Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers

C. S. Wilding, R. K. Butlin and J. Grahame*

Centre for Biodiversity and Conservation, School of Biology, The University of Leeds, Leeds LS2 9JT

*Author for correspondence: Tel. 0113 233 2852, FAX 0113 233 2835, Email j.w.grahame@leeds.ac.uk

Word count: 6066 including tables and references.

Running head: AFLP markers of differential gene exchange

Key words: speciation, disruptive selection, clines, AFLP, *Littorina*.

1   **Abstract**

2   Speciation requires the acquisition of reproductive isolation, and the circumstances

3   under which this could evolve are of great interest. Are new species formed after the

4   acquisition of generalized incompatibility arising between physically separated

5   populations, or may they arise as a result of the action of disruptive selection

6   beginning with the divergence of a rather restricted set of gene loci? Here we apply

7   the technique of Amplified Fragment Length Polymorphism (AFLP) analysis to an

8   intertidal snail whose populations display a cline in shell shape across vertical

9   gradients on rocky shores. We compare the F_{ST} values for 306 AFLP loci with the

10 distribution of F_{ST} estimated from a simulation model using values of mutation and

11 migration derived from the data. We find that about 5% of these loci show greater

12 differentiation than expected, providing evidence of the effects of selection across the

13 cline, either direct or indirect through linkage. This is consistent with expectations

14 from non-allopatric speciation models that propose an initial divergence of a small

15 part of the genome driven by strong disruptive selection while divergence at other loci

16 is prevented by gene flow. However, the pattern could also be the result of differential

17 introgression after secondary contact.

18

Introduction

The process of speciation requires the acquisition of reproductive isolation. If populations are separated by a physical barrier to dispersal, speciation may follow: the acquisition of intrinsic reproductive isolation is then an incidental consequence of the accumulation of genetic differentiation (Mayr, 1963). Increasingly, attention has shifted to the possibility that reproductive barriers might arise in populations not separated by major physical features (Bush & Howard, 1986), i.e. that speciation might begin with genetic diversification in spite of some gene exchange between constituent populations. Empirical evidence shows, for example, that a single founder population in a lake may diversify and undergo speciation following use of different niches (Schliewen *et al.*, 1994; Schluter, 1996; Wilson *et al.*, 2000), and theoretical work suggests that gene flow can be less of a cohesive force than previously thought (Barton, 1988).

Barton (1988) and Rice & Hostert (1993) have reviewed the literature on speciation mechanisms, showing that there are plausible and simple models of non-allopatric speciation. In these models, genetic divergence may be initiated by disruptive selection without a period of extrinsic isolation. This requires strong selection and either pleiotropy or linkage of the genes involved in the adaptive polymorphism with those affecting the probability of gene exchange. For parapatric populations, where gene exchange is restricted, an initial level of differentiation may be modified to increase isolation by the accumulation of different alleles in the diverging genetic backgrounds. Strong selection also is needed here if gene flow is other than negligible. Nevertheless, Rice & Hostert (1993) concluded that laboratory experiments on the development of isolation strongly support the idea that

43 reproductive isolation can evolve between sympatric or parapatric populations if

44 divergent selection is strong relative to gene flow.

45   Although the conclusion of Rice & Hostert (1993) is well supported by laboratory

46 experiments, there is less evidence from natural populations. Host races provide the

47 best examples, especially *Rhagoletis* (Feder *et al.*, 1994; Feder *et al.*, 1997). Host

48 fidelity provides the major barrier to gene exchange, permitting further differentiation

49 under selection on the alternative hosts. Some markers (presumably those linked to

50 selected loci, or perhaps under selection themselves) show allele frequency

51 differentiation, while others do not, suggesting that gene exchange is more restricted

52 in some parts of the genome than others. This may be viewed as a signature of non-

53 allopatric speciation and is in contrast to the generalised barrier to gene flow that

54 results from physical isolation. The uniform divergence across the genome that

55 evolves in allopatry may be maintained following secondary contact due to the

56 accumulation of genetic incompatibility at many loci that is revealed in some hybrid

57 zones (Barton & Hewitt, 1981; Szymura & Barton, 1991). However, it may be eroded

58 by introgression.

59   We address the issue of uniform versus restricted differentiation using a system

60 where divergent populations are parapatric. They are likely to be exchanging genes

61 only in the region of contact, and the selection gradient on which they exist is

62 imposed by the physical environment and by predation. *Littorina saxatilis* (Olivi) (the

63 'rough periwinkle') is widespread on North Atlantic shores, exhibits high

64 morphological and allozyme variability, and is ovoviviparous and of low vagility - see

65 Reid (1996) for review. In Britain it is found as two morphological forms ('H' and 'M')

66 (Hull *et al.*, 1996) that show good evidence of partial reproductive isolation. This

67 interpretation was based on reduced fertility in females inferred to be hybrids, and is

68   supported by the observation of assortative mating (Hull, 1998; Pickles & Grahame,

69   1999). The observed differentiation could be attributed to secondary contact between

70   populations that had been undergoing allopatric divergence. Alternatively we may be

71   seeing divergence *in situ* due to strong selection, despite gene flow (Endler, 1977;

72   Rice & Hostert, 1993). In either case, the current pattern of differentiation is probably

73   maintained by a balance between gene flow and selection, where the selection is due,

74   at least in part, to environmental pressures rather than genetic incompatibility.

75   Predation by crabs is thought to exert strong selection on periwinkle shell form

76   (Heller, 1976; Raffaelli, 1978; Janson, 1983; Johannesson, 1986), and among

77   molluscs more widely - see Vermeij (1987) for review. Both thickness and form of the

78   periwinkle shell may vary adaptively in response to differing predation pressures, and

79   inducible phenotypic responses are considered to be involved for thickness changes in

80   at least some species (Trussell & Smith, 2000). However, there is abundant evidence

81   that in *L. saxatilis* some of the variation is genotypic (Newkirk & Doyle, 1975;

82   Grahame & Mill, 1993; Johannesson & Johannesson, 1996), and this is especially

83   likely for shell shape. Because crab predation increases down the shore in most sites,

84   clines in shell shape are often found (Grahame *et al.*, 1997). In the upper shore,

85   *L. saxatilis* H are thin-shelled, wide-apertured animals with relatively low spires. This

86   shape may come about simply as a result of the constraints on shell shape when the

87   aperture is large (Clarke *et al.*, 1999) thus affording greater foot area (Grahame &

88   Mill, 1986) for adhesion and leading to greater gravitational stability (Heller, 1976).

89   Therefore, this is probably the optimum shape for maintaining a grip on wave or

90   wind-affected substrates in the absence of crab predation. In the lower shore

91   *L. saxatilis* M are thicker shelled, with relatively smaller apertures; these features are

92   likely to be adaptive in reducing the risk of crab predation (Johannesson, 1986;

93   Boulding *et al.*, 1999).

94   Primary and secondary origins of clines are notoriously difficult to distinguish

95   (Barton & Hewitt, 1985). Wilding *et al.* (2000) considered it probable that the current

96   distribution of mitochondrial haplotypes in *L. saxatilis* in the British Isles indicated

97   expansion from different glacial refugia. However, the distribution of the H and M

98   forms is quite different from that described for these haplotypes (Wilding *et al.*,

99   2000), and Wilding *et al.* (2001) concluded that the current haplotype distribution was

100   unrelated to whether populations were H or M morph. We tentatively suggest that the

101   *L. saxatilis* H-M cline has evolved *in situ*.

102   Here we examine putative loci (hereafter, simply 'loci') revealed by the Amplified

103   Fragment Length Polymorphism technique (AFLP) (Vos *et al.*, 1995) in samples from

104   four locations on the coast of Yorkshire, England. We compare observed F_{ST}

105   distributions across loci between populations of *L. saxatilis* H and M with F_{ST}

106   distributions in within-morph comparisons, and with expected distributions. These

107   expected distributions were derived from simulations of F_{ST} values in the absence of

108   selection, using an approach analogous to that of Beaumont & Nichols (1996). We ask

109   whether the barrier to gene exchange between H and M populations is uniformly

110   effective across loci.

Materials and methods

Sampling

Periwinkles were collected from rocky shores at Thornwick Bay, Flamborough

(British Grid reference TA 233724), Filey Brigg (TA 132815), Old Peak (NZ 982024)

and Robin Hood's Bay (NZ 955055). The coast trends overall northwesterly in this

region, the straight line distances between the sites are: Flamborough - Filey Brigg, 15

km (we estimate that 60% of the intervening shore represents suitable habitat for

L. saxatilis); Filey Brigg - Old Peak, 26 km (80% suitable habitat); Old Peak - Robin

Hood's Bay, 4 km (90% suitable habitat). At each site snails were collected from each

of two locations (one in an area occupied by the H morph and one in an area occupied

by the M morph, except at Robin Hood's Bay), individual snails were taken from an

area of about 2 m^2. H and M animals were characterized on the basis of sample

location and shell form (by eye), and only brooding females were used to avoid

contaminating the H samples with specimens of *Littorina arcana* Hannaford Ellis

(which lay eggs on the shore). Sampling locations were 5 m apart at Flamborough, 15

m apart at Filey, 300 m apart at Old Peak, and 75 m apart at Robin Hood's Bay. In the

first three instances, these distances were dictated by the presence of workable

abundances of the animals, the aim being to sample from H and M populations which

were as close to one another as possible. At Robin Hood's Bay the samples were of M

animals only, 75 m was chosen as a distance likely to be considerably in excess of

migration distance (Janson, 1983).

DNA isolation

Genomic DNA was purified from head-foot tissue of individual *Littorina saxatilis*

using a modified version of Winnepenninckx *et al*. (1993). Tissue was macerated in

300μl 60°C CTAB buffer (2% CTAB, 1.4M NaCl, 20mM EDTA, 100mM Tris-HCl

pH 8, 0.2% β-mercaptoethanol) to which 20mg proteinase K was added and incubated

at 60°C for 3-16hr. Subsequently, two extractions with chloroform:isoamyl alcohol

(24:1) were performed, and the DNA further purified with Promega's Wizard DNA

Clean-Up System following the manufacturer's instructions. Concentration was

assessed by spectrophotometry and adjusted to 100 ng.μl^{-1}.

AFLP analysis

AFLP analysis was performed using a modified version of Vos *et al.* (1995). Adapter

and primer sequences are given in Table 1. For each sample genomic DNA (500ng)

was digested with 5U *Eco*RI (NEB) and 3U *Mse*I (NEB) in 25μl total volume of 1x

NEB buffer #2 supplemented with 100μg.ml^{-1} BSA, for 3 hours at 37°C. Following

enzyme inactivation at 65°C, 25μl of a solution containing 5pmol *Eco*RI adapter,

50pmol *Mse*I adapter, 200U DNA ligase (NEB) and 5μl 10x ligase buffer (NEB), was

added and samples incubated for 16 hours at 16°C. Preselective PCRs were then

performed on 5μl diluted ligation (1:9 with 0.1xTE) in 50μl volumes containing

200μM each dNTP, 25pmol *Eco*+(C/A) primer, 25pmol *Mse*+(C/A) primer, 1.5mM

MgCl$_2$ and 1U *Taq* in manufacturer's buffer. PCR conditions were 20x(94°C 30secs,

56°C 1min, 72°C 1min). Selective *Eco*+3 primers were labeled in 0.5μl volumes

containing 1xT4 PNK buffer, 0.2μl T4 PNK (Promega) 5ng *Eco*+3 primer and 0.1μl

γ^{33}P ATP. Selective PCRs were undertaken in 20μl volumes containing 30ng *Mse*+3

155   primer (see Table 1), 5ng labeled *Eco*+3 primer, 200µM each dNTP, 1.5mM MgCl$_2$,

156   1x buffer (Promega) and 0.4U *Taq*. Cycling conditions in the first cycle were 94°C

157   30secs, 65°C 30sec, 72°C 1min with the annealing temperature reduced by 0.7°C over

158   next 12 cycles, then 23x(94°C 30secs, 56°C 30sec, 72°C 1min). On completion, 20µl

159   STOP solution (95% formamide, 10mM EDTA pH8.0, 0.025% w/v bromophenol

160   blue, 0.025% w/v xylene cyanol) was added. AFLP products were separated on 6%

161   polyacrylamide gels (Sequagel, Flowgen), for 2-2½ hours at 55W then fixed, and

162   dried to the glass plate. Kodak Biomax MR-1 film was exposed to the gel for 48

163   hours. An initial study of reproducibility showed absolute consistency of banding

164   patterns between repeated reactions. Subsequent monitoring where ≈5% reactions

165   were repeated has confirmed this.

Table 1 about here.

166   **Data analysis**

167   Gels were scored manually for band presence/absence. The frequency of the band

168   presence allele was estimated from the band presence/absence matrix for each sample

169   as $p = 1-((N-C)/N)^{0.5}$ where N = sample size and C = number of individuals with the

170   band. This calculation assumes Hardy-Weinberg genotypic frequencies and

171   dominance of band presence over absence.

172   We wish to use the allele frequency data for H and M samples to distinguish two

173   possibilities: 1. That all loci reflect mutation/drift/dispersal balance, perhaps

174   influenced by some general intrinsic barrier to gene exchange between H and M

175   populations, or 2. That strong differentiation is maintained by selection at some

176   proportion of loci, against a background of less-differentiated loci. We followed the

177   approach developed by Bowcock *et al*. (1991) and Beaumont & Nichols (1996) by

178   using simulations to predict the expected distribution of differentiation across loci for

179   a given average divergence. Differentiation is measured by F_{ST}, calculated for each

180   locus by the method of Nei (1977) with the correction suggested by Nei & Chesser

181   (1983). Simulation is necessary because the distribution of F_{ST} across loci is

182   influenced by historical sampling in the natural populations (i.e. by genetic drift) and

183   by experimental sampling. Here there is the added complication that AFLP loci are

184   dominant and, therefore, the experimental sampling error of F_{ST} is greater for high

185   mean allele frequencies (of the 'presence' allele) than for low frequencies. This is

186   because the allele frequencies have to be estimated from the proportion of 'absence'

187   homozygotes and the errors are greatest when this proportion is low.

188   We have used a simple simulation of two populations of size N diploid individuals,

189   with mutation rate μ and migration rate m, per generation. Allele frequencies for 500

190   simulated bi-allelic loci were initiated with a uniform random distribution, equal in

191   the two populations and then allowed to drift for $10N$ generations. Samples of 50

192   individuals were then taken from each simulated population and mean allele

193   frequencies and F_{ST} values were calculated in exactly the same way as for the

194   observed data (with the band presence allele dominant to the absence allele). The

195   simulation was checked by comparing the F_{ST} calculated in this way with both the F_{ST}

196   expected from theory and the F_{ST} calculated from the whole simulated population (i.e.

197   without sampling effects). The theoretical F_{ST} was calculated from

198   $F_{ST}=1/[1+16Nm+16N\mu]$ since only two populations are considered and the mutation

199   rate may be high relative to the migration rate (see below) (Crow & Aoki, 1984). The

200   simulated values calculated from the whole population agreed precisely with this

201   expectation but the simulated sample values showed a consistent upward bias of

202   0.0093 over the range of values of Nm relevant to this study. This bias is consistent

203 with previous simulation studies using Nei's method for calculation of F_{ST} (Slatkin &

204 Barton, 1989).

205 For each comparison between observed samples, Nm in the simulation was set to a

206 value expected to return the observed mean F_{ST} allowing for the estimation bias. The

207 simulation was then repeated 50 times to generate a total of 25000 values of mean

208 allele frequency and F_{ST} (minus those loci that were monomorphic in the simulated

209 samples, approximately 5%). Simulated mean F_{ST} values differed from observed

210 means by up to 6.77% but were always higher, making the test for loci with

211 unexpectedly high levels of differentiation conservative. Observed F_{ST} values were

212 compared with the 0.99 quantile of the simulated values determined for each of 20

213 categories of mean allele frequency, because the distribution of F_{ST} values is expected

214 to vary with mean allele frequency (see below and Fig. 1).

215 **Results**

216 **Levels of polymorphism**

217 A total of 306 fragments (loci) were scored from five primer combinations for 50

218 individuals per sample (Table 2). Additional, variable fragments could not be scored

219 unambiguously and were not considered further. Levels of polymorphism were

220 particularly high with 94.8% of loci polymorphic (a locus was considered

221 polymorphic if at least one individual showed a variant pattern). There was some

222 variation in the number of scorable loci per primer combination with the *Eco*+CTC-

223 *Mse*+CGA yielding 43 polymorphic bands and *Eco*+CAG-*Mse*+CGA yielding 80.

Table 2 about here

224 This high level of polymorphism suggests a value for $N\mu$ of the order of 10^{-1}, using

225 Kimura's (1968) formula for bi-allelic loci. This formula assumes symmetrical

226 mutation, which may not be true for AFLP bands, and ignores the possible existence

227 of many loci that are monomorphic for the 'absence' allele. This may mean that $N\mu$

228 has been overestimated. We have used $N\mu = 0.1$ ($N = 10^3$, $\mu = 10^{-4}$) in the simulations

229 reported below but other runs have demonstrated that neither the mean nor the

230 variance of F_{ST} is sensitive to these parameters (as also observed by Beaumont &

231 Nichols 1996). We have also run simulations with the mutation rate from presence to

232 absence ten times greater than the reciprocal rate. This increases the proportion of loci

233 monomorphic for the absence allele but has no effect on the distribution of F_{ST}.

234 **Detection of differentiated loci**

235 Ten loci had F_{ST} values higher than the 0.99 quantile of the initial simulation results

236 for all three individual H-M comparisons. Since these loci are implicated as being

237 under selection or linked to areas of the genome that are under selection, Nm was

238 recalculated after their removal, simulations were repeated, and the data compared

239 with new 0.99 quantiles. This process was carried out four times. At this stage, no

240 further locus showed observed values of F_{ST} lying above the 0.99 quantiles in all three

241 H-M comparisons, and 15 loci were identified as lying above the 0.99 quantile (Fig.

242 1). If the three H-M comparisons were independent, one would expect to see <<1

243 locus falling outside the 0.99 quantile in all three cases (0.01^3 x 306). However, gene

244 exchange between sites potentially means that allele frequencies do not vary

245 independently. Therefore, we repeated the analysis making the alternative extreme

246 assumption that the three H samples come from one population and the three M

247 samples from another. In this case, all 15 of the loci previously identified fell outside

248 the 0.99 quantile (now based on sample sizes of 150).

249   In all three H-M comparisons on the same shore, the same 15 loci lie above the

250 0.99 quantile, together with a much smaller number of other loci whose behaviour is

251 erratic. In comparisons within morphs, mostly also between shores, there are fewer

252 loci above the 0.99 quantile, they are nearer to this limit, and rarely are any of the 15

253 loci identified above involved (see Fig. 1).

Fig. 1 about here

254   Table 3 shows that when F_{ST} is calculated using all loci, values are usually higher

255 for H-M comparisons than they are for H-H or M-M comparisons. The few within

256 morph comparisons which are as large as the smallest between morph ones are from

257 samples at or near the extremes of the sample range, e.g. Old Peak H-Thornwick Bay

258 H (0.0318). Yet overall, F_{ST} seems to be independent of distance, thus the F_{ST} for H-

259 M at Thornwick Bay is 0.0378 (spatial distance 5 m) while the values for H at

260 Thornwick Bay compared with the two M samples at Robin Hood's Bay (distance 45

261 km) are 0.0350 and 0.0340. The lack of relationship between all F_{ST} values and linear

262 distance is further suggested by a randomization test (Manly, 1996; Manly, 1997)

263 (1000 permutations) when the value of P for association was 0.3690. However if F_{ST}

264 is estimated after removal of the 15 loci considered to be differentiated between H and

265 M (Fig. 2), there is evidence of association with distance, P = 0.0020. In the figure,

266 and for the randomization tests, distance was transformed by taking base 10

Table 3 about here

267 logarithms, F_{ST} by taking $F_{ST}/(1- F_{ST})$ as recommended by Rousset (1997).

268   Two-sample randomization tests (Manly, 1996; Manly, 1997) were carried out on

269 the F_{ST} data in Table 3 either when the values were calculated with, or without, the 15

270 loci considered as likely to be differentiated. For values including these 15 loci, the

271 probability that within morph and between morph F_{ST} values were the same was P =

272 0.001. When these 15 loci were excluded from the F_{ST} estimates, this probability

273 became 0.1450, indicating no difference between the two groups of F_{ST} estimates.

274   Mean F_{ST} values after removal of these 15 loci imply that Nm between H and M

275　morphs within shores is in the range 5.5 at Old Peak, 6.3 at Thornwick Bay and 308

276　individuals per generation (respective *Nm* values were 1.9, 2.0 and 3.9 before

277　removal). *Nm* between M morphs at Robin Hood's Bay is estimated as infinity ($F_{ST} =$

278　0).

279　**Genetic variation among *L. saxatilis* populations**

280　Nei's genetic distances between samples of *L. saxatilis* H and M were used to

281　construct a neighbour-joining tree (Fig. 3a). The three samples of *L. saxatilis* H form

282　one cluster separated from the five samples of *L. saxatilis* M by the greatest internal

283　branch length and with high bootstrap support. When we omitted the data for the 15

284　loci identified as potentially under selection from the three comparisons of *L. saxatilis*

285　H and M, the revised tree showed radically altered structure (Fig. 3b). Now, instead of

286　populations clustering by morphotype (H and M), they cluster by site, with Filey H

287　and M clustering together, Old Peak H and M together, etc.

Fig. 3 about here

Discussion

This study asks whether the *Littorina saxatilis* H-M cline represents a general barrier

to gene exchange or reflects divergence at a limited number of loci under selection.

By generating a large number of marker loci using AFLPs, and using the analytical

approach of Beaumont & Nichols (1996), we have identified at least 15 loci (from a

total of 306 studied; 5%) that seem either to be under selection or (more likely) linked

to loci that are. However, none of the 306 loci is implicated as under selection when

two populations of *L. saxatilis* M are compared from the same shore (Robin Hood's

Bay). It is interesting that our H-M comparisons show differentiation at these loci

regardless of whether they are spatially widely separate (300 m at Old Peak) or close

together (5 m at Flamborough). Within morph comparisons do not show such

differentiation, and now there is evidence of isolation-by-distance. F_{ST} values for

between morph comparisons are evidently higher than for within morph comparisons

when all loci are considered. The F_{ST} values after removal of these exceptional loci

are more nearly similar, but still imply that there is a general barrier to gene exchange

between H and M populations that is greater than would be expected from their spatial

separation.

Our simulation assumes free recombination among loci. In reality, this is clearly

not the case with 300 loci randomly distributed across the genome. In the extreme,

some AFLP bands may be allelic or very tightly linked and so their levels of

differentiation will not be independent. This will be detectable in hybridizing

populations because it will generate strong disequilibrium between differentiated loci.

We are currently analyzing such populations. However, in the present analysis, any

effect of linkage would apply equally to all comparisons and so cannot explain the

difference in distribution of F_{ST} between H-M and within-morph comparisons.

313   Thus, while there are no fixed differences between morphs in any of the

314   populations we have investigated, in appropriate comparisons (H versus M

315   populations), there is a small group of loci which show considerable differentiation

316   against a background of a majority where differentiation is weak. We suggest that this

317   is the most striking aspect of the data reported above: that there is a consistent group

318   of loci apparently differentiated. This point is further supported by comparing trees in

319   which the samples group by morphotype when the differentiated loci are included in

320   the analysis, but by shore when they are excluded. From this we infer that the majority

321   of the AFLP loci are in mutation/drift/dispersal equilibrium, although we cannot

322   exclude the possibility of a general reduction in gene exchange between H and M

323   populations relative to populations of the same morph. Against this background, we

324   suggest that differentiation is being maintained for the small number of differentiated

325   loci by selection on the loci themselves, or on closely linked loci. These findings are

326   consistent with earlier work demonstrating morphological, ecological and behavioural

327   differences between *L. saxatilis* H and M (Hull *et al.*, 1996; Hull, 1998; Pickles &

328   Grahame, 1999) but imply that the genetic differences underlying these characters

329   involve only a small proportion of the genome. This is what would be expected in a

330   case of non-allopatric speciation in progress. However, it could also be the result of

331   differential introgression following secondary contact resulting in homogenization of

332   allele frequencies at all loci except those under selection, or closely linked to loci

333   under selection.

334   The H and M forms of *L. saxatilis* represent one of several cases of divergence in

335   shell shape in this species. Similar variation is reported for shores in Sweden (Janson

336   & Sundberg, 1983), where it is considered to be phenotypic. It has been shown that

337   some allozyme loci are under selection, or linked to selected loci, in Swedish

338   populations (Johannesson *et al.*, 1995a; Johannesson & Tatarenkov, 1997), although

339   this has not been explicitly associated with shell form. On the Galician coast of Spain

340   very different shell forms occur in populations between which there is some

341   restriction of gene flow and evidence of selection on shell form (Johannesson *et al.*,

342   1995b; Rolán-Alvarez *et al.*, 1997). We do not have direct evidence of selection

343   operating on H and M forms on the Yorkshire coast, but it seems reasonable to infer

344   that it does. The findings from Britain and Spain suggest that a pervasive influence in

345   habitat use and subsequent diversification in *L. saxatilis* is the vertical shore gradient.

346   In turn, this suggests an unusually simple physical background (a spatially very

347   restricted cline, limited by the extent of the intertidal zones occupied by the animals)

348   against which to study speciation processes.

349   Whether the differentiation of the small proportion of loci between H and M is

350   primary (the result of divergent selection) or secondary (the result of renewed

351   contact), the main point is that differentiation is maintained for a small portion of the

352   genome, while gene exchange continues to prevent divergence at the majority of loci.

353   Detailed investigation of these loci in particular may provide important insights into

354   the nature of the barrier between these two forms of intertidal snail, and into the

355   evolution of barriers to gene exchange in general.

Acknowledgements

356

Table 1. Adapters and selective primer sequences used for AFLP analysis.

Primer/adapter

Adapters

*Eco*RI	5'-CTCGTAGACTGCGTACC-3'
	3'-CATCTGACGCATGGTTAA-5'
*Mse*I	5'-GACGATGAGTCCTGAG-3'
	3'-TACTCAGGACTCAT-5'

Primers	Sequence (5′-3′)
Eco+1	
Eco+A	GACTGCGTACCAATTCA
Eco+C	GACTGCGTACCAATTCC
Mse+1	
Mse+A	GATGAGTCCTGAGTAAA
Mse+C	GATGAGTCCTGAGTAAC
Eco+3	
Eco+ACG	GACTGCGTACCAATTCACG
Eco+CAG	GACTGCGTACCAATTCCAG
Eco+CTC	GACTGCGTACCAATTCCTC
Mse+3	
Mse+AGT	GATGAGTCCTGAGTAAAGT
Mse+ATC	GATGAGTCCTGAGTAAATC
Mse+CAA	GATGAGTCCTGAGTAACAA
Mse+CGA	GATGAGTCCTGAGTAACGA

Combinations

A	*Eco*+CTC-*Mse*+CGA
B	*Eco*+CAG-*Mse*+CGA
D	*Eco*+CAG-*Mse*+CAA
E	*Eco*+ACG-*Mse*+ATC
F	*Eco*+ACG-*Mse*+AGT

Table 2. Levels of polymorphism of scored AFLP markers.

| Restriction site | Primer combination | | | | | |
	A	B	D	E	F	Total
No. of variable bands	43	80	54	54	59	290
No. of fixed bands	0	3	8	1	4	16
Total	43	83	62	55	63	306

Table 3. F_{ST} (below diagonal) between populations of *Littorina saxatilis* (mean over 290 loci). Above diagonal, F_{ST} following removal of 15 loci. Standard errors of F_{ST} estimates range from 9.80-22.15% (below diagonal) and 10.19-22.20% (above diagonal) of the mean. TH, Thornwick Bay; OP, Old Peak; FY, Filey Brigg; RB, Robin Hood's Bay (two samples, M only).

	TH-H	TH-M	OP-H	OP-M	FY-H	FY-M	RB1	RB2
TH-H		0.0190	0.0311	0.0372	0.0242	0.0227	0.0301	0.0298
TH-M	0.0378		0.0397	0.0293	0.0322	0.0217	0.0260	0.0255
OP-H	0.0318	0.0633		0.0204	0.0223	0.0256	0.0254	0.0247
OP-M	0.0489	0.0292	0.0396		0.0309	0.0241	0.0221	0.0212
FY-H	0.0247	0.0551	0.0238	0.0480		0.0095	0.0204	0.0207
FY-M	0.0328	0.0236	0.0402	0.0244	0.0247		0.0148	0.0142
RB1	0.0350	0.0308	0.0347	0.0241	0.0286	0.0156		0.0055
RB2	0.0340	0.0299	0.0339	0.0230	0.0291	0.0153	0.0052	

References

Barton, N.H. & Hewitt, G.M. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.* **16**: 113-148.

Barton, N.H. 1988. Speciation. In: *Analytical biogeography* (A.A. Myers & P.S. Giller, eds), pp. 185-218. Chapman & Hall, London.

Barton, N.H. & Hewitt, G.M. 1981. The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity.* **47**: 367-383.

Beaumont, M.A. & Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B.* **263**: 1619-1626.

Boulding, E.G., Holst, M. & Pilon, V. 1999. Changes in selection on gastropod shell size and thickness with wave-exposure on Northeastern Pacific shores. *J. exp. mar. Biol. Ecol.* **232**: 217-239.

Bowcock, A.M., Kidd, J.R., Mountain, J.L., Hebert, J.M., Carotenuto, L., Kidd, K.K. & Cavalli-Sforza, L.L. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* **88**: 839-843.

Bush, G.L. & Howard, D.J. 1986. Allopatric and non-allopatric speciation; assumptions and evidence. In: *Evolutionary processes and Theory.* (S. Karlin & E. Nevo, eds), pp. 411-438. Academic Press Inc., San Diego.

Clarke, R.K., Grahame, J. & Mill, P.J. 1999. Variation and constraint in the shells of two sibling species of intertidal rough periwinkles (Gastropoda: *Littorina* spp.). *J. Zool. Lond.* **247**: 145-154.

Crow, J.F. & Aoki, K. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proc. Natl. Acad. Sci. USA* **81**: 6073-6077.

Endler, J.A. 1977. *Geographic variation, speciation, and clines.* Princeton University Press, Princeton, New Jersey.

Feder, J.L., Opp, S.B., Wlazlo, B., Reynolds, K. & Go, W. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proc. Natl. Acad. Sci. USA* **91**: 7990-7994.

Feder, J.L., Roethele, J.B., Wlazlo, B. & Berlocher, S.H. 1997. Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proc. Natl. Acad. Sci. USA* **94**: 11417-11421.

Felsenstein, J. 1993. *PHYLIP: Phylogenetic Inference Package.* University of Washington, Seattle.

Grahame, J., Hull, S.L., Mill, P.J. & Hemmingway, R. 1997. Discovering unrecognised diversity among marine molluscs. In: *Marine Biodiversity: Patterns and Processes* (R.F.G. Ormond, J.D. Gage & M.V. Angel, eds), pp. 293-318. Cambridge University Press, Cambridge, U.K.

Grahame, J. & Mill, P.J. 1986. Relative size of the foot of two species of *Littorina* on a rocky shore in Wales. *J. Zool. Lond.* **208**: 229-236.

Grahame, J. & Mill, P.J. 1993. Shell shape variation in rough periwinkles: genotypic and phenotypic effects. In: *Quantified phenotypic responses in morphology and physiology* (J.C. Aldrich, ed), pp 25-30. JAPAGA, Ashford.

Heller, J. 1976. The effects of exposure and predation on the shell of two British winkles. *J. Zool. Lond.* **179**: 201-213.

Hull, S.L. 1998. Assortative mating between two morphs of *Littorina saxatilis* on a shore in Yorkshire. *Hydrobiologia.* **378**: 79-88.

Hull, S.L., Grahame, J. & Mill, P.J. 1996. Morphological divergence and evidence for reproductive isolation in *Littorina saxatilis* (Olivi) in northeast England. *J. moll. Studs.* **62**: 89-99.

Janson, K. 1983. Selection and migration in two distinct phenotypes of *Littorina saxatilis* in Sweden. *Oecologia.* **59**: 58-61.

Janson, K. & Sundberg, P. 1983. Multivariate morphometric analysis of two varieties of *Littorina saxatilis* from the Swedish west coast. *Mar. Biol.* **74**: 49-53.

Johannesson, B. 1986. Shell morphology of *Littorina saxatilis* Olivi: the relative importance of physical factors and predation. *J exp mar Biol Ecol.* **102**: 183-195.

Johannesson, B. & Johannesson, K. 1996. Population differences in behaviour and morphology in the snail *Littorina saxatilis*: phenotypic plasticity or genetic differentiation? *J. Zool. Lond.* **240**: 475-493.

Johannesson, K., Johannesson, B. & Lundgren, U. 1995a. Strong natural selection causes microscale allozyme variation in a marine snail. *Proc Natl Acad Sci USA* **92**: 2602-2606.

Johannesson, K., Rolán-Alvarez, E. & Ekendahl, A. 1995b. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution.* **49**: 1180-1190.

Johannesson, K. & Tatarenkov, A. 1997. Allozyme variation in a snail (*Littorina saxatilis*) - deconfounding the effects of microhabitat and gene flow. *Evolution.* **51**: 402-409.

Kimura, M. 1968. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**: 247-269.

Manly, B.F.J. 1996. *RT, a Program for Randomization Testing. Version 2.0.* Centre for Applications of Statistics and Mathematics, University of Otago, Otago.

Manly, B.F.J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology.* Chapman & Hall, London.

Mayr, E. 1963. *Animal Species and Evolution.* The Bellknap Press of Harvard
University Press, Cambridge, Massachusetts.

Nei, M. 1977. *F*-statistics and analysis of gene diversity in subdivided populations.
Ann. Hum. Genet. **41**: 225-233.

Nei, M. & Chesser, R.K. 1983. Estimation of fixation indices and gene diversities.
Ann. Hum. Genet. **47**: 253-259.

Newkirk, G.F. & Doyle, R.W. 1975. Genetic analysis of shell-shape variation in
Littorina saxatilis on an environmental cline. *Mar. Biol.* **30**: 227-237.

Pickles, A. & Grahame, J. 1999. Mate choice in divergent morphs of *Littorina
saxatilis* (Olivi): speciation in action? *Anim. Behav.* **58**: 181-184.

Raffaelli, D. 1978. The relationship between shell injuries, shell thickness and habitat
characteristics of the intertidal snail *Littorina rudis* Maton. *J. moll. Studs.* **44**:
166-170.

Reid, D.G. 1996. *Systematics and evolution of Littorina.* The Ray Society, Dorchester.

Rice, W.R. & Hostert, E.E. 1993. Laboratory experiments on speciation: what have
we learned in 40 years? *Evolution.* **47**: 1637-1653.

Rolán-Alvarez, E., Johannesson, K. & Erlandsson, J. 1997. The maintenance of a
cline in the marine snail *Littorina saxatilis*: the role of home site advantage
and hybrid fitness. *Evolution.* **51**: 1836-1847.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from *F*-statistics
under isolation by distance. *Genetics.* **145**: 1219-1228.

Schliewen, U.K., Tautz, D. & Paabo, S. 1994. Sympatric speciation suggested by
monophyly of crater lake cichlids. *Nature.* **368**: 629-632.

Schluter, D. 1996. Ecological speciation in postglacial fishes. *Phil. Trans. R. Soc.
Lond. B.* **351**: 807-814.

Slatkin, M. & Barton, N.H. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution.* **43**: 1349-1368.

Szymura, J.M. & Barton, N.H. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution.* **45**: 237-261.

Trussell, G.C. & Smith, L.D. 2000. Induced defenses in response to an invading crab predator: an explanation of historical and geographic phenotypic change. *Proc. Natl. Acad. Sci. USA* **97**: 2123-2127.

Vermeij, G.J. 1987. *Evolution and escalation: an ecological history of life.* Princeton University Press, Princeton, New Jersey.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407-4414.

Wilding, C.S., Grahame, J. & Mill, P.J. 2000. Mitochondrial DNA CoI haplotype variation in sibling species of rough periwinkles. *Heredity.* **85**: 62-74.

Wilding, C.S., Grahame, J. & Mill, P.J. 2001. Correlation of morphological diversity with molecular marker diversity in the rough periwinkle *Littorina saxatilis* (Olivi). *J. Shellfish. Res.* **xx**: xxx.

Wilson, A.B., Noack-Kunnmann, K. & Meyer, A. 2000. Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proc. R. Soc. Lond. B.* **267**: 2133-2141.

Winnepenninckx, B., Backeljau, T. & De Wachter, R. 1993. Extraction of high-molecular-weight DNA from molluscs. *Trends. Genet.* **9**: 407.

Figure legends

Fig. 1

F_{ST} values estimated from 290 variable AFLP loci plotted against mean allele

frequency in three H-M comparisons (Thornwick Bay, Old Peak and Filey), one M-M

comparison (Robin Hood's Bay) and six representative between shores, within morph

comparison. The solid line represents the 0.99 quantile estimated from a simulation

model (see methods). The position of each of the 15 loci implicated as linked to a

region under selection due to their presence outside the region defined by the 0.99

quantile consistently in all three H-M comparisons (see Results) is labeled with the

locus identifier (where A-F = primer combination, see Table 1).

Fig. 2

Relationship of mean F_{ST} with linear distance, 15 differentiating loci removed.

Fig. 3.

Neighbour-Joining tree calculated from allele frequency data in PHYLIP (Felsenstein,

1993). Data were bootstrapped (x100) using SEQBOOT and Nei's genetic distance

calculated using GENDIST. Distances were clustered with NEIGHBOR and the

consensus tree constructed with CONSENSE. Bootstrap values are shown at the

nodes of a representative non-consensus tree in order to retain branch length

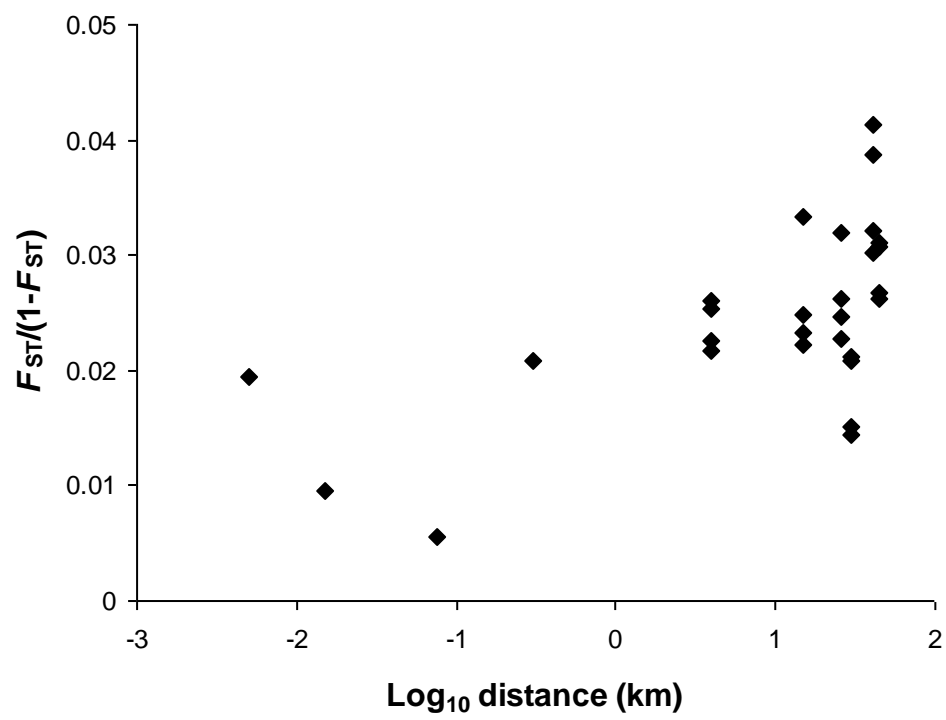information. a), using all 290 AFLP loci; b), after removal of 15 differentiated loci.
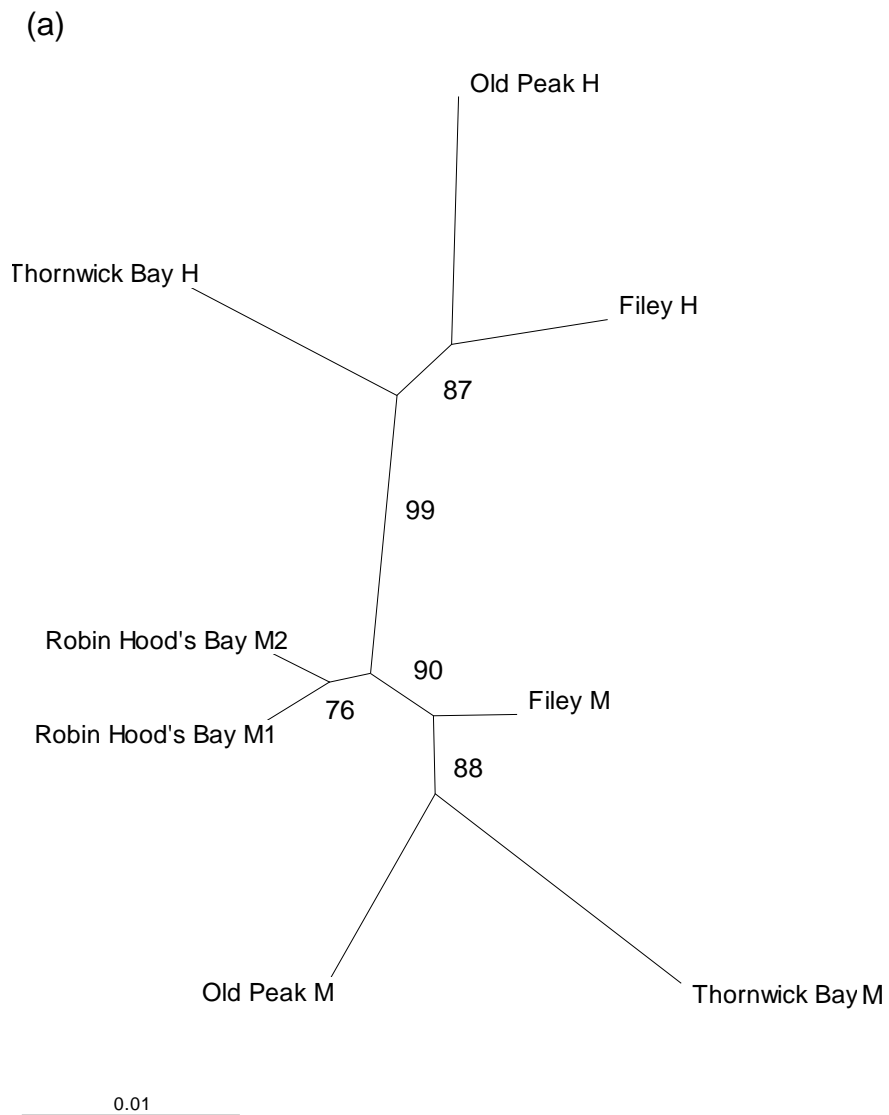
Fig. 2

Fig. 3.

(a)

Old Peak H

Thornwick Bay H

Filey H

87

99

Robin Hood's Bay M2

90

Filey M

76

Robin Hood's Bay M1

88

Old Peak M

Thornwick Bay M

0.01

Fig. 3

(b)

Thornwick Bay M

Filey H

Filey M

73

99

67

Old Peak H

Thornwick Bay H

99

81

Robin Hood's Bay 2

Robin Hood's Bay 1

Old Peak M

0.01