Intron-length polymorphism at the actin gene locus mac-1: a genetic marker for population studies in the marine mussels *Mytilus galloprovincialis* Lmk. and *M. edulis* L.

Marc Ohresser
Laboratoire Génome et Populations, UPR CNRS 9060, Génétique des Populations marines, Station Méditerranéenne de l'Environnement Littoral, Quai de la Daurade, 34200 Sète, France

Philippe Borsa
ORSTOM-Zoologie, B.P. 5045, 34032 Montpellier cedex 1, France, and Laboratoire Génome et Populations, UPR CNRS 9060, Station Méditerranéenne de l'Environnement Littoral, Quai de la Daurade, 34200 Sète, France

Claude Delsert*
IFREMER, Laboratoire de Genétique, Aquaculture et Pathologie, BP 133, 17390 La Tremblade, France, and Laboratoire Génome et Populations, UPR CNRS 9060, Station Méditerranéenne de l'Environnement Littoral, Quai de la Daurade, 34200 Sète, France

Abstract

A novel intron-length polymorphism at the actin gene locus mac-1 is here reported and used as a genetic marker for population studies in mussels of the genus *Mytilus*. Two closely related genes subsequently identified as alleles, *mac-la* and *mac-lb*, from a genomic library of *M. galloprovincialis* were partially cloned and sequenced. They mainly differed from each other by a 65-bp insertion within their first intron. Polymerase chain reaction (PCR) primers were designed outside the insertion. The PCR analysis of 166 individual mussels from *M. galloprovincialis* and *M. edulis* populations revealed three size-classes of alleles or allelomorphs, two of which were of the expected sizes for *mac-la* and *mac-lb*. One allelomorph was absent from *M. edulis* samples, although it was present at substantial frequencies in *M. galloprovincialis* populations. The frequencies of the two other allelomorphs significantly differed between *M. galloprovincialis* and *M. edulis* populations. The comparison of six mac-1 intron sequences over 277 bp showed at once that allelomorphs encompassed alleles differing from one another by substantial numbers of mutations, and that identical alleles were present in both *M. galloprovincialis* and *M. edulis* individuals, a probable result of the recent introgression between the two species.

Introduction

Genetic studies of natural populations usually involve large sets of individual genotypic data, either inferred from individual phenotypes with known genetic determination, or directly obtained by DNA typing. Molecular markers for population studies are particularly useful if they fulfill the following criteria: they should exhibit a reasonable amount of polymorphism; be selectively neutral; be codominant for direct scoring of diploid genotypes, be inherited, then, in a Mendelian fashion; and from a practical point of view, they should be cost- and time-effective.

The polymerase chain reaction technique (PCR; Saiki et al., 1988) has provided population geneticists with a powerful yet simple and cost-effective method for amplification of DNA fragments (Skibinski, 1994; Bonhomme et al., 1995). PCR enables multiple-locus genotyping from minute tissue samples including microbiopsies and larvae (for such examples in marine bivalves, see Côté-Real et al., 1994a), and also from alcohol-preserved or dried specimens.

As a preliminary step to PCR, specific oligonucleotides are deduced from the targeted DNA sequence and serve as primers during the amplification. Determination of nucleotide sequence often necessitates the construction and screening of a genomic DNA library although this step can be circumvented by using the amino acid sequences of proteins known for their high degree of phylogenetic conservation. This technique allows the use of intron poly-
morphism as a genetic marker, provided the intron is limited in size. First, degenerate oligonucleotides have to be deduced from the amino acid sequences encoded by the two exons flanking the targeted intron and used for PCR amplification. Then, following sequence determination of the amplified DNA fragment, new specific primers are designed. This approach (Lessa, 1992) has already been successfully used, in particular, to provide molecular markers in mussels (Côrte-Real et al., 1994b).

The intron-length polymorphism reported for the *Mytilus calmodulin CaM-1* gene (Côrte-Real et al., 1994a) revealed only weak differences in allelic frequencies between samples from various localities in Europe, including the British Isles and the Adriatic Sea. Another intron-length polymorphism was visualized from an internal transcribed spacer between the 18S and 28S nuclear rDNA coding regions in *M. californianus* and members of what has been referred to as the *M. edulis* complex (*M. edulis*, *M. galloprovincialis*, and *M. trossulus*) (Heath et al., 1996). In the same study, the coding sequence at the sperm-specific protein PHI-1 locus was found to be polymorphic between two species groups and monomorphic within each group. Another, better-characterized sequence polymorphism at a locus encoding byssal adhesive proteins was shown to occur intraspecifically and interspecifically in the genus *Mytilus* (Inoue et al., 1995). Other markers, such as microsatellite sequences (Weber and May, 1989), have been developed in the flat oyster *Ostrea edulis* (Naciri et al., 1995).

Length polymorphism at a noncoding intron locus is likely to be selectively neutral (Lessa, 1992), which makes it potentially useful for the analysis of population genetic structure. Alleles can be easily characterized on agarose gel according to their size, which allows the analysis of large samples, whereas other types of markers including some of those mentioned above usually require restriction mapping or DNA sequencing of the PCR-amplified fragment. Moreover, the sequences of PCR-amplified fragments that differ in size can be obtained by direct sequencing after they have been purified from agarose gels.

Here we report the identification of an intron of variable length in one of the actin genes of marine mussels of the genus *Mytilus, M. galloprovincialis* and *M. edulis*. A mussel genomic DNA library was constructed and screened for actin genes. Sequence analysis allowed the identification of two clones of the same actin gene differing by the presence of an insert in their first intron. PCR amplification was done on the DNA of individual *M. galloprovincialis*, using primers designed in the regions flanking the insert. Analysis of the products of this amplification by electrophoresis on agarose gel revealed the existence of three size-classes of alleles. The distribution of individual phenotype frequencies in *M. galloprovincialis* and *M. edulis* natural populations did not depart from the expectations of the Hardy-Weinberg model, suggesting that the size polymorphism at this locus is Mendelian. One size class of alleles (allelomorph) was present at a substantial frequency in *M. galloprovincialis* populations but totally absent in *M. edulis* samples, while frequencies of the two other allelomorphs significantly differed between populations of the two species. We believe that this novel genetic marker is of much value for population studies in marine mussels of the genus *Mytilus*.

**Results**

**Sequence analysis of mac-1 intron 1**

We isolated two closely related actin genes, *mac-1a* and *mac-1b*. The 5′ end region of the first intron of each gene was sequenced (Figure 1). The two sequences differed by single-nucleotide or dinucleotide substitutions and by a large 65-bp insertion (Figure 1). The insertion in *mac-1b* consists of a 57-bp fragment flanked by the direct repeat of the octanucleotide 5′-AATCTGCT-3′ that is present as a single element in *mac-1a* (Figure 1). The rate of nucleotide substitution outside the insertion was about 1.8%, and one- or two-nucleotide gaps were present at a rate of about 0.9%.

**Allelism of Intron sequences**

In order to determine whether *mac-1a* and *mac-1b* were alleles, we PCR-amplified the genomic DNA of individual mussels using primers located in the region outside the insert (Figure 1). If *mac-1a* and *mac-1b* were alleles, no more than two PCR products would ever be characterized in an individual (because mussels are diploid; Thiiriot-Quivréx and Ayraud, 1982); the expected phenotypes might be DNA fragments corresponding to either *mac-1a* (324 bp) or *mac-1b* (388 bp), or both, or combinations of either with fragments of other sizes. Preliminary PCR runs on 28 *M. galloprovincialis* fit the allelic model, which was further supported by the subsequent analysis of larger samples of individuals.
Figure 1. Sequence polymorphism of Intron 1 of the *Mytilus galloprovincialis* actin gene *mac-1*. (A) 5' to 3' nucleotide sequence alignment for *mac-1* alleles *a*, *b*, and *c*. *mac-a*, and *mac-b*, were isolated from a mussel genomic DNA library; *mac-c*, was PCR-amplified. Coordinates (in bp) are relative to the *mac-a* translation start site. Dots indicate matches; Dashes indicate gaps; Bold lower cases for the octonucleotide direct repeat; Arrows underline the PCR primer regions. (B) Simplified representation of *mac-1* alleles. The corresponding DNA fragments were PCR-amplified using primers 1 (forward) and 2 (reverse). Numbers in bp are the sizes of *mac-a*, *mac-b*, and, presumably, *mac-c*, and also correspond to approximate sizes of other *mac-a*, *mac-b* and *mac-c* alleles, respectively. Open rectangle corresponds to the 65-bp insert; Thick line corresponds to the 28-bp deletion; Arrowheads correspond to the octonucleotide direct repeat.

as detailed below. This preliminary step also revealed the existence of a third size-class of alleles (allelomorph) at approximately 360 bp (Figure 2), hereafter called *mac-c*. One *mac-c* allele, *mac-c*, was subsequently sequenced (Figure 1) using the direct sequencing method on the purified PCR product of one heterozygous individual. It differed from *mac-b*, by a 28-bp deletion, 3 bp upstream from the 65-bp insertion, and by nucleotide substitutions and short gaps at rates of about 0.7% and 1.0%, respectively. The 28-bp deletion removed one of the duplicated octonucleotides flanking the 65-bp insert.

**Sequence polymorphism**

Another three bands, corresponding to the PCR products of heterozygous individuals each having one allele of each of the three allelomorphs detected so far, *mac-a*, *-b*, and *-c*, were excised from agarose gels, purified, and sequenced using the direct sequencing method. In total, 6 *mac-1* intron 1 sequences were obtained. Nucleotide differences
Variability within and between populations

Electrophoresis on agarose gel of the PCR products of 166 individual *Mytilus* DNA extracts from four populations revealed either one- or two-banded phenotypes with each band of approximately the expected size for *mac-la*, *mac-lb*, or *mac-1c* (Figure 2). These individual patterns always were in accordance with the co-dominant allelic model. Allelomorphic frequencies and estimated gene diversities in each population are reported in Table 1. Individual phenotype frequencies in three of four samples (all except Polzeath Bay) did not depart from the expectations of the Hardy-Weinberg model (see f column in Table 2), suggesting (although not proving) that the size polymorphism at locus *mac-1* may be Mendelian and selectively neutral, and that the samples represent panmictic or nearly panmictic populations.

The overall θ value between samples from different regions was highly significant, reflecting strong differences between populations (Table 2, “Europe”). A small fraction of this structure may be explained by differences between subpopulations as inferred from the θ value estimated at the regional scale (Table 2, “Charente”). Within-subpopulation structure was not evident, but the moderately strong (yet not significant) θ value in the Polzeath Bay sample may reflect some level of heterozygote deficit in the population (Table 2).

Discussion

The length polymorphism reported here provides a new nuclear marker to discriminate between *Mytilus galloprovincialis* and *M. edulis* populations. This is an interesting feature in view of the thorny systematics of the genus *Mytilus* (Skibinski and Beardmore, 1979; Varvio et al., 1988; Beaumont et al., 1989), of the process of introgression in areas where the two species hybridize (Skibinski et al., 1978; Skibinski and Beardmore, 1979; Skibinski et al., 1983; Coustau, 1991), and of the selective processes inferred to affect for example, edulis-like genomes in some areas of overlap (Gardner and Skibinski, 1990; Coustau, 1991). The forelisted studies have made extensive use of mainly two discriminatory enzyme loci. The selective neutrality of allozyme markers, however, has been questioned (e.g., Avise, 1994; but see Skibinski and Ward, 1982). Also, addressing the phylogenetic relations among electromorphs at an enzyme locus is a difficult task, as opposed to nucleotide sequencing of nuclear DNA fragments that can be PCR-amplified. Nevertheless, the use of different classes of genetic mark-
Table 1. *Mytilus galloprovincialis* and *M. edulis*. mac-1 Intron 1 allelomorph frequencies in four populations.

<table>
<thead>
<tr>
<th>Allelomorph</th>
<th>Size (bp)</th>
<th>Boyard</th>
<th>Aiguillon</th>
<th>Sète</th>
<th>Polzeath</th>
</tr>
</thead>
<tbody>
<tr>
<td>mac-1a</td>
<td>324</td>
<td>0.83</td>
<td>0.95</td>
<td>0.12</td>
<td>0.53</td>
</tr>
<tr>
<td>mac-1b</td>
<td>388</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
<td>0.09</td>
</tr>
<tr>
<td>mac-1c</td>
<td>360</td>
<td>0.17</td>
<td>0.05</td>
<td>0.61</td>
<td>0.38</td>
</tr>
<tr>
<td>(N)</td>
<td></td>
<td>(15)</td>
<td>(33)</td>
<td>(59)</td>
<td>(59)</td>
</tr>
<tr>
<td>h*</td>
<td></td>
<td>0.29</td>
<td>0.10</td>
<td>0.15</td>
<td>0.58</td>
</tr>
</tbody>
</table>

N, sample size; h*, gene diversity, with correction for individuals' sampling bias.

Table 2. *Mytilus galloprovincialis* and *M. edulis*. Summary statistics on genetic structure at Locus mac-1, at two geographic scales.

<table>
<thead>
<tr>
<th>Geographic scale, population</th>
<th>(N)</th>
<th>f</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charente</td>
<td></td>
<td>0.072*</td>
<td></td>
</tr>
<tr>
<td>Boyard</td>
<td>(15)</td>
<td>-0.167</td>
<td></td>
</tr>
<tr>
<td>Aiguillon</td>
<td>(33)</td>
<td>-0.033</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td></td>
<td>0.386***</td>
<td></td>
</tr>
<tr>
<td>Charente</td>
<td>(48)</td>
<td>-0.080</td>
<td></td>
</tr>
<tr>
<td>Sète</td>
<td>(59)</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Polzeath</td>
<td>(59)</td>
<td>0.171</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05, Fisher's exact test on allelic frequencies

***p < 0.001, test on 1000 permutations

Population genetic studies of marine mussels, and of marine bivalves in general, have often revealed heterozygote deficits in comparison with Hardy-Weinberg expectations at allozyme loci (e.g., Zouros and Foltz, 1984; Borsa et al., 1991; Gaffney, 1994) and other nuclear markers (Karl and Avise, 1993). Whenever possible, crossing experiments should be performed to test for the Mendelian inheritance of the nuclear markers used in genetic studies of natural populations. However, for marine bivalves such crosses can be difficult to control and costly. Still, some uncertainties about the causes of departure from Mendelian expectations have been reported in the progenies of experimental crosses, at allozyme, microsatellite, and other noncoding nuclear DNA loci (Beaumont et al., 1983; Hvislom and Theisen, 1984; Foltz, 1986; Borsa, 1990; Côrte-Real et al., 1994a). These were attributed to null alleles or to selection before the metamorphosis of larvae. Here, both the Mendelian inheritance and selective neutrality of the size polymorphism at locus mac-1 were suggested (although not proven) by the general lack of departure from Hardy-Weinberg expectations in the samples.

Heterozygote deficits were possibly present in the Polzeath Bay sample. If this proves to be the correct inference from the comparatively high f value estimated here (a larger sample size would...
be desirable, though, for a more powerful test), then
these could be best explained by some admixture
of *M. galloprovincialis* with individuals bearing *M.
edulis* mac-1 genes. Some observations may support
this hypothesis: the occurrence of typical *M. edulis*
and presumably backcrossed populations in the vi-
cinity of Polzeath Bay (Skibinski et al., 1983; Bea-
umont et al., 1989); the presumed occurrence of
edulis-like individuals within samples from Pol-
zeath Bay (Beaumont et al., 1989); and the allelo-
morph frequency values at Polzeath Bay, which are
intermediate between those of typical *M. galloprov-
incialis* and typical *M. edulis* populations (Table 1).

From the sequences of mac-1 haplotypes ana-
alyzed so far, we found substantial, hence presum-
ably old, sequence divergence between alleles
within an allelomorph (mac-1a, of *M. galloprovin-
cialis* and mac-1a, of *M. edulis*.) Also, identical al-
leles (mac-1c) occur in both species, probably as a
result of recent introgression. No further conclusion
can be drawn from these limited data. We expect
that future in-depth phyleogeographic studies in-
volving the sequencing of large numbers of mac-1
alleles in samples of individuals from different *M.
galloprovincialis* and *M. edulis* populations will re-
veal the apportion of ancestral polymorphism ver-
sus recent introgression in the genus *Mytilus*.

**Experimental Procedures**

*Isolation, cloning and sequencing of a mussel actin
gene*

Genomic DNA from 10 mussels, *Mytilus galloprov-
incialis* from Sète, France, was partially digested
with the restriction enzyme Sau3A and cloned into
phage λ FIX2 (Stratagene Cloning Systems, San
Diego, Calif., U.S.A.). Screening was done with a
radioactively labeled probe corresponding to the 5’
end region of the coding sequence of the cytoskele-
tic actin A3 gene of the silkworm *Bombyx mori*
(Mounier and Prudhomme, 1986).

Positive clones were subcloned into plasmid
pBluescript (Stratagene Cloning Systems). Double-
strand sequence analysis (using the T7 sequencing
kit; Pharmacia Biotech, Upppsala, Sweden) of series
of subclones containing the 5’ end of the actin gene
revealed that one clone, here termed mac-1a, was
present in the library together with a closely related
but different clone, mac-1b. The sequences, encom-
passing the actin first exon and the first 500 bp of
the first intron (M. Ohresser and C. Delsert, unpub-
lished results), of mac-1a, and mac-1b, differed by
a 65-bp insertion, by single-base substitutions, and
by a few short gaps in the intron.

To determine whether these genes were allelic,
the DNA extracts of individual mussels were PCR-
amplified using primers chosen outside the inser-
tion and designed from the sequences of mac-1a,
and mac-1b, (Figure 1A). Four other PCR products
(one for each allelomorph, mac-1a and mac-1b, and
two for the newly identified mac-1c) of heterozy-
gous individuals were purified from agarose gels
and sequenced using the direct sequencing method
of Kusukawa et al. (1990).

**DNA extraction and PCR procedures**

Genomic DNA from single mussels was prepared
as follows. Approximately 1–3 mg of macerated ad-
ductor muscle tissue was taken from live or frozen
mussels using a hand micropipette, placed into an
Eppendorf microtube containing 5% chelexing resin
(Chelco, Biorad, Richmond, Calif., U.S.A.) in 500
μl 1 mM Tris/0.1 mM EDTA, pH 8.0 buffer, shaken,
and heated overnight at 55°C, then vortexed, heated
at 95°C for 15 minutes, vortexed again and centri-
fuged for 2 minutes at 10,000 × g. The supernatant
was stored at −20°C until it was used as template
for PCR.

Within-intron-specific PCR primers (5’-CATATT-
N-N-GAAACATCAAGTCG-3’ and 5’-CATATTTATG-
TTTCAGATGG-3’) were designed from the partial
sequences of intron 1 of mac-1a and mac-1b (Figure
1). Approximately 1 μl of DNA solution was added
to 12.5 μl PCR mixture containing 25 ng of each
primer, 2.5 mM MgCl₂, 0.2 mM of each dNTP, and
0.25 unit of Goldstar Taq polymerase (Eurogentec,
Liège, Belgium). Thirty cycles of PCR were run in a
Crocodile III thermocycler (Appligène, Strasbourg,
France) following a three-minute start at 96°C. The
consecutive steps within each cycle were 94°C for
1.5 minutes, 45°C for 1 minute, and 72°C for 0.5
minute, and the ramping time between steps was
set as minimal. PCR products were electrophoresed
on 1.5% agarose gel in parallel with 123-bp and
1-kb DNA ladders (Life Technology, Gaithersburg,
Md., U.S.A.) and were visualized and photographed
under ultraviolet light after staining with ethidium
bromide.

**Sampling**

Samples of mussels were collected along the water-
front at the Station Méditerranéenne de l’Environ-
ment Littoral in Sète, south France (June 1996; N
= 59 individuals) and on the rocks on the shore of Polzeath Bay on the northern shore of the Camel River estuary, southwest England (July 1996; N = 59). Samples of locally recruited, farm-produced mussels from Boyardville on the eastern shore of Oléron Island, west France (December 1995; N = 15) and from nearby Baie de l’Aiguillon (July 1996; N = 33) were purchased at the market. The morphology of the mussels from Sète and Polzeath Bay conformed to the *Mytilus galloprovincialis* morphotype, while those from Boyardville and Baie de l’Aiguillon conformed to the *M. edulis* morphotype (Lubet et al., 1984). Also, samples of mussels from Sète and Polzeath Bay have been referred to as *M. galloprovincialis* by Coustau (1991) and Beaumont et al. (1989), respectively.

**Analysis of data**

Two geographic scales were addressed for the analysis of population structure: the regional scale ("Charente," grouping subpopulations Boyardville and Baie de l’Aiguillon) and the Northeast Atlantic–Mediterranean scale ("Europe," grouping populations Charente, Sète, and Polzeath Bay). Wright’s (1951) fixation indices $F_{st}$ and $F_{is}$ were estimated using Weir and Cockerham’s (1984) estimators $f$ and $\theta$ and the permutation tests were made using the computer package *GENETIX* (Belkhir et al., 1996).

**Acknowledgments**

We are grateful to F. Bonhomme for initiating this project in Station Méditerranéenne de l’Environnement Littoral in Sète, to D. Hedgecock for helpful comments on the manuscript, to D. Jollivet and P. Le Maho for participating in the collection of samples, and to C. Coustau and M. Raymond for earlier discussions. We are indebted to A. Gérard for his support during the course of this work, which was supported by grant CT-2487 from the European Commission.

**References**


