LJMU Research Online

Wilding, CS, Grahame, J and Mill, PJ

A GTT microsatellite repeat motif and differentiation between morphological forms of Littorina saxatilis: speciation in progress?

http://researchonline.ljmu.ac.uk/2707/

Article

A GTT microsatellite repeat motif and differentiation between morphological forms of *Littorina saxatilis*: speciation in progress?

Craig S. Wilding*, John Grahame, Peter J. Mill

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

ABSTRACT: A size (and sequence) variable microsatellite has been identified in the gastropod *Littorina saxatilis* following sequencing of products obtained through randomly amplified polymorphic DNA amplification. Size frequency distributions for this GTT repeat motif have been produced from a total of 439 *L. saxatilis*. Although there is evidence for a high prevalence of null alleles, consistent, largely significant, differences are found between the average size of this repeat in high shore, thin shelled morphs (*L. saxatilis* H) compared to mid shore, thick shelled animals (*L. saxatilis* M), with the former having consistently larger allele sizes. These findings are evident on both a large scale (around the coast of Britain) and a microgeographic scale (replicated sampling of a single shore). Unusually large alleles (>400 bp) are also considerably more prevalent in *L. saxatilis* H, and sequencing of these, including DNA flanking the GTT repeat, indicates that they do not represent independent expansion events. The significant difference in GTT repeat size, and increased prevalence of such related alleles in the H form compared to the M morph suggests that either gene flow is greater within, than between, morphs or that selection is acting on this locus or a closely linked locus. We argue for the latter scenario and discuss why this indicates that these forms are diverging across their range in the British Isles.

KEY WORDS: Speciation · *Littorina saxatilis* · Microsatellite · GTT · Repeat · Gastropod

INTRODUCTION

Because of the substantial morphological and ecological diversity of the intertidal, ovoviviparous gastropod mollusc *Littorina saxatilis* (Olivi), numerous names have been proposed as varieties, forms or monstrosities (Reid 1996). Many of these varieties are of restricted occurrence, but two are widespread on rocky coastlines and often found parapatrically within a shore. On high shore massive boulders and cliffs, *L. saxatilis* are thin shelled and patulous, whilst mid shore individuals, found on smaller boulders and bedrock, are thick shelled with smaller apertures. These have sometimes been regarded as separate species, e.g. Maton (1797) described the thick shelled animals as *L. rudis*, and whilst a variety of synonyms have been attributed to the

higher shore form (see Reid 1996), *L. patula* has been widely used (see Heller 1975). However, later work, particularly on shell, radula and penis morphology, led to Raffaelli (1979) concluding that these represent morphological variants of *L. saxatilis*, and this has become the accepted opinion (Reid 1996).

More recently, Hull et al. (1996) have studied populations of these 2 morphs on the east coast of England. They found that the high shore, thin shelled animals (termed *Littorina saxatilis* H for high shore) differed in reproductive characteristics from mid shore individuals (termed *L. saxatilis* M for mid shore). Hull et al. (1996) found that uncleaved embryos of *L. saxatilis* H were significantly larger than those of *L. saxatilis* M, and this was consistent at a number of sample sites. Crucially, investigation of rare *Littorina* of intermediate shape showed that uncleaved embryos within a single individual fell into 2 size classes, corresponding to the sizes characteristic of *L. saxatilis* H and M, and

---

that a high proportion of these embryos were aborted. This suggests that these intermediates are hybrids between *L. saxatilis* H and M, and that a real barrier to gene flow between these 2 forms exists. This view is also supported by behavioural studies demonstrating assortative mating (Hull 1998, Pickles & Grahame 1999).

Allozyme studies on *Littorina saxatilis* of similar form in Sweden concluded that very little of the observed variation is associated with the morphologically different forms (Janson & Ward 1984). Wilding et al. (1998) have also applied randomly amplified polymorphic DNA (RAPD) methodology to an analysis of *L. saxatilis* H and M, and found that differentiation of these morphs was detectable at 2 sites separated by approximately 30 km (Old Peak and Filey) but that differentiation was much greater at Old Peak. Mitochondrial DNA analysis (Wilding et al. 2000a) and analysis of 4 nuclear DNA RFLPs (restriction fragment length polymorphisms; Wilding et al. 2000b) has also been undertaken on these H and M forms but consistent differences were not detected.

The high mutability and rapid evolution of microsatellites (Freimer & Slatkin 1996) suggests that such markers could be invaluable for the study of genetic differentiation of these molluscs. Here, variation within a triplet (GTT) microsatellite repeat identified from a RAPD band (Grahame et al. 1997) is analysed in populations of *L. saxatilis* H and M from a variety of locations around the coast of Britain and Ireland.

Fig. 1. Collection sites for *Littorina saxatilis* used in this study

MATERIALS AND METHODS

Littorina saxatilis were collected from 8 locations around the coast of Britain and Ireland (Fig. 1). At 7 of these, *Littorina saxatilis* H were found on high shore cliffs or massive boulders, whilst *L. saxatilis* M were taken from mid shore bedrock and boulders. However, at Ballynahown, Ireland, *L. saxatilis* M were found on high shore small boulders, whilst on mid shore bedrock they more closely resembled the H form in shell shape. The factors responsible, and the classification of the *L. saxatilis* at this last site, have not yet been fully examined.

Littorina saxatilis H and M were characterised on the basis of shell morphology. Only brooding females were used so as to avoid confusing the H samples with specimens of the morphologically similar, oviparous *L. arcana* Hannaford Ellis, or the M animals with the oviparous *L. compressa* Jeffreys.

DNA extraction. Total genomic DNA was extracted from the head-foot region of individual *Littorina saxatilis*, using the protocol (number 47) of Ashburner (1989). DNA concentrations were measured by fluorometry and adjusted to 10 ng μl^{-1}.

Discovery of the repeat motif. Through RAPD screening of populations using the primer RAPD-H (5′-GCC GTG GTT A-3′), a band was identified that was amplified in *Littorina saxatilis* M, but not in *L. saxatilis* H (Grahame et al. 1997). This was subsequently cloned into the TA cloning vector (Invitrogen) following the manufacturer′s instructions. Twenty-four clones were screened by hybridisation analysis and 7 unique clones were identified and subsequently sequenced. One of these (CLONE) contained a microsatellite sequence with a GTT repeat motif (Fig. 2). No other clone contained any repetitive sequences, and no match was found for any of the 7 sequences using BLAST searches.

Amplification of the repeat. The primer pair RAPDRPT1 and RAPDRPT2 (Table 1, Fig. 3) were designed to the sequence flanking the repeat. RAPDRPT2B and RAPDRPT2C, primers that are positioned 3 bases and 4 bases respectively 3′ of RAPDRPT2 (Table 1), were also

Table 1. Sequences of primers (given 5′ to 3′) used in amplifications of the GTT repeat. Primers RAPDRPT2, RAPDRPT2B and RAPDRPT2C are aligned to demonstrate overlaps

Primer	Sequence
RAPDRPT1	TTT ACC TCA TCT GTT GCG TCC
RAPDRPT2	AAT AAC ATA CAG CGA CGA CGG
RAPDRPT2B	AC ATA CAG CGA CGA CGG CAA
RAPDRPT2C	A CAG CGA CGA CGG CAA C
RAPDRPT3	AGT AAC GGC AGA CGC CAT

```
CLONE    RR1→GATA(GTT)₂₅TTTGTT
MM107    RR1→GATA(GTT)₁₇
SH121    RR1→GATA(GTT)₁₇
SM121    RR1→GATA(GTT)₁₈
MM112    RR1→GATA(GTT)₁₅TTT(GTT)₁₀
FM171    RR1→GATA(GTT)₄₁
FH208    RR1→GATA(GTT)₅GTC(GTT)₂₂TTT(GTT)₂TTT(GTT)₄GTG(GTT)₃TT(GTT)₈(GCT)₂(GTT)₄
MM118    RR1→GATA(GTT)₂₄
MH130    RR1→GATA(GTT)₂₁TTT(GTT)₉
FM208    RR1→GATA(GTT)₅₂
```

$\text{MH485}\quad \text{RR1}\rightarrow\text{GATA}(GTT)_{19}TT(GTT)_{5}TT(GTT)_{10}TT(GTT)_{17}\quad\quad GCT\quad\quad (GTT)_5\ GCT\quad (GTT)_4\ GGT\quad (GTT)_5(GCT)_2(GTT)_{10}TT(GTT)_{19}(GCT)_7\quad\quad (GTT)_{31}(GCT)_2(GTT)_4$

$\text{BM400}\quad \text{RR1}\rightarrow\text{GATA}(GTT)_{18}TT(GTT)_{5}TT(GTT)_{10}TT(GTT)_{10}\quad\quad GCT\quad\quad (GTT)_5\ GCT\quad (GTT)_4\ GGT\quad (GTT)_5(GCT)_2(GTT)_{10}TT(GTT)_{16}(GCT)_7(GTT)_{22}GAT(GTT)_9\ (GCT)_2(GTT)_4$

$\text{BM446}\quad \text{RR1}\rightarrow\text{GATA}(GTT)_{21}TT(GTT)_{5}TT(GTT)_{8}\ TT(GTT)_{7}TT(GTT)_{10}TT(GTT)_{5}GCT\quad (GTT)_3TT(GTT)_3ATT(GTT)_2\ GCT\quad GTT)_4\ GGT\quad (GTT)_6(GCT)_2(GTT)_5\ TT(GTT)_{19}(GCT)_3\quad\quad (GTT)_9\ (GCT)_2(GTT)_4$

$\text{GH422}\quad \text{RR1}\rightarrow\text{GATA}(GTT)_{18}TT(GTT)_{5}TT(GTT)_{10}TT(GTT)_{11}\quad\quad (GCT)_2\quad\quad (GTT)_4(GCT)_2(GTT)_3(GGT)_2(GTT)_4(GCT)_3(GTT)_9\ GT(GTT)_{13}(GCT)_9\quad\quad (GTT)_{26}(GCT)_2(GTT)_4$

$\text{HH422}^{\dagger}\ \text{RR1}\rightarrow\text{GATA}(GTT)_{29}TT(GTT)_{5}TT(GTT)_{6}\ TT(GTT)_{9}\quad\quad TTTTT(GTT)_5GCT\quad (GTT)_3TT(GTT)_3GTC(GTT)_2\ GCT\quad (GTT)_4\ GGT\quad (GTT)_6(GCT)_2(GTT)_7\ TT(GTT)_{11}(GCT)_4\quad\quad (GTT)_{14}(GCT)_2(GTT)_4$

```
             10        20        30        40        50        60        70        80        90       100
              .         .         .         .         .         .         .         .         .         .
CLONE   GCCGTCGTCGCTGTATGTTATTGTTATAGTTGTTGATGTCAGTCTTATTCTTGTCTTCTCTGCCTCCCCCTCCTTTTTCTTCTCTTTCTCCTTCTTTTTC
MM107   .............----.........................................----------------------........
SH121   .............----.........................................----------------------........
SM121   .............----.........................................----------------------........
MM112   .............----.........................................-------------------.C......
FM171   .............----.........................................----------------------........
FH208   .............----.........................................----------------------........
MM118   .............----C........................................----------------------........
MH130   .............----.......................................A.----------------------........
FM208   .............----.........................................----------------------........
MH485   .............----.........................................----------------------........
BM400   .............----..............................C..........----------------------........
BM446   .............----.........................................----------------------........
GH422   .............----.........................................----------------------........

            110       120       130       140       150       160       170       180       190       200
              .         .         .         .         .         .         .         .         .         .
CLONE   TGCTGCCCATCTGTCCGTCTGTCTGCCTGCCTGTCTGTCTATCTGTCTTTTG-ATACTAAAACAACAAAGAGAGCTAGATGTATACATTTCACCCGGTC←RR3
MM107   ..........A...........................T.........................←RR3
SH121   ..........A...........................T.........................←RR3
SM121   ..........A...........................T.........................←RR3
MM112   ........................A.............T...........T.C...........←RR3
FM171   ........................A.............T...........T.............←RR3
FH208   .....................................-.........................←RR3
MM118   ........................A.............T.............T...........←RR3
MH130   ........................A.............T.............T...........←RR3
FM208   .....................................-.........................←RR3
MH485   ...................................G.T.........................←RR3
BM400   .....................A.............G.T.........................←RR3
BM446   ...................................G.T.........................←RR3
GH422   ...................................G.T.........................←RR3
```
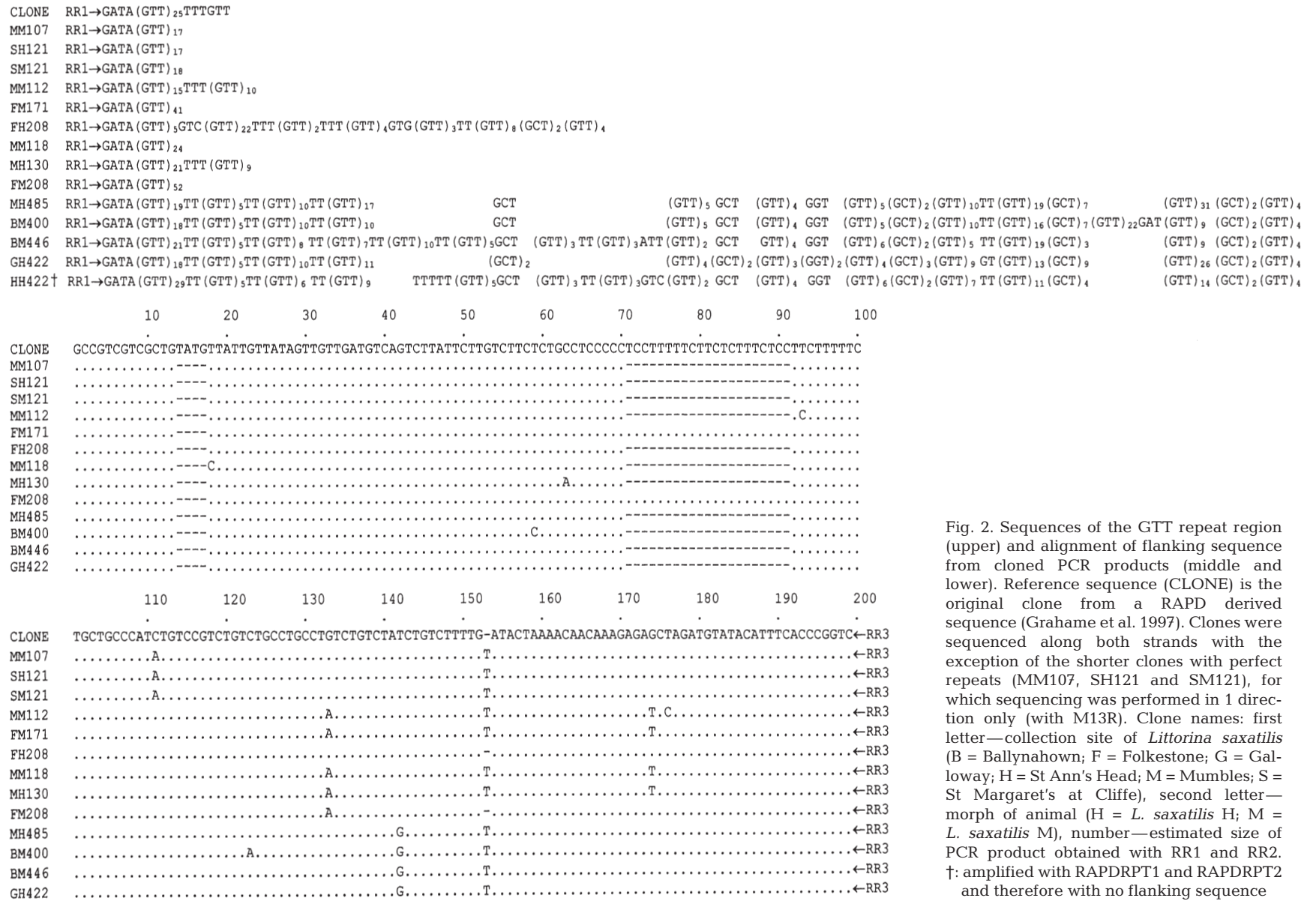
Fig. 2. Sequences of the GTT repeat region (upper) and alignment of flanking sequence from cloned PCR products (middle and lower). Reference sequence (CLONE) is the original clone from a RAPD derived sequence (Grahame et al. 1997). Clones were sequenced along both strands with the exception of the shorter clones with perfect repeats (MM107, SH121 and SM121), for which sequencing was performed in 1 direction only (with M13R). Clone names: first letter—collection site of *Littorina saxatilis* (B = Ballynahown; F = Folkestone; G = Galloway; H = St Ann's Head; M = Mumbles; S = St Margaret's at Cliffe), second letter—morph of animal (H = *L. saxatilis* H; M = *L. saxatilis* M), number—estimated size of PCR product obtained with RR1 and RR2. †: amplified with RAPDRPT1 and RAPDRPT2 and therefore with no flanking sequence
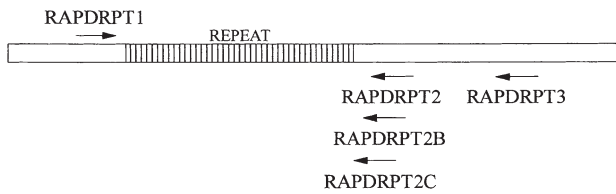
Fig. 3. Relative positions of priming sites around the repeat region

later tested in an effort to resolve occasional problems of non-specific priming with RAPDRPT2 (see 'Results'). An additional primer, RAPDRPT3, was used in amplification of the microsatellite with flanking region for sequencing (see later). Primers were used in 50 µl reactions containing 50 mM KCl, 10 mM Tris-HCl pH 9.0, 1.5 mM MgCl$_2$, 0.1% Triton X-100, 0.01% gelatin, 200 µM each dNTP, 25 pmol each primer, 25 ng DNA and 1 U *Taq* (Super*taq*, HT Biotechnologies). Ten µl of positive PCR reactions were run on 3% metaphor agarose gels (FMC Bioproducts) and stained with ethidium bromide. Band sizes were estimated from Polaroid photographs by compari-

son with a 100 bp standard. Spreadex™ gels (Elchrom Scientific, Cham, Switzerland), which have a resolution of 3 bp, were tested on some samples but gave no greater resolution than the 3% agarose method, which was used for the majority of samples.

Data analysis. Amplification with these primers produced 1 or 2 bands per individual. GENEPOP (Raymond & Rousset 1995) was used to implement a Markov chain method (Guo & Thompson 1992) for estimation of the Hardy-Weinberg (H-W) exact probability (parameters: 50 batches and 1000 iterations per batch with dememorisation 1000). For certain populations (see 'Results'), it became apparent that there was an unusual lack of heterozygotes (e.g. Langland Bay H: 7 of 10 are homozygous; Old Peak M: 10 of 14 are homozygous). This could be a real phenomenon, perhaps due simply to sampling, or could be due to non-amplification of an allele (null allele). Consequently, 2 separate approaches were taken for the analysis of these data. In the first, the sizes for both bands of each heterozygote were used in analyses, whilst samples with only a single band were assumed to be homozygous and band sizes were therefore counted twice. For the second approach, single bands were assumed to represent the heterozygous condition with a null allele and individuals for which no amplification was possible were scored as null-null homozygotes. Additionally, *Littorina saxatilis* H and M were pooled for each site and the analysis repeated in order to investigate whether there was evidence for the Wahlund effect (which would be supported by conformity to H-W in H and M individually, but non-conformity after pooling).

Differences between average sizes of the repeat in different groups (*Littorina saxatilis* H and M) were tested using a Mann-Whitney test implemented in MINITAB v9.2 (Minitab Inc., State College, PA). Two groups of allele sizes became apparent, with very few sizes in the 260 to 400 bp range (Fig. 4). In some instances bands >400 bp were discovered not to be the GTT repeat (see 'Results'). In case any of the large bands (>400 bp) were misclassified as the repeat, thus biasing this test, an additional Mann-Whitney test was performed using only alleles of sizes <300 bp (Table 2). Bonferroni corrections for multiple tests were undertaken according to Hochberg (1988).

Sequencing of the repeat. Examples of certain sizes of bands were picked for sequencing to ensure that all products were the GTT repeat, and to test whether the flanking region harboured any variation. Fourteen bands, representing most of the size range of alleles encountered (≈100 to 500 bp) were picked. Only 1 band per size was used. Sequences of repeat regions and regions flanking the repeat were obtained following cloning of products amplified using the primer pair RAPDRPT1 and RAPDRPT3. RAPDRPT3 is separated
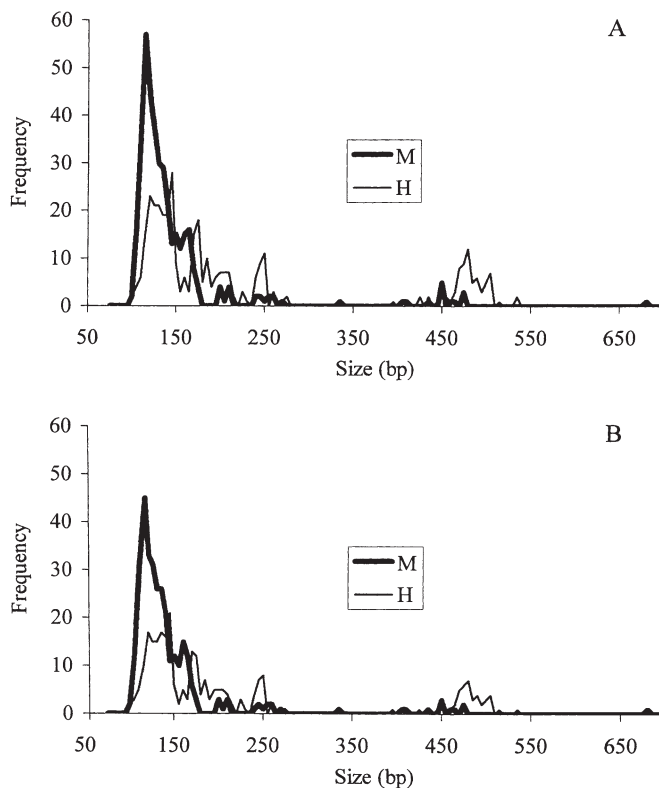

Fig. 4. Size frequency distribution of repeat products in *Littorina saxatilis* H and *L. saxatilis* M. Sizes were pooled into 5 bp categories. With single bands scored as (A) homozygotes and (B) heterozygous with a null allele

Table 2. Fit to Hardy-Weinberg equilibrium for the GTT repeat data. exp. hom. = expected homozygosity; obs. hom. = observed homozygosity; exp. het. = expected heterozygosity; obs. het. = observed heterozygosity; F_{IS} = measure of heterozygote deficiency within populations; p = probability of rejecting hypothesis of H-W equilibrium; SE p = standard error of this estimate; N = sample size. In the first analysis, single bands were assumed to represent homozygous individuals. In the second analysis, single bands were assumed to represent the heterozygous condition with a null allele

Population	exp. hom.	obs. hom.	exp. het.	obs. het.	F_{IS}	p	SE p	N
Single band = homozygote								
Galloway M	4.144	11	44.856	38	0.154	0.0000	0.0000	98
Galloway H	2.389	19	45.611	29	0.367	0.0000	0.0000	96
Mumbles M	7.116	25	66.884	49	0.269	0.0000	0.0000	148
Mumbles H	2.98	52	71.02	22	0.692	0.0000	0.0000	148
Langland Bay M	1.821	6	18.179	14	0.235	0.0150	0.0077	40
Langland Bay H	1.421	7	8.579	3	0.662	0.0000	0.0000	20
Old Peak M	1.111	10	12.889	4	0.698	0.0000	0.0000	28
Old Peak H	1.407	6	12.593	8	0.373	0.0078	0.0078	28
Folkestone M	2.059	4	6.941	5	0.292	0.3622	0.0301	18
Folkestone H	1.474	5	8.526	5	0.427	0.0002	0.0002	20
St Margaret's at Cliffe M	1.579	2	8.421	8	0.053	0.2053	0.0190	20
St Margaret's at Cliffe H	1	5	9	5	0.458	0.0000	0.0000	20
Ballynahown M	1.351	10	17.649	9	0.497	0.0000	0.0000	38
Ballynahown H	1.410	6	18.590	14	0.252	0.0000	0.0000	40
St Ann's Head (slope bottom)	3.544	17	25.456	12	0.533	0.0000	0.0000	58
St Ann's Head (slope top)	2.281	23	26.719	6	0.779	0.0000	0.0000	58
Single band = heterozygote with null allele								
Galloway M	3.907	0	45.093	49	−0.088	0.7099	0.0365	98
Galloway H	3.211	0	44.789	48	−0.073	0.9239	0.0323	96
Mumbles M	6.871	0	67.129	74	−0.103	0.3352	0.0414	148
Mumbles H	11.902	3	65.098	74	−0.138	0.9914	0.0086	154
Langland Bay M	1.487	0	18.513	20	−0.083	0.3779	0.0413	40
Langland Bay H	2.048	1	8.952	10	−0.124	0.6695	0.045	22
Old Peak M	1.926	0	12.074	14	−0.167	0.7091	0.0447	28
Old Peak H	1.148	0	12.852	14	−0.093	0.1987	0.0423	28
Folkestone M	1.412	0	7.588	9	−0.200	1	0	18
Folkestone H	1.053	0	8.947	10	−0.125	0.2032	0.0301	20
St Margaret's at Cliffe M	1.316	0	8.684	10	−0.161	0.4921	0.0265	20
St Margaret's at Cliffe H	0.842	0	9.158	10	−0.098	0.3921	0.0424	20
Ballynahown M	1.676	0	17.324	19	−0.100	0.5716	0.0635	38
Ballynahown H	1.179	0	18.821	20	−0.064	0.2791	0.0523	40
St Ann's Head (slope bottom)	3.754	0	25.246	29	−0.152	0.0558	0.0199	58
St Ann's Head (slope top)	5.729	1	24.271	29	−0.199	0.965	0.0193	60
Single band = heterozygote with null allele. H and M combined								
Galloway	6.264	0	90.736	97	−0.069	0.5194	0.0622	194
Mumbles	15.827	3	135.173	148	−0.095	0.8856	0.0389	302
Langland Bay	2.803	1	28.197	30	−0.065	0.405	0.0562	62
Old Peak	2.891	0	25.109	28	−0.118	0.7023	0.0576	56
Folkestone	2.351	0	16.649	19	−0.146	0.2986	0.0356	38
St Margaret's at Cliffe	1.949	0	18.051	20	−0.111	0.2409	0.0344	40
Ballynahown	2.831	0	36.169	39	−0.079	0.1075	0.0378	78
St Ann's Head	9.299	1	49.7	58	−0.169	0.0052	0.0052	118

from RAPDRPT2 by up to 199 bp (based on the original cloned product, Figs 2 & 3). PCR reactions containing representative bands were cleaned using the QIAQUICK PCR clean-up kit (QIAGEN) and eluted in 30 µl. From this, 2 µl were subsequently cloned into Promega's pGEM-T EASY vector and transformed into JM109 competent cells. Plasmids from positive colonies were extracted by alkaline lysis and sequenced using automated sequencing techniques with *Taq* terminator mix on an ABI373. Sequencing was undertaken with standard sequencing primers (M13-20 and M13R).

RESULTS

Organisation and nature of the repeat

Identity searches using BLASTn and BLASTx on sequences flanking the repeat failed to reveal any significant matches; thus there is, as yet, no indication as to whether this repeat is associated with any known functional gene. The original sequence from the cloned RAPD band (Fig. 2) indicated that the repeat was not perfect, with a point mutation changing 1 GTT to TTT. Sequencing of repeat amplicons from a number of additional individuals confirmed that these also contained the triplet motif but that, although the GTT repeat is often perfect, in some instances a variety of interruptions are recognised, including apparent point mutations and single base pair deletions (Fig. 2). No perfect repeats have been found for alleles >400 bp.

Comparison with Hardy-Weinberg expectations

Because alleles had not been sized exactly, precise genotype frequencies were not available. Thus, comparison to H-W expectations using the raw data was not possible. In order to make a comparison, alleles were 'binned' into 5 bp classes (95.001 to 100 bp, 100.001 to 105 bp etc.) and each 'binned' group classed as an allele. Only 3 samples that were originally scored as heterozygotes needed to be designated as homozygous in this analysis. GENEPOP was then utilised to calculate heterozygosity and F_{IS} statistics (Table 2). When individuals with single bands were classified as homozygotes, no population conforms to H-W equilibrium, with the exception of the Folkestone *Littorina saxatilis* M and St Margaret's at Cliffe *L. saxatilis* M. This is due to a deficiency of heterozygotes in each population, and even the Folkestone and St Margaret's at Cliffe *L. saxatilis* M had non-significant deficiencies of heterozygotes. This deficiency of heterozygotes could be due to the presence of null alleles (Hare et al. 1996). To check for this, the H-W comparison was repeated with samples with single bands classed as heterozygotes with the null alleles, and individuals with no amplification as null homozygotes. In this analysis no test is significant, thus all populations are in H-W equilibrium (Table 2).

When *Littorina saxatilis* H and M are pooled, and the analysis repeated, assuming single bands represent null allele heterozygotes, all populations remain in H-W equilibrium. Thus, there is no evidence of a Wahlund effect.

Frequency of size fractions in *Littorina saxatilis* H and M

Amplification with RAPDRPT1 and RAPDRPT2 produced 1 or 2 bands in the range of 95 to 675 bp. When only a single band was amplified from an individual, this was initially assumed to represent the homozygous condition, and size frequency distributions were calculated after accounting for both bands of each heterozygous individual and scoring every assumed homozygous band twice (Fig. 4A). Comparisons of the repeat size distribution between the 2 forms were then made at 3 scales: between *Littorina saxatilis* H and M over all samples; between populations of H and M at individual sites; and between replicate samples of H and M at 1 site (Nynian's Cave, Galloway). Over all 3 scales, amplification products from *L. saxatilis* H have a higher median size than those from *L. saxatilis* M (Table 3), except at Ballynahown. However, some comparisons are non-significant. Because of uncertainties concerning the nature of some of the bands in the >400 bp size range (see below), additional comparisons of *L. saxatilis* H and M were undertaken using only bands smaller than 300 bp. In most cases this gave an identical outcome, thus demonstrating that, regardless of the high frequency of larger alleles in *L. saxatilis* H, alleles are still, on average, of a larger size in this form.

When individuals with only a single band were counted as heterozygotes with the null allele, the size frequency distribution was similar in shape (Fig. 4B) although the number of observations was curtailed, thus diminishing the power of the comparative tests. Nevertheless, Mann-Whitney comparisons produced a similar outcome to the first analysis (individuals with 1 band counted as homozygotes), although now, despite the samples of *Littorina saxatilis* H having a consistently larger median allele size, fewer comparisons were significant (Table 3).

Incidence and prevalence of large (>400 bp) alleles

Most alleles of the triplet repeat were between 95 and 300 bp. However, some were >400 bp. Two classes of large alleles (>400 bp) were recognised. One type of product found within the 400 to 675 bp range was of a similar intensity and diffuseness to the recognised triplet repeat products of a smaller size: whilst the other (>600 bp; not shown in Fig. 4 and not included in analyses) was much rarer, and produced a very tight, compact band when run on an agarose gel. Sequencing of representative bands revealed that, whilst the bands of 'normal' intensity were large repeat alleles, the tight bands had been primed by RAPDRPT2 at both ends and sequences did not contain the triplet repeat, nor any recognisable sequence from the flanking region of typical alleles. Two of these have been cloned and their ends sequenced. BLAST searches revealed no homology matches for one, whilst the other had a strong match to the sequence of ORS571 from *Azorhizobium* (Kaplowski et al. 1991). Attempts to prevent amplification of these non-specific products using alternatives to RAPDRPT2

(RAPDRPT2B and RAPDRPT2C) did not stop amplification. Since these bands were recognisable, of apparently low frequency, and could be discarded from further analyses, all results are based on continued amplification with primer pair RAPDRPT1 and RAPDRPT2. In some populations, the 'normal' alleles of size >400 bp were found at quite high frequency, and were particularly prevalent in *Littorina saxatilis* H (see Table 3).

Sequence of flanking regions in the 2 morphs

In some instances, regions flanking microsatellites have been shown to harbour variation which can aid in, for example, detecting homoplasy of the repeat (Grimaldi & Crouau-Roy 1997). To investigate whether the regions flanking the GTT repeat exhibited differences between *Littorina saxatilis* H and M, a number of products were sequenced. Flanking regions from 13 samples together with the original clone are presented in Fig. 2. Of the 199 bp, 10 sites are variable due to point mutations. There are also 3 indel events, of 1, 4 and 21 bp. For the 13 flanking region sequences shown, there is no evidence that any of these variable positions are linked to the form of *L. saxatilis* from which they were amplified. The alleles greater than 400 bp had several commonalities. Firstly, at position 142 these sequences have a G in contrast to an A in smaller alleles. These large alleles also have a common core repeat sequence (Fig. 1), indicating that these large alleles are not independently derived from independent expansions from smaller alleles. For example, although MH485, BM400, BM446, GH422 and HH422 may differ in the numbers of certain triplets, there is essentially the same pattern of repeat types.

Table 3. Results of Mann-Whitney test for comparison of repeat allele sizes between *Littorina saxatilis* H and M. At Galloway, comparisons were made between morphs in a single collection in 1997 (1), in collections made in 1998 from 4 separate transects (2 to 5) and combining all these data. Four analyses are presented for each population, for all allele sizes, and for sizes <300 bp. Each was undertaken for 2 data treatments. In the first analysis, single bands were assumed to represent homozygous individuals. In the second analysis, single bands were assumed to represent the heterozygous condition with a null allele. Sample size (alleles) in parentheses. Levels at which tests are significant (p) are reported for each comparison; significance after correction for multiple tests is portrayed with an asterisk (α = 0.05). Bonferroni correction was done separately for 'All data' analyses and for '<300 bp' analyses with 'All populations', between groups at the 8 sites, and between the 5 replicates at Galloway treated as different families (Hochberg & Tamhane 1987)

Population	Sample	Median M	Median H	p	Population	Sample	Median M	Median H	p
Single band = homozygote					**Single band = heterozygote with null allele**				
All populations	All	125.46 (390)	167.70 (372)	0.0000*	All populations	All	126.48 (324)	163.62 (271)	0.0000*
	<300	124.49 (375)	143.14 (296)	0.0000*		<300	125.13 (312)	142.64 (224)	0.0000*
Galloway	All	128.44 (98)	181.71 (96)	0.0000*	Galloway	All	129.39 (87)	174.22 (77)	0.0000*
	<300	128.44 (98)	179.10 (93)	0.0000*		<300	129.39 (87)	173.47 (75)	0.0000*
1	All	126.48 (20)	175.45 (16)	0.0041*	1	All	126.99 (18)	175.40 (12)	0.0141*
	<300	126.48 (20)	159.80 (15)	0.0077*		<300	126.99 (18)	159.80 (11)	0.0276
2	All	120.64 (20)	229.61 (20)	0.0002*	2	All	122.57 (16)	222.02 (15)	0.0006*
	<300	120.64 (20)	229.61 (20)	0.0002*		<300	122.57 (16)	222.02 (15)	0.0006*
3	All	138.39 (18)	170.56 (20)	0.0252	3	All	137.00 (16)	167.65 (17)	0.0243*
	<300	138.39 (18)	170.56 (20)	0.0252		<300	137.00 (16)	167.65 (17)	0.0243
4	All	132.25 (20)	182.59 (20)	0.0005*	4	All	132.25 (20)	182.59 (15)	0.0018*
	<300	132.25 (20)	182.59 (20)	0.0005*		<300	132.25 (20)	182.59 (15)	0.0018*
5	All	128.44 (20)	153.14 (20)	0.0719	5	All	128.76 (17)	153.14 (18)	0.0477*
	<300	128.44 (20)	143.11 (18)	0.1737		<300	128.76 (17)	149.60 (17)	0.0761
The Mumbles	All	118.03 (148)	154.14 (148)	0.0000*	The Mumbles	All	119.59 (125)	145.76 (95)	0.0000*
	<300	117.93 (145)	137.40 (94)	0.0000*		<300	118.12 (123)	137.13 (64)	0.0000*
Langland Bay	All	118.70 (40)	326.30 (20)	0.0000*	Langland Bay	All	119.54 (34)	184.30 (13)	0.0013*
	<300	118.69 (40)	129.59 (10)	0.0941		<300	119.54 (34)	129.59 (7)	0.2528
Old Peak	All	154.20 (28)	169.58 (28)	0.0012*	Old Peak	All	154.4 (18)	167.50 (22)	0.0223
	<300	154.20 (27)	167.46 (26)	0.001*		<300	154.20 (17)	166.05 (20)	0.0174
Folkestone	All	161.66 (18)	202.91 (20)	0.0659	Folkestone	All	161.66 (14)	202.91 (15)	0.1538
	<300	161.66 (18)	202.91 (20)	0.0659		<300	161.66 (14)	202.91 (15)	0.1538
St Margaret's at Cliffe	All	156.56 (20)	180.27 (20)	0.0135*	St Margaret's at Cliffe	All	156.56 (18)	162.16 (15)	0.0951
	<300	156.56 (20)	180.27 (20)	0.0135		<300	156.56 (18)	162.16 (15)	0.0951
Ballynahown	All	130.70 (38)	129.70 (40)	0.9084	Ballynahown	All	135.00 (28)	129.70 (34)	0.6204
	<300	124.32 (27)	127.70 (33)	0.4217		<300	124.32 (19)	126.78 (28)	0.6644
		Slope bottom	Slope top				Slope bottom	Slope top	
St Ann's Head	All	127.97 (58)	135.33 (58)	0.0036*	St Ann's Head	All	127.97 (42)	135.08 (35)	0.0381
	<300	127.97 (57)	134.06 (50)	0.0459		<300	127.97 (41)	133.94 (31)	0.1270

DISCUSSION

Genetic differentiation of *Littorina saxatilis* H and M

If *Littorina saxatilis* H and M share a common gene pool then the size distribution of repeat motifs should be essentially homogeneous when the 2 forms are compared. However, if there are differences detectable between H and M that are due only to genetic drift then when samples from different positions on the shore are compared, these differences should not be of a predictable nature (assuming that the repeat variation is neutral or that there is no significant linkage disequilibrium with a selected locus). In reality, homogeneity of the repeat is not observed. In addition, wherever *L. saxatilis* H and M are examined (with the exception of the Ballynahown population showing an uninvestigated position reversal of H and M), *L. saxatilis* H have, on average, larger repeats than *L. saxatilis* M, and in many cases this is a significant difference. This is irrespective of how these data are treated with respect to the possibility of null alleles in the data set. This pattern occurs at every shore examined (with the exception of Ballynahown) and consistently at a site where replicate transects have been taken. Where such replicate sampling of a single shore has been undertaken (Galloway), there are consistent differences between *L. saxatilis* at the top of the shore (H) and those at the bottom (M). Thus, the differences are not due to unpredictable microgeographic variation, as has been found for differences detectable between samples of *L. saxatilis* examined by allozyme electrophoresis (Janson & Ward 1984), but to real differences between the 2 forms. In addition to the difference in overall size frequency distribution of this repeat, there is also a difference in prevalence of large alleles (>400 bp). These alleles are found in high frequency in certain populations of *L. saxatilis* H, but only rarely in *L. saxatilis* M.

At Ballynahown (Ireland), *Littorina saxatilis* M is found higher up the shore than *L. saxatilis* H. This position reversal is unique amongst the shores hitherto examined. At this site, the repeat size in the M form is not significantly different from that of the H form. Because of the unique nature of this site, we do not know how this affects our findings.

These data raise important questions over how this repeat allele distribution has arisen. Two options are plausible—that gene flow is greater within, than between, morphs, or that there is selection on the repeat or a locus linked to it. Both possibilities can be accommodated, and neither is proven with these data. However, on the balance of evidence, we suggest that the latter is more likely. Firstly, because if restricted gene flow were the cause, then combining H and M data for each population would result in a Wahlund effect detectable by non-conformity to H-W equilibrium; this is not seen. Secondly, gene flow between distant populations will be very rare for these low-vagility, ovoviviparous animals. Indeed, the differences in average repeat size, even between close populations (e.g. Folkestone M and St Margaret's at Cliffe M) suggests that gene flow within morphs is not sufficient to homogenise allele frequencies. Despite this, the same pattern of repeat distribution (with greater average size of the repeat in *L. saxatilis* H compared to M) is seen on different coasts where these microsatellite data, and mtDNA frequencies (Wilding et al. 2000), have shown gene flow to be restricted. A selection argument can be more readily entertained. In the early phase of non-allopatric speciation, models predict that strong disruptive selection can result in divergence of a small part of the genome linked to the locus/loci on which the disruptive selection acts. We have demonstrated that in *L. saxatilis*, 5% of loci detected by amplified fragment length polymorphism (and by inference 5% of the genome) seem to be influenced by selection, although the magnitude of differentiation varies between populations, presumably influenced by the strengths of disruptive selection (Wilding et al. 2001). These GTT allele data fit well with this system. Here, the locus could be under selection, but more likely is linked to one that is, and differing strengths of selection result in differing degrees of differentiation. This would explain why not all populations show the same degree of difference. This selection argument is not at odds with the observation of distinct geographic structure of mtDNA, nor with the data from St Ann's Head, a unique site with *L. saxatilis* H only found on a steep slope.

The considerable deficiencies of heterozygotes in some populations are a feature of this analysis, the basis of which is not yet fully understood. A high frequency of null alleles (Hare et al. 1996) is a potential source, and since the 4 bp indel event does fall towards the 5′ end of the RAPDRPT2 priming site, this could explain the existence of null alleles. Modifications occurring within the priming sites have resulted in null alleles reported in other studies (Grimaldi & Crouau-Roy 1997). However, this seems an unlikely explanation in this study as very few samples produced no amplification products that could be attributable to null-null homozygotes (5 individuals compared to 439 where amplification succeeded). Allele dropout, whereby 1 allele of a heterozygote does not amplify (Pemberton et al. 1995) could produce such a pattern, and has been observed in microsatellite surveys (e.g. Gerloff et al. 1995). However, allele dropout is typically only a problem when low quantities of template DNA are used (Taberlet et al. 1996). In our study, DNA was not limiting. Whatever the cause of this heterozygote defi-

ciency, it is not limited to either *Littorina saxatilis* H or M and so should not affect the finding that the 2 forms have different size frequency distributions.

Evolution of the repeat

The mechanism of evolution of the repeat does not seem to fit with the standard slippage-mispairing model (Freimer & Slatkin 1996). The size frequency distribution of repeat allele sizes has a number of peaks, but these are confined to 2 major regions at 95 to 260 and 400 to 525 bp. The effective absence of repeat sizes between these regions is not compatible with mutation of only single repeat units. Valdes et al. (1993) have shown that the stepwise mutation model is capable of producing distributions with peaks instead of smooth distributions, but this model would not explain the jump in repeat size from 260 to 400 bp. However, Di Rienzo et al. (1994) suggested that these kinds of distributions were more aptly modelled by a 2-phase process in which most mutations involve changes of 1 repeat unit, but where occasional larger changes in repeat number can also occur. Such a process may explain how the large shift between alleles of 95 to 260 bp and those alleles over 400 bp in size could occur. Similar allelic distributions have been noted previously. Imbert et al. (1993), analysing a CTG repeat within a putative protein kinase locus causing myotonic dystrophy, have shown that the disease is predisposed when the typical $(CTG)_5$ allele mutates to an allele with 19 to 30 repeats. Recently, Davison (1999) has demonstrated differences in allele sizes of over 150 repeat units for the gastropod mollusc *Cepaea nemoralis*. Thus, large transitions in repeat number within triplet repeats are known and the 2-phase model of Di Rienzo et al. (1994) may be appropriate for the GTT locus analysed in the current study.

The repeat is often, but not always, perfect. In some instances, the sequence of the repeat shows mutations due to point mutations and/or deletions. It is assumed that these are not errors introduced during PCR amplification (Ennis et al. 1990). Such mutations may be capable of stabilising the repeat through prevention of polymerase slippage (Pépin et al. 1995). Whether the occurrence of such interruptions has driven the distribution of the repeat sizes to the observed size frequency distribution is uncertain.

LITERATURE CITED

Ashburner M (1989) *Drosophila*: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Davison A (1999) Isolation and characterisation of long compound microsatellite repeat loci in the land snail, *Cepaea nemoralis* L. (Mollusca, Gastropoda, Pulmonata). Mol Ecol 8:1760–1761

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Ennis PD, Zemmour J, Salter RD, Parham P (1990) Rapid cloning of HLA-A, B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc Natl Acad Sci USA 87:2833–2837

Freimer NB, Slatkin M (1996) Microsatellites: evolution and mutational processes. In: Chadwick D, Cardew G (eds) Variation in the human genome. Wiley & Son Ltd, Chichester, p 51–67

Gerloff U, Schlötterer C, Rassmann K, Rambold I, Hohmann G, Fruth B, Tautz D (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). Mol Ecol 4: 515–518

Grahame J, Hull SL, Mill PJ, Hemmingway R (1997) Discovering unrecognised diversity among marine molluscs. In: Ormond RFG, Gage JD, Angel MV (eds) Marine biodiversity: patterns and processes. Cambridge University Press, Cambridge, p 293–318

Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequence. J Mol Evol 44:336–340

Guo SW, Thompson EA (1992) A Monte-Carlo method for combined segregation and linkage analysis. Am J Hum Gen 51:1111–1126

Hare MP, Karl SA, Avise JC (1996) Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. Mol Biol Evol 13:334–345

Heller J (1975) The taxonomy of some British *Littorina* species with notes on their reproduction (Mollusca: Prosobranchia). Zool J Linn Soc 56:131–151

Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75:800–802

Hochberg Y, Tamhane A (1987) Multiple comparison procedures. Wiley, New York

Hull SL (1998) Assortative mating between two morphs of *Littorina saxatilis* on a shore in Yorkshire. Hydrobiologia 378:79–88

Hull SL, Grahame J, Mill PJ (1996) Morphological divergence and evidence for reproductive isolation in *Littorina saxatilis* (Olivi) in northeast England. J Molluscan Stud 62:89–99

Imbert G, Kretz C, Johnson K, Mandel JL (1993) Origin of the expansion mutation in myotonic dystrophy. Nat Genet 4: 72–76

Janson K, Ward RD (1984) Microgeographic variation in allozyme and shell characters in *Littorina saxatilis* Olivi (Prosobranchia: Littorinidae). Biol J Linn Soc 22:289–307

Kaplowski K, Klosse U, Brujn FJ (1991) Characterization of a novel *Azorhizobium caulinodans* ORS571 two-component regulatory system, NtrY/NtrX, involved in nitorgen fixation and metabolism. Mol Gen Genet 231:124–138

Maton WG (1797) Observations relative chiefly to the natural history, picturesque scenery, and antiquities, of the western counties of England, made in the years 1794 and 1796, Vol 1. Salisbury

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Non-amplifying alleles at microsatellite loci—a caution for parentage and population studies. Mol Ecol 4:249–252

Pépin L, Amigues Y, Lépingle A, Berthier JL, Bensaid A, Vaiman D (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74:53–61

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

Raffaelli D (1979) The taxonomy of the *Littorina saxatilis* species-complex, with particular reference to the systematic status of *Littorina patula* Jeffreys. Zool J Linn Soc 65: 219–232

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Reid DG (1996) Systematics and evolution of *Littorina*. The Ray Society, London

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V,

Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res 24:3189–3194

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Wilding CS, Grahame J, Mill PJ (1998) Rough periwinkle polymorphism on the east coast of Yorkshire: comparison of RAPD-DNA data with morphotype. Hydrobiologia 378: 71–78

Wilding CS, Grahame J, Mill PJ (2000a) Mitochondrial DNA CoI haplotype variation in sibling species of rough periwinkles. Heredity 85:62–74

Wilding CS, Grahame J, Mill PJ (2000b) Nuclear DNA restriction site polymorphisms and the phylogeny and population structure of an intertidal snail species complex (*Littorina*). Hereditas 133:9–18

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. J Evol Biol 14:611–619