biology of the threespine stickleback.* Oxford University

661   Press, Oxford.

662

663   Bentzen P, McPhail JD (1984) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*):

664   specialization for alternative trophic niches in the Enos Lake species pair. *Canadian Journal of*

665   *Zoology*, **62**, 2280-2286.

666

667   Bentzen P, Ridgway MS, McPhail JD (1984) Ecology and evolution of sympatric sticklebacks

668   (*Gasterosteus*): spatial segregation and seasonal habitat shifts in the Enos Lake species pair. *Canadian*

669   *Journal of Zoology*, **62**, 2436-2439.

670

671   Bettles CM, Docker MF, Dufour B, Heath DD. Hybridization dynamics between sympatric species of

672   trout: loss of reproductive isolation. *Journal of Evolutionary Biology,* in press.

673

674   Boecklen WJ, Howard DJ (1997) Genetic analysis of hybrid zones: numbers of markers and power of

675   resolution. *Ecology*, **78**, 2611-2616.

676

677   Boughman JW (2001) Divergent sexual selection enhances reproductive isolation in sticklebacks.

678   *Nature*, **411**, 944-948.

679

680   Colosimo PF, Hosemann KE, Balabhadra S *et al*. (2005) Widespread parallel evolution in sticklebacks

681   by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928-1933.

682

683   Colosimo PF, Peichel CL, Nereng K *et al*. (2004) The genetic architecture of parallel armor plate

684   reduction in threespine sticklebacks. *PLoS Biology*, **2**, 635-641.

685

686   Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus

687   genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989-2000.

688

689   COSEWIC (2004) *Canadian species at risk*. Committee on the Status of Endangered Wildlife in

690   Canada. Available from COSEWIC Secretariat, Canadian Wildlife Service, Ottawa, Ontario, Canada.,

691   K1A 0H3.

692

693   Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc. Sunderland, Mass.

694

695   Cresko WA, Amores A, Wilson C *et al*. (2004). Parallel genetic basis for repeated evolution of armor

696   loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences of*

697   *the USA*, **101**, 6050–6055.

698

699   Day T, Pritchard J, Schluter D (1994) Ecology and genetics of phenotypic plasticity: a comparison of

700   two sticklebacks. *Evolution*, **48**, 1723-1734.

701

702   Feldman MS, Kumm J, Pritchards J (1999) Mutation and migration in models of microsatellite

703   evolution. In: *Microsatellites: evolution and applications* (eds. Goldstein DB, Schlötterer C), pp. 98-

704   115. Oxford University Press, Oxford.

705

706   Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of*

707   *Heredity*, **86**, 485-486.

708

709   Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version

710   2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

711

712   Hatfield T (1997) Genetic divergence in adaptive characters between sympatric species of stickleback.

713   *The American Naturalist*, **149**, 1009–1029.

714

715   Hatfield T (2001) Status of the stickleback species pair, *Gasterosteus* spp., in Hadley Lake, Lasqueti

716   Island, British Columbia. *Canadian Field Naturalist*, **115**, 579-583.

717

718   Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: Environment dependent hybrid

719   fitness. *Evolution* **53**:866–873.

720

721   Page: 20

722   Hendry A.P, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by

723   gene flow? An evaluation using lake-stream stickleback pairs. *Evolution*, **58**, 2319-2331.

724

725   Kingsley DM, Zhu BL, Osoegawa K *et al*. (2004) New genomic tools for molecular studies of

726   evolutionary change in threespine sticklebacks. *Behaviour*, **141**, 1331-1334.

727

728   Kraak SBM, Mundwiler B, Hart PJB (2001) Increased number of hybrids between benthic and limnetic

729   three-spined sticklebacks in Enos Lake, Canada: the collapse of a species pair? *Journal of Fish*

730   *Biology*, **58**, 1458-1464.

731

732   Larson GL (1976) Social behaviour and feeding ability of two phenotypes of *Gasterosteus aculeatus* in

733   relation to their spatial and trophic segregation in a temperate lake. *Canadian Journal of Zoology*, **54**,

734   107-121.

735

736   Lavin PA, McPhail JD (1985) The evolution of freshwater diversity in the threespine stickleback,

737   *Gasterosteus aculeatus*: site-specific differentiation of trophic morphology. *Canadian Journal of*

738   *Zoology*, **63**, 2632-2638.

739

740   Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**,

741   23-35.

742

743   McKinnon JS, Mori S, Blackman BK *et al*. (2004) Evidence for ecology's role in speciation. *Nature*,

744   **429**, 294-298.

745

746   McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems.

747   *Trends in Ecology & Evolution*, **17**, 480- 488.

748

749   McPhail, JD (1984) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): morphological

750   and genetic evidence for a species pair in Enos Lake, British Columbia. *Canadian Journal of Zoology*,

751   **62**, 1402-1408.

752

753   McPhail JD (1992) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): evidence for a

754   species pair in Paxton Lake, Texada Island British Columbia. *Canadian Journal of Zoology*, **70**, 361-

755   369.

756

757   McPhail JD (1993) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): origin and

758   evolution of the sympatric pairs. *Canadian Journal of Zoology*, **71**, 515-523.

759

760   McPhail JD (1994) Speciation and the evolution of reproductive isolation in the sticklebacks

761   (Gasterosteus) of south-western British Columbia. In: *The evolutionary biology of the threespine*

762   *stickleback* (eds. Bell MA, Foster SA), pp. 399-437. Oxford University Press, Oxford.

763

764   Nagel L, Schluter D (1998). Body size, natural selection, and speciation in sticklebacks.

765   *Evolution*, **52**, 209–218.

766

767   Nielsen EE, Hansen MM, Bach LA (2001) Looking for a needle in a haystack: discovery of indigenous

768   Atlantic salmon (*Salmo salar* L.) in stocked populations. *Conservation Genetics*, **2**, 219-232.

769

770 Ostberg CO, Slatton SL, Rodriguez RJ (2004) Spatial partitioning and asymmetric hybridization among

771 sympatric coastal steelhead trout (*Oncorhynchus mykiss irideus*), coastal cutthroat trout (*O. clarki*

772 *clarki*) and interdpecific hybrids. *Molecular Ecology*, **13**, 2773-2788.

773

774 Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time

775 estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular*

776 *Ecology*, **13**, 55-65.

777

778 Peichel CL, Nereng KS, Ohgi KA *et al.* (2001) The genetic architecture of divergence between

779 threespine stickleback species. *Nature*, **414**, 901– 905.

780

781 Piry S, Alapetite A, Cornuet J-M *et al.* (2004) GeneClass2: A Software for Genetic Assignment and

782 First-Generation Migrant Detection. *Journal of Heredity*, **95**, 536-539.

783

784 Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of*

785 *the National Academy of Sciences of the USA*, **94**, 9197-9201.

786

787 Raymound M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests

788 and ecumenicism. *Journal of Heredity*, **86**, 248-249.

789

790 Raymound M, Rousset F (2001) GENEPOP (version 3.3) Available from http://www.cefe.cnrs-

791 mop.fr/. Updated from Raymound & Rousset (1995).

792

793   Redenbach Z, Taylor EB (2003) Evidence for bimodal hybrid zones between two species of char

794   (*Salvelinus*) in northwestern North America. *Journal of Evolutionary Biology*, **16**, 1135-1148.

795

796   Reed KM, Sites JW (1995) Female fecundity in a hybrid zone between two chromosome races of the

797   *Sceloporus grammicus* complex (Sauria, phrynosomatidae). *Evolution*, **49**, 61-69.

798

799   Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

800

801   Rico C, Zadworny D, Kuhnlein U, Fitzgerald GJ (1993) Characterization of hypervariable

802   microsatellite loci in the threespine stickleback *Gasterosteus aculeatus*. *Molecular Ecology*, **2**, 271-

803   272.

804

805   Ridgway MS, McPhail JD (1984) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*):

806   mate choice and reproductive isolation in the Enos Lake species pair. *Canadian Journal of Zoology*, **62**,

807   1813-1818.

808

809   Rubridge EM, Taylor EB (2004) Hybrid zone structure and the potential role of selection in hybridizing

810   populations of native westslope cutthroat trout (*Oncorhynchus clarki lewisi*) and introduced rainbow

811   trout (*O. mykiss*). *Molecular Ecology*, **??**, ???? - ?????

812

813   Rundle HD (2002) A test of ecologically dependent postmating isolation between sympatric

814   sticklebacks. *Evolution*, **56**, 322-329.

815

816   Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336-352.

817

818   Rundle HD, Schluter D (1998) Reinforcement of stickleback mate preferences: Sympatry breeds

819   contempt. *Evolution*, **52**, 200–208.

820

821   Rundle HD, Vamosi SM, Schluter D (2003) Experimental evidence of predation's effect on divergent

822   selection during character displacement in sticklebacks. *Proceedings of the National Academy of*

823   *Sciences of the USA*, **100**, 14943-14948.

824

825   Sanderson SL, Cech JJ, Patterson MR (1991) Fluid dynamics in suspension-feeding blackfish. *Science*,

826   **252**, 1346-1348.

827

828   Schlötterer C (2003) Hitchhiking mapping – functional genomics from the population genetics

829   perspective. *Trends in Genetics*, **19**, 32-38.

830

831   Schluter D (1993) Adaptive radiation in sticklebacks: Size, shape, and habitat use efficiency. *Ecology*,

832   **74**, 699–709.

833

834   Schluter D (1994) Experimental evidence that competition promotes divergence in adaptive radiation.

835   *Science*, **266**, 798–800.

836

837   Schluter D (1995) Adaptive radiation in sticklebacks: Trade-offs in feeding performance

838   and growth. *Ecology*, **76**, 82–90.

839

840   Schluter D (2003) Frequency dependent natural selection during character displacement in

841   sticklebacks. *Evolution*, **57**, 1142-1150.

842

843   Schluter D, Clifford EA, Nemethy M, McKinnon JS (2004) Parallel evolution and inheritance of

844   quantitative traits. *The American Naturalist*, **163**, 809-822.

845

846   Schluter D, McPhail JD (1992) Ecological character displacement and speciation in sticklebacks. *The*

847   *American Naturalist*, **140**, 85–108.

848

849   Schluter D, Nagel LM (1995) Parallel speciation by natural selection. *The American Naturalist*, **146**,

850   292-301.

851

852   Seehausen O, van Alphen JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that

853   curbs sexual selection. *Science*, **277**, 1808-1811.

854

855   Shapiro MD, Marks ME, Peichel CL *et al*. (2004) Genetic and developmental basis of evolutionary

856   pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.

857

858   Shimoda N, Knapik EW, Ziniti J *et al*. (1999) Zebrafish genetic map with 2000 microsatellite markers.

859   *Genomics*, **58**, 219-232.

860

861   Taylor EB (1998) Microsatellites isolated from the threespine stickleback *Gasterosteus aculeatus*.

862   *Molecular Ecology*, **7**, 930-931.

863

864   Taylor EB, Groenenboom M, Sniatynski M *et al*. In press. Evolution in reverse: morphological and

865   genetic evidence of the collapse of a threespine stickleback (*Gasterosteus aculeatus*) species pair.

866   *Molecular Ecology*.

867

868   Taylor EB, McPhail JD (1999) Evolutionary history of an adaptive radiation in species

869   pairs of threespine sticklebacks (*Gasterosteus*): Insights from mitochondrial DNA. *Biological Journal*

870   *of the Linnean Society*, **66**, 271–291.

871

872   Taylor EB, McPhail JD (2000) Historical contingency and ecological determinism interact to prime

873   speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society of London B*, **267**, 2375–

874   2384.

875

876   Vamosi SM, Schluter D (1999) Sexual selection against hybrids between sympatric sticklebacks:

877   Evidence from a field experiment. *Evolution*, **53**, 874–880.

878

879   Vamosi SM, Schluter D (2004) Character shifts in the defensive armor of sympatric sticklebacks.

880   *Evolution*, **58**, 376-385.

881

882   Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure.

883   *Evolution*, **38**, 1358-1370.

884

885   Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$.

886   *Heredity*, **82**, 17-125.

887

888   Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97-159.

889

Author Information Box

With strong interests in the role of gene flow in ecological speciation, Jennifer Gow has been investigating hybridization and introgression in threespine stickleback species pairs as part of her postdoctoral studies. Catherine Peichel is broadly interested in the genetic basis of traits that underlie reproductive isolation in threespine sticklebacks. Eric Taylor has strong interests in the evolution and conservation of native fishes, and employs molecular and ecological methods in studies of the origins and persistence of biodiversity.

909 **Figure 1** Two-dimensional multivariate scatter plots of nine morphological traits from 96 threespine
910 stickleback from (**a**) Paxton Lake and (**b**) Priest Lake. The squared correlation in distances (r^2)
911 indicates the proportion of variance of the data that is accounted for by the corresponding distances in
912 the figure.
913
914 **Figure 2** Categorization of threespine sticklebacks sampled from Priest (n = 198), Paxton (n = 192) and
915 Enos lakes (n = 192) in 2003 using a model-based Bayesian method implemented by NewHybrids
916 (Anderson & Thompson 2002).
917
918 **Figure 3** Two-dimensional Factorial Correspondence Analysis illustrating relationships among the
919 multilocus genotypes of individual threespine sticklebacks from (**a**) Priest, (**b**) Paxton and (**c**) Enos
920 lakes. I and II are the first and second principal factors of variability, respectively. Large circles
921 encompass individuals categorised as 'pure' benthic or limnetic by NewHybrids, and filled symbols
922 represent hybrids.
923

923 **Table 1** Suite of microsatellites showing non-overlapping allele ranges between benthics and limnetics
924 in at least one of the species pairs. The sample sizes for the species pair to which the allele range refers
925 are 48 for Priest and Paxton benthics and limnetics, 25 for Enos Lake benthics and 26 for Enos Lake
926 limnetics. Number of discrepancies refers to the number of individuals from the screening panel ('L'
927 suffix for limnetic, 'B' for benthic) which carry an allele from the other species range.
928

Locus	Reference or GenBank accession number	Species pair	Allele range (base pairs):		Number of discrepancies
			benthics	limnetics	
*Stn*388	BV678141	Paxton	185	199 – 215	4L
*Stn*295	BV678106	Paxton	151	163 – 185	2L
*Stn*142	Peichel *et al*. 2001	Paxton	199 – 219	179 – 187	1B; 1L
*Stn*383	BV212282	Paxton	192 – 208	178 – 182	1B; 3L
*Stn*387	BV678140	Priest	205 – 235	165 – 175	1B
		Enos	201 – 239	165 – 173	10L
*Stn*254	BV678079	Priest	249 – 279	225 – 227	
		Enos	255 – 275	225 – 227	1B; 9L
*Stn*216	Colosimo *et al*. 2004	Enos	195 – 209	177	1B; 7L
*Stn*386	BV678139	Enos	210 – 223	233 – 241	1B; 10L
*Stn*43	Peichel *et al*. 2001	Enos	148 – 166	132 – 136	9L

929

Table 2 Number of mis-assigned individuals during NewHybrids simulations. Categories of samples of known origin include: limnetic (L) or benthic (B) parent, F_1, F_2, and limnetic (LBx) or benthic (BBx) backcross from Priest (Pr) and Paxton (Pa) lakes. Total sample sizes of each category run during five repeated simulations for each species pair included in parenthesis.

Sample origin	Number & category of mis-assignments for species pair:	
	Priest	Paxton
Parents: B (460 Pr, 455 Pa) & L (465 Pr, 370 Pa)	3 BBx	6 BBx
F_1(50)	0	0
F_2(50)	15 BBx	8 BBx 2 F_1
LBx (50) & BBx (50)	6 L 9 B 3 F_1 1 F_2	15 L 3 B 10 F_1

Table 3 Long term gene flow estimates between benthics (B) and limnetics (L) in Priest and Paxton lakes, and between benthics and hybrids (H) in Enos Lake. Population estimates calculated using a maximum-likelihood coalescent method implemented by MIGRATE (Beerli & Felsenstein 1999) are summed to estimate the total effective population size (N_e); the total proportion of migrants (m), highlighted in bold print, is calculated by dividing the sum of the two N_em estimates by the total effective population size. Approximate 95 % confidence intervals are given in parenthesis for population estimates of N_e and m.

Lake	N_e			m		
	L or H	B	Total	L or H	B	Total (N_em[L or H]+ N_em[B] / N_e[Total])
Priest	915 (873 - 930)	1334 (1272 - 1356)	2249	0.00092 (0.00083 - 0.00095)	0.00062 (0.00056 - 0.00064)	0.84182 + 0.82799 / 2249 = **0.00074**
Paxton	908 (864 -924)	1160 (1110 - 1178)	2068	0.00205 (0.00192 - 0.00209)	0.00156 (0.00147 - 0.00159)	1.86092 + 1.80840 / 2068 = **0.00177**
Enos	926 (868 - 947)	1342 (1295 - 1358)	2268	0.00417 (0.00398 - 0.00424)	0.00259 (0.00247 - 0.00263)	3.86246 + 3.47533 / 2268 = **0.00324**