Vol. 21: 25–36, 2014
doi: 10.3354/ab00566

AQUATIC BIOLOGY
Aquat Biol

Published online June 3

New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea

Małgorzata Zbawicka[1,*], Tomasz Sańko[1], Jakob Strand[2], Roman Wenne[1]

[1]Institute of Oceanology PAS, Powstańców Warszawy 55, 81-712 Sopot, Poland
[2]Arhus University, Department of Bioscience, 4000 Roskilde, Denmark

ABSTRACT: Environmental conditions such as a pronounced salinity gradient and postglacial history make the Baltic Sea a suitable area for studying how selection and gene flow affect genetic differentiation in marine species. A cDNA library was used to identify new single nucleotide polymorphisms (SNPs) in Baltic populations of *Mytilus* spp. mussels. Sixty polymorphic SNPs were used to genotype 642 individual mussels from the inner Baltic, Danish Straits, northwest Denmark and a population of the northeast Pacific. We characterized 49 novel SNP markers that differentiate the populations of the North and Baltic Sea areas. Concordant narrow clines were observed at the entrance of the Baltic Sea for most of these markers. Considerable variance of hybrid index scores was observed in populations with intermediate allele frequencies within the hybrid zone, e.g. in Gedser and Hjelm. The presented results are in accordance with the existence of strong reproductive isolation, probably caused by a combination of exogenous (e.g. adaptation to brackish waters) and endogenous pre- and post-zygotic factors (e.g. selection against hybrids). The overwhelming majority of new SNPs markers showed a larger representation of *M. trossulus* than *M. edulis* genes in the nuclear DNA of Baltic *Mytilus* species. Finally, we identified a few markers with an elevated level of introgression of *M. edulis* alleles in the Baltic Sea *M. trossulus* populations in comparison to the reference *M. trossulus* population of the Pacific.

KEY WORDS: Population structure · Hybrid index score · Single locus clines · Diagnostic markers

INTRODUCTION

Mytilus taxa are an important component of food chains in marine ecosystems and a dominant benthic filter-feeder in the Baltic. The Baltic Sea was colonized by *Mytilus* spp. mussels about 7000 yr ago, after the most recent freshwater period and the emergence of this area as a marine (brackish) ecosystem (Zillén et al. 2008). The Baltic Sea has a salinity ranging from almost freshwater in the northern part (Gulf of Bothnia) and brackish (6 to 8 PSU) in the inner Baltic to full salinity (from 20 to 30 PSU) in the northern Danish Straits (Kattegat) rising to 33 PSU in the North Sea. Sharp salinity gradients are built up in the Sound and the Belts.

Baltic populations are usually smaller, less genetically variable and more isolated than those in the Atlantic and may also be subject to genetic bottlenecks (Johannesson & André 2006). However, it has been observed that the inner-Baltic populations of *Mytilus* spp. mussels are more genetically variable than populations outside the Baltic (North Sea and Kattegat). A similar situation has been observed for *Macoma balthica* (Nikula et al. 2008). This is because the Baltic was first inhabited by genetically distinct lineages of these species (*Mytilus trossulus*, *Macoma balthica balthica*) which hybridized with Atlantic *Mytilus edulis* and *Macoma balthica rubra,* respectively, at the entrance to the Baltic Sea (Väinölä & Hvilsom 1991, Luttikhuizen et al. 2012). Baltic popu-

*Corresponding author: mzbawicka@iopan.gda.pl

Publisher: Inter-Research · www.int-res.com

lations of *Mytilus* spp. have a unique genetic composition of loci derived from *M. trossulus* and *M. edulis* genomes due to multiple processes: the history of the taxa involving divergence over more than 3 million yr, a hybrid zone at the entrance of the Baltic, intrinsic post-zygotic and pre-zygotic isolation (Bierne et al. 2011) and local adaptation caused by location-specific salinity regimes (i.e. locus-specific selection evidenced by impeded gene flow).

Differential introgression of loci observed across the contact zone may be caused not only by natural selection (McDonald 1994) but also by a large amount of stochastic variation at neutral loci (Bierne et al. 2003). Hybrid zones are very interesting areas for research because they make it possible to study gene flow and reproductive isolation, and provide an opportunity to examine the functional tests of different combinations of mutations. As mentioned above, some genomic regions in species from hybrid zones are not only influenced by neutral processes. Payseur (2010) tried to identify genomic regions involved in speciation using differential introgression in hybrid zones. F_{ST} outlier analysis can be performed to find the regions (loci) influenced by selection or linked to selected loci (Luttikhuizen et al. 2012). Selection against hybrids reducing their fitness has been observed in hybrid zones (Barton & Hewitt 1989, Burke & Arnold 2001).

According to Beaumont (2005), analysis of estimated F_{ST} can be a first step to identifying genes that might be under selection. However, Bierne et al. (2011, 2013) indicate some caveats in interpreting the increasing the number of F_{ST} outliers, independent of population structure, and pointed out that genome-wide genetic barriers are often multifactorial. For example, many outlying F_{ST} loci observed in gene scan may be attributable to local adaptation as well as to endogenous genetic incompatibilities (caused by, for example, cryptic hybrid zones involving multiple loci involved in pre- and post-zygotic isolation). This approach to the problem indicates that F_{ST} outliers should be considered as being candidate loci requiring further detailed investigation (Gosset & Bierne 2013).

Initially, North Sea and Baltic Sea mussels of the genus *Mytilus* were characterized using allozymes (McDonald et al. 1991, Väinölä & Hvilsom 1991). Baltic mussels were classified as being very similar and closely related to Pacific *M. trossulus*. Study of the genetic transition through the Sound and the other Belt Sea straits between the Baltic and North Sea populations using the *Gpi* allozyme locus was recently summarized by Väinölä & Strelkov (2011). The work of Väinölä & Strelkov (2011) corroborates the origin of the Baltic *Mytilus* spp. from the North Pacific. Earlier work on Baltic populations based on gene-targeted PCR markers (e.g. Bierne et al. 2003, Riginos & Cunningham 2005, Kijewski et al. 2006, Väinölä & Strelkov 2011) has shown the Baltic–North Sea contact (hybrid) zone to have a thoroughly mixed genetic composition of *M. edulis* and *M. trossulus*, unimodal in character in the Baltic Sea (introgressed *M. trossulus*) with bimodality observed in Øresund and the Danish Straits. Recent work (Zbawicka et al. 2012) based on single nucleotide polymorphisms (SNPs) detected 5 markers located in genes of the histone family and *p53* differentiating the European populations of mussels.

SNPs are variations that occur at the nucleotide level when a single nucleotide differs among or within individuals of a species. They explain 90% of the genetic differences between individuals and are thus very suitable for genetic research and selective breeding applications (Brookes 2007). SNPs located in coding regions (non-synonymous and synonymous substitutions) can be used to distinguish loci under selective pressure from neutral loci (Morin et al. 2004). Synonymous changes may not be neutral because of the efficiency of same codon translation: many organisms preferentially use certain synonymous codons (Carlini & Stephan 2003). Depending on the rate of recombination and the nature of selection, variation at synonymous SNPs can also be influenced by selection on neighboring regions (Nielsen et al. 2005). Multilocus scans used to compare different populations for several loci can identify regions in the genome carrying a mutation arising from local adaptation (Schlötterer 2002). SNPs have been widely used for differentiation studies at the individual, population and species level (Pariset et al. 2009, Quintela et al. 2010, Williams et al. 2010), and also for ecological and conservation studies (Vignal et al. 2002).

New SNPs, discovered based on the expressed sequence tag (EST) sequences of the Baltic mussel, were used as markers for *Mytilus* taxa in an analysis of populations from the Baltic Sea region. F_{ST} outlier analysis was used to identify candidate loci for which gene flow is impeded across the Danish Straits. The purpose of this study was to apply SNPs to identify markers and genes affected by hybridization in the Danish Straits and to examine the variability between them in the geographic cline (along the hybrid zone). The study also investigated the concordance of multigene clines at the Baltic Sea entrance and the extent of the barrier to gene flow.

MATERIALS AND METHODS

Identification and genotyping of SNPs

EST sequences (deposited in GenBank under accession numbers: KJ871031–KJ871077) were used for SNP discovery using the Staden computer programs (Staden et al. 2001). These sequences were aligned with appropriate GenBank EST sequences (mainly from *Mytilus galloprovincialis* and *M. edulis*) using the ClustalX program v.1.83 (Thompson et al. 1997) with default settings. The possible effect of the SNPs, resulting in change or no change in amino acid sequence (non-synonymous or synonymous changes, respectively), was predicted on the basis of the open reading frames (ORFs) identified (see Table S1 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf).

Genotyping of SNPs was performed using the Sequenom MassARRAY iPLEX platform (Gabriel et al. 2009). PCR and extension primers were designed using the Assay Design 3.1 program (Sequenom); analysis and scoring were performed using Typer 3.4 software (Sequenom). SNPs were classified, based on manual inspection, as 'failed assays' (meaning that the majority of genotypes could not be scored and/or the samples did not cluster well according to genotype), 'monomorphic SNPs', or 'polymorphic SNPs'. Assays were designed for 84 candidate SNPs (77 based on EST sequences and 7 already identified by Zbawicka et al. 2012); 5 of which (SNP1B, SNP2A, SNP2B, SNP3D and SNP6B) were characteristic of the *M. trossulus* genome and 1 of *M. galloprovincialis* (SNP1C). Seven SNPs, already identified, were renamed BM201B, BM201C, BM202A, BM202B, BM203C, BM203D and BM206B.

DNA sampling

Mytilus spp. samples, consisting of 630 individuals in total, were collected from 27 localities (Fig. 1, Table S2 in the Supplement). Adult mussels of mixed age and size (15 to 50 mm) were collected. Two samples, Veno Bight (VEM) and Logstor Bredning (LOG), were obtained from the Limfjord (northwest Denmark), which is highly influenced by North Sea water. Sixteen samples were obtained from the Kattegat and northern part of the Belt Sea: Aarhus Bight (ARH), Augstenborg Fjord (AUG), Gilleleje (GIL),

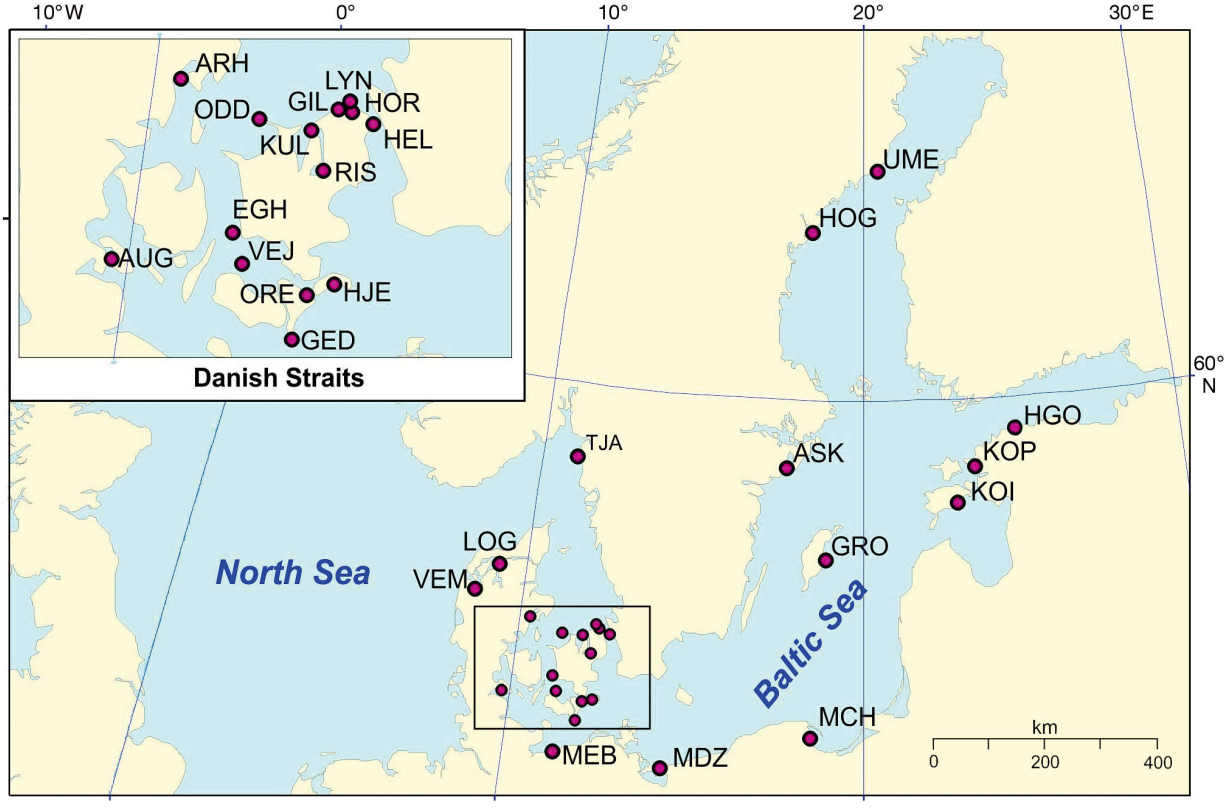

Fig. 1. Geographic locations of 27 *Mytilus* spp. sampling sites in the Baltic and North Seas. See Table 1 for definitions of site abbreviations

Helsingor (HEL), Hornbaek (HOR), Kulhuse (KUL), Lynetten (LYN), Odden (ODD), Riso (RIS), and Tjarno (TJA), including 6 from the southern part of Belt Sea: Egholm flak (EGH), Mecklenburg Bight (MEB), Ore Falster (ORE), Vejro (VEJ), Gedser (GED), and Hjelm Bight (HJE). Nine samples were obtained from the southern, central and northern Baltic Sea: Międzyzdroje (MDZ), Mechelinki (MCH), Grogarn (GRO), Asko (ASK), Hgona (HGO), Koiguste (KOI), Kopli Laht (KOP), Hoga Kusten (HOG) and Umeå (UME). Additionally, a reference population of *M. trossulus* was sampled on Vancouver Island (CAN). Most of the samples were collected in 2010; 6 were collected between 2003 and 2006 (Table S2). Samples were stored frozen at −20°C until the SNP analyses. Twenty-seven samples consisted of 16 to 29 individuals each; the sample from Canada consisted of 12 individuals. DNA was isolated from mantle tissue using a modified CTAB method according to Hoarau et al. (2002) and suspended in sterile, filtered, distilled water.

Analysis of genotype data

SNP markers and populations were analyzed for genetic diversity, proportion of polymorphic SNP (P_o), observed (H_o) and expected (H_e) heterozygosity, and inbreeding coefficient (F_{IS}) using Arlequin v.3.5 (Excoffier & Lischer 2010). The statistical significance of F_{IS} (>0) was tested by 10 000 permutations of alleles between individuals. Departures from Hardy-Weinberg equilibrium (HWE) were tested by exact test, and significance was determined by Markov chain Monte Carlo simulations. Pairwise analysis in Arlequin was used to calculate mean pairwise F_{ST} values defining population differentiation. F_{ST} values at individual SNPs were calculated using the AMOVA function of the same program. Permutation testing with 1000 iterations was used to calculate p-values for mean and locus-by-locus F_{ST} values. Arlequin was also used to detect loci under selection through genetic structure analysis. The neutral distribution of F_{ST} was simulated with 30 000 iterations at a 95% confidence level, and characterized by estimating the 0.05 and 0.95 quantiles of the distribution. In the first step, the neutral expectation was based on the overall mean value of F_{ST} calculated for all SNPs. First, the analysis was conducted for all SNPs, then SNPs with F_{ST} values outside the 0.95 limits corresponding to the null hypothesis were removed and a new analysis was performed with the recalculated value of F_{ST}. SNPs with F_{ST} values above the 0.95 limits after the second analysis were also considered to be outlier loci (Acheré et al. 2005). This procedure reduces bias in the estimation of F_{ST} by removing the highly diverged loci. The F_{ST} distance matrix of sampling sites obtained in Arlequin was used to construct a neighbour-joining tree illustrating the genetic relatedness of all populations, using MEGA v.4 (Tamura et al. 2007). Correction for multiple tests of overall significance values was carried out using the Bonferroni procedure of Rice (1989).

The frequency distribution of the score for a hybrid index, giving the percentage of *M. trossulus* characteristic alleles, was calculated for all loci with F_{ST} > 0.5 (Table S1). A score of 0 indicates a pure *M. edulis*, whereas a score of 1 indicates a pure *M. trossulus*. Only one, the most differentiated SNP coming from the same fragment, was used.

GENEPOP 4.1 (Rousset 2008) was used to test for linkage disequilibrium (LD) between all pairwise combinations of 51 loci (1 SNP per contig or fragment, marked in Table S1) for all populations separately, and for 2 groups of populations: (1) populations from the North Sea and the northern Danish Straits, and (2) populations from the inner Baltic Sea grouped together. Correspondence analysis (CA; Benzécri 1992), implemented in GENETIX (Belkhir et al. 2003), was used for visualizing the genetic substructure at population and individual levels. The result is presented as a scatter plot, with the axes representing the contribution of inertia of the data matrix in a way that can be considered analogous to the total variance in allelic frequency (Benzécri 1992).

RESULTS

SNP validation

In total, 340 putative SNPs were found, distributed across 146 contigs. Eighty-four candidate SNPs were chosen to genotype 642 individual mussels from 28 populations. The assay design failed for 16 SNPS, 6 did not provide an acceptable quality score and 2 were monomorphic in all samples. The remaining 60 SNPs were polymorphic with a high quality score (above 90%) in most (98%) of the sampled individuals. SNP annotation is presented in Table S1 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf. Seven SNPs were identified in contigs that included the *histone* and *hsp70* genes (Table S1), as reported in the study of European populations of the *Mytilus* spp. mussels (Zbawicka et al.

2012). ORFs were identified in the majority of fragments. ORF identification was not possible in only 6 contigs (6 SNPs). Of the 60 SNPs, 47 (78.3%) were located in coding regions, and the majority (40; 85.1%) were synonymous. Of the SNPs used, 7 (11.6%) were located in noncoding regions (Table S1).

Of the 60 SNPs, only 3 (BM20A, BM21C and BM9C) had more than 2 alleles each (Table S1). The mean (±SD) minor allele frequency (MAF) for SNPs with 2 alleles was 0.14 ± 0.09. SNPs with MAF < 0.2 were highly represented in the study data. The highest average MAF values were observed for samples from the southern Danish Straits (Table 1).

Very little LD between pairs of 51 markers was found. Only 7 pairs of loci (BM202A vs. BM203D, BM11A and BM201B; BM203D vs. BM11A and BM201B; BM11A vs. BM201B; BM12A vs. BM67B) out of a total of 1275 were in highly significant LD ($p < 0.0001$) in 3 populations from the southern part of the Belt Sea: 1 pair in EGH, 7 in GED and 4 in HJE. One pair of SNPs (BM12A vs. BM67B) was in highly significant LD in the specified 3 samples and in one group of populations of predominantly the *Mytilus trossulus* type. For 3 pairs of loci with BM201B, highly significant LD was found in 1 population (GED).

The frequency distribution of the hybrid index (HI) in the study populations is presented in Fig. 2. The percentage of *M. edulis* characteristic alleles (at loci with $F_{ST} > 0.5$), showed clinal variation for the Kattegat-Danish Straits-Baltic Sea region. The greatest variance of HI was observed in populations GED and HJE with the highest number of loci with departures from HWE (Table 1).

Table 1. Genetic parameters of the 28 populations of *Mytilus* mussels. F_{IS}: inbreeding coefficient; *H:* Hardy-Weinberg equilibrium; H_o: observed heterozygosity; H_e: expected heterozygosity; MAF: minor allele frequency; V_{HI}: variance of hybrid index. VEM: Veno Bight; LOG: Logstor Bredning; ARH: Aarhus Bight; AUG: Augstenborg Fjord; GIL: Gilleleje; HEL: Helsingor; HOR: Hornbaek; KUL: Kulhuse; LYN: Lynetten; ODD: Odden; RIS: Riso; TJA: Tjarno; EGH: Egholm flak; MEB: Mecklenburg Bight; ORE: Ore Falster; VEJ: Vejro; GED: Gedser; HJE: Hjelm Bight; MDZ: Międzyzdroje; MCH: Mechelinki; GRO: Grogarn; ASK: Asko; HGO: Hgona; KOI: Koiguste; KOP: Kopli Laht; HOG: Hoga Kusten; UME: Umeå; CAN: Canada. Values with $p < 0.05$ after Bonferroni correction are marked in **bold**; $^{\star}p < 0.05$

Population	F_{IS}	No. of loci with HWE departure ($p < 0.01$)	Loci with HWE departure after Bonferroni correction	H_o	H_e	MAF	Average gene diversity over loci	Average no. of pairwise differences within population	No. of individuals	V_{HI}
VEM	0.2191*	2		0.0591	0.0834	0.0554	0.0517	4.28	28	0.0080
LOG	**0.2558***	4		0.0767	0.1037	0.0623	0.0818	5.27	24	0.0230
ARH	0.1223*	3		0.1229	0.1444	0.0916	0.1353	7.74	29	0.0384
AUG	0.1226*	1		0.0958	0.1146	0.0735	0.0938	5.98	28	0.0275
GIL	0.0574	1		0.1433	0.1583	0.1025	0.1431	8.45	27	0.0376
HEL	**0.1786***	4		0.1257	0.1601	0.1037	0.1534	8.81	25	0.0671
HOR	**0.1598***	2		0.1055	0.1303	0.0802	0.1181	6.80	27	0.0279
KUL	**0.2383***	1	BM203C	0.1103	0.1454	0.0916	0.1329	7.52	23	0.0298
LYN	0.1403*	1		0.1065	0.1327	0.0793	0.1209	6.74	20	0.0373
ODD	0.0676	0		0.1408	0.1556	0.0956	0.1437	8.05	17	0.0422
RIS	0.2152*	2	BM203C	0.0877	0.1113	0.0721	0.0973	5.84	18	0.0290
TJA	**0.2181***	3		0.1035	0.1364	0.0847	0.1278	7.27	26	0.0297
EGH	0.0968*	2		0.1892	0.2128	0.1365	0.1953	11.64	26	0.0595
MEB	0.1262*	1		0.1924	0.2327	0.1486	0.2155	11.09	18	0.0807
ORE	**0.2379***	5		0.1926	0.2637	0.1719	0.2530	14.13	21	0.1598
VEJ	0.1248*	2		0.1844	0.2195	0.1369	0.2124	11.86	27	0.0834
GED	**0.1637***	7	BM17B	0.2866	0.3584	0.2787	0.3494	19.03	28	0.2001
HJE	**0.1541***	7	BM33B, BM98C	0.2809	0.3444	0.2506	0.3429	18.92	26	0.1741
MDZ	0.0786	2		0.2924	0.3367	0.2429	0.3177	16.94	18	0.0952
MCH	0.1485*	3		0.2219	0.2863	0.1923	0.2554	12.20	16	0.0737
GRO	0.1207*	2	BM17B	0.2474	0.2937	0.2136	0.2726	15.52	23	0.0704
ASK	0.1400*	5	BM17B, BM98C	0.2160	0.2565	0.1804	0.2410	13.43	29	0.0449
HGO	0.1348*	4	BM55A	0.2401	0.2873	0.2104	0.2694	15.07	23	0.0567
KOI	**0.1578***	5	BM17B, BM33B	0.2252	0.2767	0.1963	0.2494	13.23	23	0.0532
KOP	0.1509*	2		0.2258	0.2854	0.2058	0.2733	14.62	16	0.0513
HOG	0.1252*	6	BM17B, BM33B	0.2083	0.2486	0.1702	0.2286	12.96	22	0.0584
UME	0.0574	1		0.2337	0.2647	0.1889	0.2405	13.68	22	0.0400
CAN	0.1493*	0		0.1125	0.1395	0.0959	0.1085	6.21	12	0.0238
Average	0.1486			0.1724	0.2101	0.1433	0.1937	10.83		

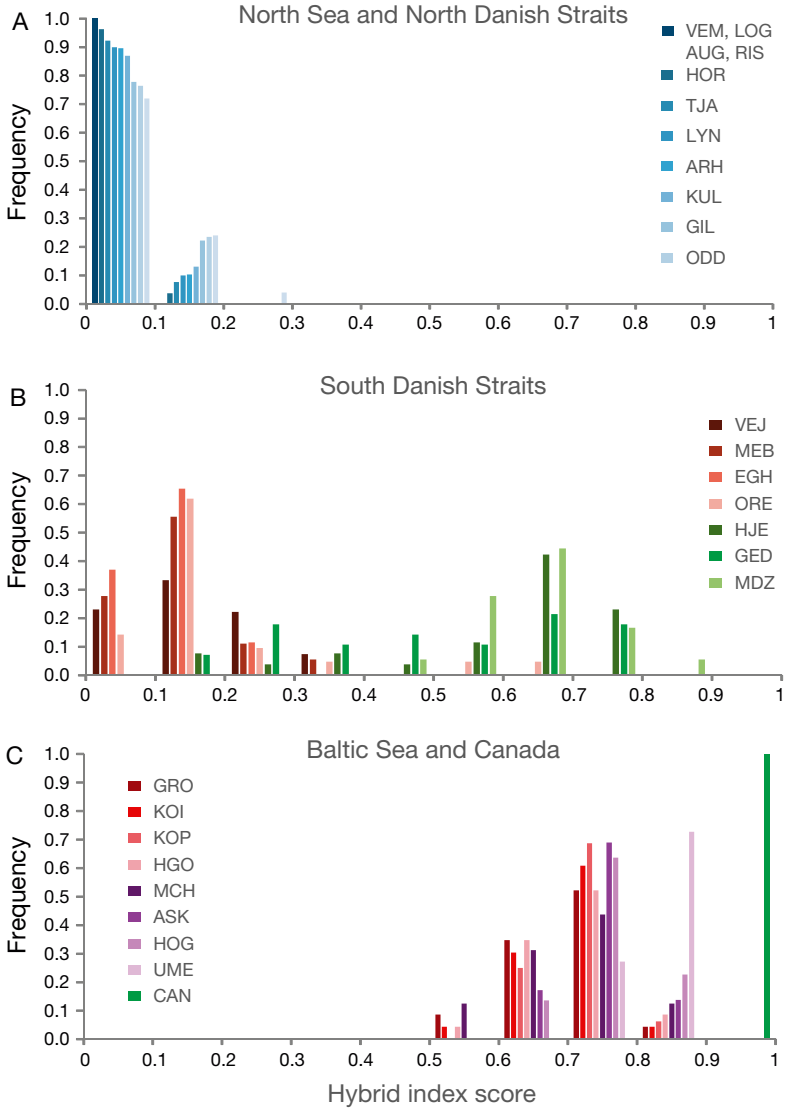

Fig. 2. Frequency distribution of the score for a hybrid index giving the percentage of *Mytilus trossulus* characteristic alleles. A score of 0 is a pure *M. edulis*, whereas a score of 1 is a pure *M. trossulus*. Analysis is presented for 3 groups of populations: (A) North Sea and northern Danish Straits, (B) southern Danish Straits, and (C) Baltic Sea and Canada. See Table 1 for definitions of site abbreviations

Genetic diversity and Hardy-Weinberg equilibrium

The percentage of polymorphic SNPs (P_o) ranged from 35.0 to 98.3% among populations, and was highest in populations from the southern Danish Straits. The most homogenous population was the *M. edulis* population from the North Sea, whereas almost all loci in 2 populations from the southern Danish Straits (GED and HJE) were polymorphic.

Most loci were in HWE in the different populations. The biggest fraction of SNPs that were not in HWE ($p < 0.01$) was observed in samples from the southern Danish Straits (GED and HJE) (Table 1). Only 5 loci showed significant departures from HWE after a Bonferroni correction. BM17B (ribosomal protein) was not in HWE in 5 populations from the southern Danish Straits and Baltic Sea, while BM33B (cytochrome *c* oxidase) was not in HWE in 3 populations from the same region because some subunits are encoded in the mitochondrial genome. H_o for 60 loci among all populations was in general lower than H_e. Heterozygosity among populations was generally lowest in the North Sea and the northern Danish Straits, increased toward the south, and decreased slightly in the inner Baltic (Table 1). The mean within-population fixation index F_{IS} (averaged over all polymorphic loci in each population) was 0.15, showing a significant excess of homozygotes in 9 populations and in all study regions (Table 1).

Genetic differentiation among populations

We investigated the potential of SNP markers for discriminating between populations of *Mytilus* spp., concentrating on mussels from the Baltic Sea region. The level of genetic differentiation among the studied populations was high ($F_{ST} = 0.257$). F_{ST} values were significantly greater than zero between most of the pairs of samples (reaching values as high as 0.874) (Table S3 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf). Values not significantly different from zero (indicating the absence of differentiation) were observed in comparisons within groups of 10 northern Danish Straits samples as well as the 8 inner Baltic Sea samples. The mean pairwise F_{ST} values were 0.06, 0.191 and 0.344 for the North Sea vs. northern Danish Straits, northern vs. southern Danish Straits and southern Danish Straits vs. inner Baltic, respectively, all of which were highly significant ($p < 0.01$). AMOVA also revealed significant differences between these 3 groups. The highest differentiation was observed between the North Sea samples and the Baltic Sea with the Canada group ($F_{ST} = 0.704$). However, comparing the Baltic Sea and Canada samples, F_{ST} values were also high (0.258) and significant. In order to assess the introgression of Baltic populations, F_{ST} values between *M. edulis* and Baltic *M. trossulus* were

contrasted with this value between *M. edulis* and Pacific *M. trossulus*, and the results obtained were 0.686 and 0.863, respectively. The highest gene diversity and pairwise differences were observed in the more eastern populations from the southern Danish Straits (GED and HJE) (Table 1).

The neighbour-joining tree showing the genetic relationships of all samples was constructed using the F_{ST} distance matrix (Table S3 in the Supplement), and is presented in Fig. 3. The 2 main groups represent *M. edulis* and *M. trossulus*, each with *M. edulis–M. trossulus* hybrids. Six groups of populations can be distinguished based on this tree: (G1), the North Sea *M. edulis* populations; (G2), the northern Danish Straits *M. edulis* populations, closer to the hybrid zone (a little more introgressed); (G3 and G4), southern Danish Straits populations that are an admixture of various proportions of *M. edulis* and *M. trossulus*: (G3) with a predominance of *M. edulis*, and (G4) with a predominance of *M. trossulus*; (G5), *M. trossulus* from the Baltic; and (G6), the *M. trossulus* from Canada. AMOVA revealed significant differences between these 6 groups of populations.

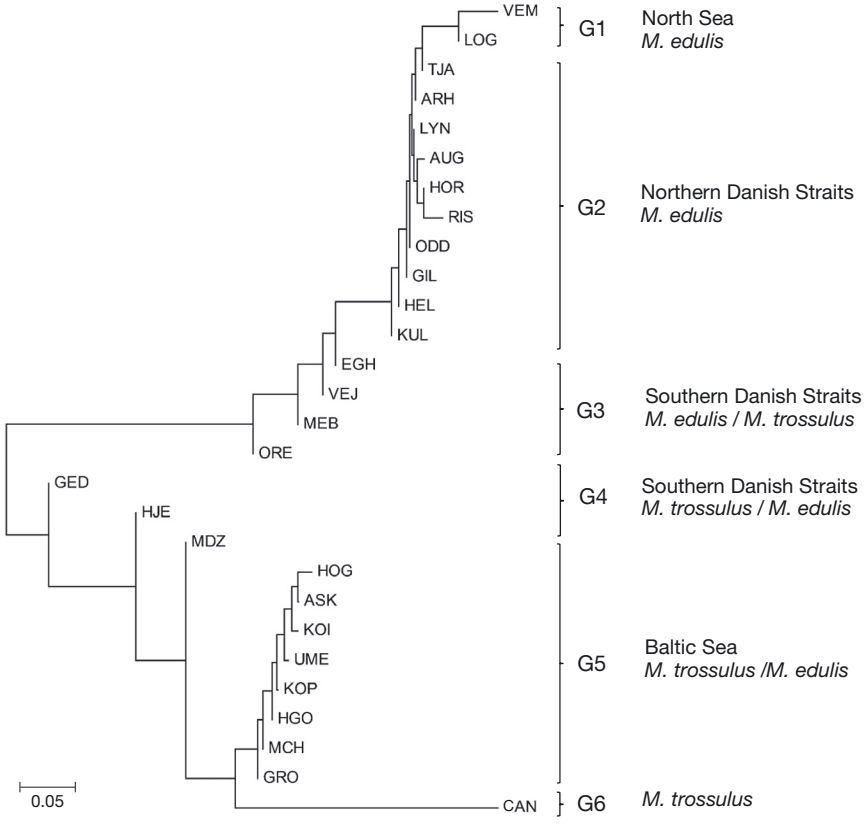

Fig. 3. Neighbour-joining tree of *Mytilus* spp. populations based on the F_{ST} distance matrix

Detection of outlier loci

Twenty SNPs (33.3%) with F_{ST} values significantly different from zero and departing from the neutral-model F_{ST} expectation were identified by outlier tests carried out for all studied populations (Table S1 in the Supplement). Initially, 2 main groups were tested to detect outlier loci: populations with a predominance of *M. edulis* genes (G1–G2) and populations with a predominance of *M. trossulus* genes (G5–G6). Populations with an admixture of both taxa with hybrids (G3–G4) were not analysed. The first test indicated 23 outlier loci (Fig. 4). These 23 loci, representing 20 genes, were significantly involved in the differentiation between the 2 main groups of the studied populations, and were candidates for mussel taxonomic markers (17 synonymous, 2 non-synonymous, 4 non-coding). Eight of them corresponded to ribosomal genes, 4 were in histone genes, 1 was in the *hsp70* gene and another in an ETC subunit and COX subunit, and the location of the other 8 was unknown. Two SNPs (BM79B and BM206B) demonstrated non-synonymous substitution. The number of outlier loci dropped to 15 when 2 populations at the edge of the hybrid zone (EGH and MDZ) were subsequently included.

In the following stage, outlier tests were carried out for 2 smaller groups of populations to identify SNPs significantly involved in their differentiation. Fifty-eight SNPs were polymorphic for North Sea and northern Danish Straits populations in groups G1 and G2. Compared to the previous test for outlier loci, only one SNP (BM15C) was identified by this outlier test as being significantly differentiated between the 2 groups. Five of 59 polymorphic SNPs were involved in significant differentiation between groups G5 and G6 of the inner Baltic and Canada populations. Three SNPs (BM2G, BM206B and BM86A) were significant also in the analysis between G1–G2 and G5–G6 groups, while BM17B and BM98C were new. In the analyses described above, 26 SNPs were significantly involved in the differentiation between study groups of populations. The single locus clines for the most differentiated 20 loci were further described by a plot of allele frequency with distance from North Sea populations (Fig. 5).

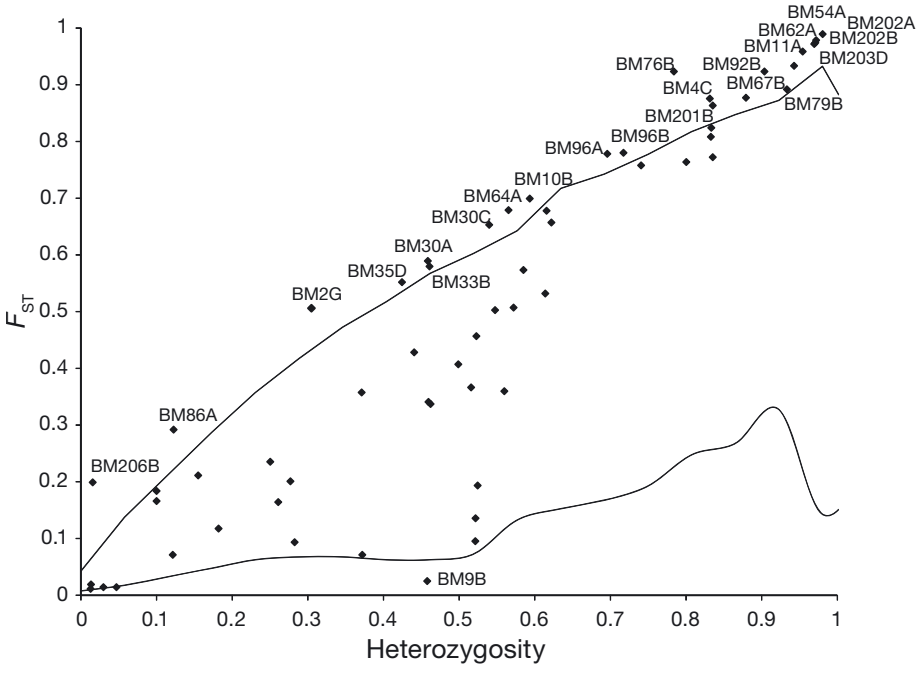

Fig. 4. Detection of outlier loci. Distribution of F_{ST} values as a function of the heterozygosity under a neutral model for the 60 single nucleotide polymorphisms (SNPs) analysed. Analysis was done for 2 groups: G1–G2 (populations with a predominance of *Mytilus edulis* genes) and G5–G6 (populations with a predominance of *M. trossulus* genes). SNPs that had an F_{ST} value above the 0.95 or below the 0.05 quantile (upper and lower solid lines) were considered to be outlier loci. Each SNP is indicated by a diamond

Population structure

A CA test was carried out to characterize population structure. CA results for individuals are shown in Fig. 6; for higher resolution, only samples from the Baltic Sea environments are presented. The first 2 axes accounted for 91% of the total variation. Axis 1 shows a separation between the North Sea and Danish Straits populations with a predominance of *M. edulis* (groups G1, G2 and G3) and populations possessing the *M. trossulus* genes (G4 and G5). The North Sea and northern Danish Straits individuals formed a very tight group, partially overlapping with some southern Danish Straits individuals, while the inner Baltic group had a greater dispersion. Individuals from the 3 most south-eastern Danish Straits populations GED, HJE and ORE either grouped with *M. edulis*, *M. trossulus* or were situated between them. Results for all populations are shown on Fig. S1 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf.

DISCUSSION

SNP features

We described and characterized 60 SNPs as potential markers for *Mytilus* spp. mussels from the North and Baltic Sea regions. These SNPs were used to genotype 642 mussels collected from 28 localities. Genotyped data very clearly differentiated populations and allowed an accurate description to be made of the *M. edulis* × *M. trossulus* hybrid zone in the Danish Straits.

Based on the analysis of linkage disequilibrium, it can be concluded that almost all studied SNPs were independent characters. Seven pairs of loci, all with highly significant LD, were characteristic of the *M. trossulus* genome and could thus account for the linkage signature. The highest level of LD (as was to be expected even without physical linkage) was observed in populations from the centre of hybridization (southern part of the Belt Sea), in particular in the GED and HJE samples. The greatest variance of hybrid index score, which is a straightforward multilocus measure of linkage disequilibrium in hybrid zones (Kruuk et al. 1999), was observed in the same populations. Analysis of inter-marker disequilibria in mixed populations often reveals LD between any 2 alleles at different loci with different allele frequencies, even when the loci are unlinked (Gorroochurn et al. 2007). Such a situation was observed in the study of SNP markers in Scottish and Norwegian *M. edulis*–*M. trossulus* populations (Zbawicka et. al 2012).

Baltic *Mytilus* hybrid zone characteristics and SNPs as genetic markers

The variability between markers and genes that could be affected by hybridization in the geographic

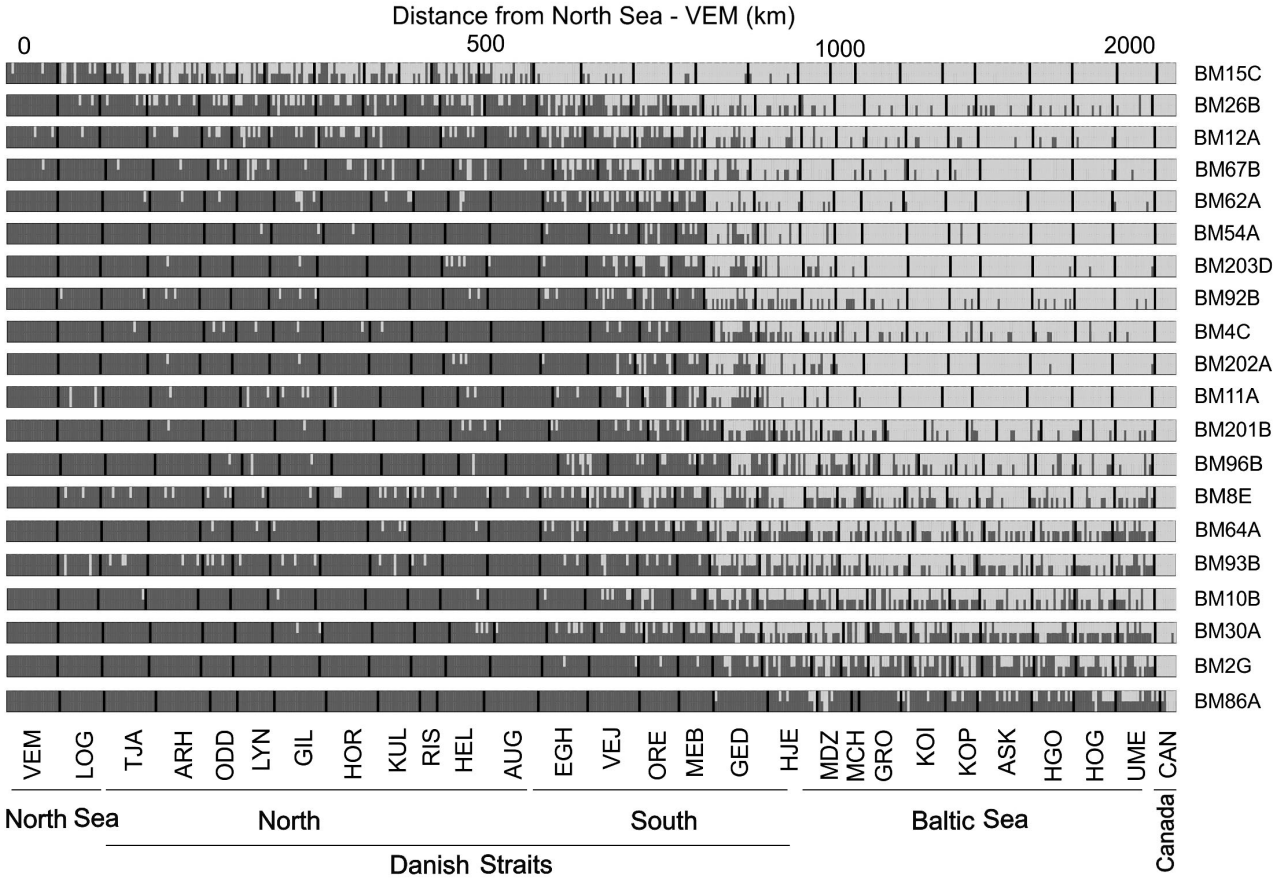

Fig. 5. The single locus clines in the Kattegat through the Sound for the most differentiated 20 loci described by allele frequency as a function of distance from North Sea populations

cline were examined. When the results from an analysis of allele frequencies (Table S4 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf) and outlier analysis were compared, the 49 SNPs found in 4 DNA fragments and 40 EST contigs were candidates for markers differentiating Baltic Sea region populations and affected by hybridization in the Danish Straits. Grouping of the populations had an impact on the identification of outlier markers. Groups were defined on the basis of the F_{ST} distance matrix and CA analysis results. Twenty six SNPs were significantly involved in the differentiation between mussels from the North Sea and Baltic area (Fig. 4), showing very sharp differences in frequencies (Fig. 5) in a similar way to allozymes (Väinölä & Hvilsom 1991, Väinölä & Strelkov 2011) and 2 of the 4 nuclear DNA markers studied by Stuckas et al. (2009): lysin M7 and male mtDNA. An evenly distributed upward trend of the number of individuals with *M. trossulus* genes can be observed from the North Sea to the inner Baltic and Canada for another 23 SNPs (Table S4). Although specific alleles at 3 SNPs (BM22A, BM60A and BM3B) were more frequent in the inner Baltic than in the other study region, the differences were not significant. Adding the introgressed populations at the border of the hybrid zone to the groups results in overinflating the heterogeneity in differentiation level and consequent reduction of outliers.

Introgressive hybridization is very strong for maternally inherited mtDNA and a little weaker for the nuclear ITS and SNP5B markers, resulting in complete or almost complete replacement of *M. trossulus* with *M. edulis* mtDNA or fragments of nuclear DNA (Quesada et al. 1995, Kijewski et al. 2006, Zbawicka et al. 2012). In the present work, a similar phenomenon was observed, with a high frequency of the alleles characteristic of pure *M. edulis* occurring in 5% of the SNPs (e.g. BM2G, BM86A and BM98C) (Table S4). In contrast, about 23% of the SNPs that were characteristic of a putatively pure Canadian *M. trossulus* population were observed in Baltic populations at a very high frequency (90 to 100%) (e.g. BM11A, BM12A and BM203D) (Table S4).

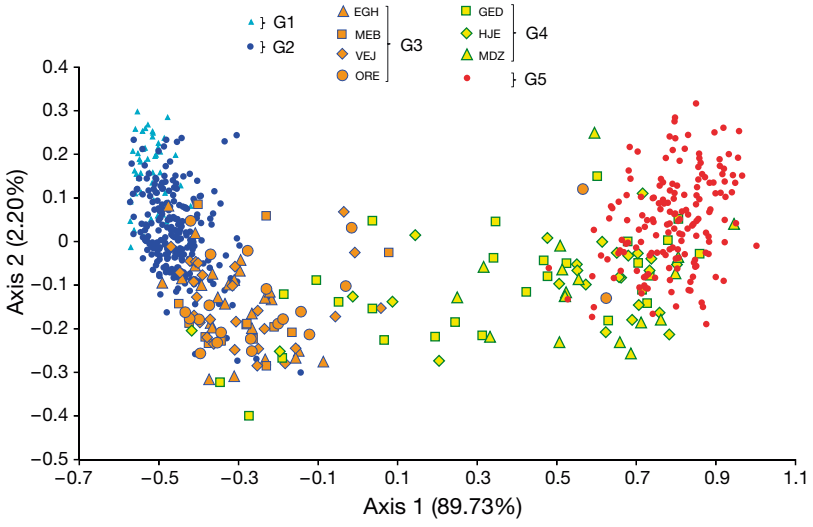

Fig. 6. The first 2 axes of the correspondence analysis (CA) computed from the single nucleotide polymorphism (SNP) data on *Mytilus* spp. populations from Baltic Sea region. Each dot (point) is an individual. Seven populations of *M. edulis*–*M. trossulus* from the southern Danish Straits are indicated with a different symbol (G3 and G4), whereas populations from the North Sea (G1), north Danish Straits (G2) and inner Baltic Sea (G5) are shown with the same symbol for each group

A similar effect had already been observed for the EF-bis marker and for allozyme data, and was described as a cline (e.g. *Gpi* locus) (Väinölä & Hvilsom 1991, Kijewski et al. 2006, Stuckas et al. 2009, Väinölä & Strelkov 2011). This may have resulted from selection and differential introgression (Bierne et al. 2003). None of the inner Baltic individuals had only *M. trossulus*- or *M. edulis*-specific SNP alleles.

Another characteristic of this region was differing levels of introgression observed at some loci over relatively short distances (southern Danish Straits, e.g. VEJ and HJE, ORE and GED). Stuckas et al. (2009) argued for discordance of the shape of multilocus clines across the Baltic contact zone between *M. edulis* and *M. trossulus*, based on 3 nuclear and 2 mitochondrial markers with very small sample sizes, while Väinölä & Hvilsom (1991) and Väinölä & Strelkov (2011) both reported concordant allozyme clines with more extensive sampling. The analysis of SNP frequency in this study showed clinal variation for the region comprising the Kattegat, Danish Straits and Baltic Sea. Alleles predominant or common in the North Sea area were nearly absent from the other areas. In this work, most of the studied SNPs showed a concordant and abrupt genetic shift around the Øresund and Danish Belts, which confirms the data presented for some selected SNPs (Fig. 5).

The presence of genome-wide genetic barriers to gene flow and the appearance of the F_{ST} outliers is partly the result of the history of these mussels and the hybrid zone that, with long periods of allopatric isolation, would have permitted the accumulation of many intrinsic incompatibilities throughout the genome (Bierne et al. 2011). Local adaptation caused by different salinity that leads to non-neutrality, favouring one gene type inside the Baltic and another outside it, should also be taken into account. Thus, the barriers to gene flow can have a multifactorial character, both exogenous and endogenous. The existence of a semi-permeable genetic barrier is confirmed by the trend of evenly distributed SNP frequency across the hybrid zone (BM6C, BM16A, BM33B, BM73B and BM88A). Baltic populations of *M. trossulus* are all more introgressed than the Pacific population, shown by comparison of F_{ST} values between *M. edulis* and both *M. trossulus* populations. However, in a locus-specific analysis, 8 SNPs characteristic of a putatively pure Canadian *M. trossulus* population were observed in inner Baltic populations with 100% frequency (Table S4).

Similar to the mussel *Mytilus* spp., genomic clines over the Sound and the Danish Belts with selection for certain variants of genes have also been observed in the bivalve *Macoma balthica* (Luttikhuizen et al. 2012) and some fish species (e.g. André et al. 2010). Based on allozyme loci in *M. balthica*, it had also been observed that the transitions in different characters do not coincide exactly, and the extent of introgression varies among loci (Nikula et al. 2008). Barriers to Baltic-Atlantic gene flow were also observed in other species including pike, whitefish, stickleback and bladderwrack (Wennerström et al. 2013).

Characteristics of the studied *Mytilus* spp. populations and interpopulation differentiation

Two populations from the southern Danish Straits (from the eastern islands of Falster and Moen: GED and HJE) had the largest number of polymorphic loci (almost 100%), the highest gene diversity, and considerable departures from HWE (Table 1). These re-

sults indicate the location of the centre of the hybrid zone, which is characterized by strongly impeded gene flow. In some populations (KUL, ORE and HJE), approximately 17% of loci had very large (>0.6) within-population fixation index values (F_{IS} averaged across all loci) (Table S5 in the Supplement at www.int-res.com/articles/suppl/b021p025_supp.pdf). The majority of these loci were nearly homozygous for 1 allele within a population, with the absence of any heterozygotes or a single heterozygote only, and a few individuals homozygous for the alternative allele. An excess of homozygotes of the different taxa was observed. In a few cases, F_{IS} had values of 1, indicating the existence of 2 alternative homozygotes without any heterozygotes (e.g. BM11A, BM17B, BM54A and BM55A). Pure *M. edulis* populations were mostly homozygous for this alternative allele, and the *M. trossulus* population from Canada was predominately homozygous for the second allele. This phenomenon was observed to the greatest extent in the Scottish and Norwegian populations of *Mytilus* spp., and was probably caused by recent mixing of *M. trossulus* with *M. edulis* (Zbawicka et al. 2012). In contrast, in the Danish Straits there is a long history of divergence, involving endogenous and exogenous barriers to gene flow as well as local adaptation (Bierne et al. 2013). Väinölä & Hvilsom (1991) observed a similar situation for some allozymes and explained it as selection against hybrids in later generations. These phenomena can be useful in determining the adaptive differentiation of populations identified from differences in allele frequencies among populations (i.e. F_{ST}).

Based on the present study, it can be concluded that the boundary separating populations with a predominance of *M. edulis* genes from those with a predominance of *M. trossulus* genes is located around the eastern islands of Falster and Moen, where the salinity is 10 to 14 PSU. For most of the examined SNPs, allele frequencies change abruptly, creating concordant narrow clines at the Baltic Sea entrance (Kattegat and the Sound). This study revealed the complexity of the Baltic hybrid zone and clearly showed how populations of the Danish Straits and Øresund are more an admixture of the 2 taxa than a unimodal hybrid swarm in the inner Baltic. In conclusion, the majority of new SNPs showed greater frequency of *M. trossulus* than *M. edulis* genes in the nuclear DNA of Baltic *Mytilus* spp. The results indicate the high potential of the new SNPs in the study of hybridization zones, and in particular as a new tool to study population structure and illustrate admixture level of mussels of the genus *Mytilus*.

Acknowledgements. This study was partially funded by the BaltGene Bonus Plus, 03/BONUS/2009 SPB, 2011/01/B/NZ9/04352 NCN project to R.W. and Statutory topic IV in the IOPAS. We thank Dr. Hans Kautsky and Dr. Jonne Kotta for generously providing samples. The mussel sampling in Danish waters benefited from cooperation with BONUS+ project BEAST and the contaminant activities in the nationwide monitoring programme for environment and nature (NOVANA) in Denmark.

LITERATURE CITED

Acheré V, Favre JM, Besnard G, Jeandroz S (2005) Genomic organization of molecular differentiation in Norway Spruce (*Picea abies*). Mol Ecol 14:3191–3201

André C, Larsson LC, Laikre L, Bekkevold D and others (2010) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral versus putatively selected loci. Heredity 2011:270–280

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. Nature 341:497–503

Beaumont MA (2005) Adaptation and speciation: what can *F*(*st*) tell us? Trends Ecol Evol 20:435–440

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2003) GENETIX version 4.04, logiciel sous Windows™ pour la genetique des populations. Laboratoire Genome, Populations, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier

Benzécri JP (1992) Correspondence analysis handbook. In: Balakrishnan N, Schucany WR, Garvey PR (eds) Statistics: a series of textbooks and monographs, Vol 125. Marcel Dekker, New York, NY

Bierne N, Daguin C, Bonhomme F, David P, Borsa P (2003) Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. Mol Ecol 12:2505–2510

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol Ecol 20:2044–2072

Bierne N, Rose D, Welch J (2013) Pervasive selection or is it…? Why are F_{ST} outliers sometimes so frequent? Mol Ecol 22:2061–2064

Brookes, AJ (2007) Single nucleotide polymorphism (SNP). Encyclopedia of life sciences, doi: 10.1002/9780470015902.a0005006.pub2

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. Annu Rev Genet 35:31–52

Carlini DB, Stephan W (2003) *In vivo* introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. Genetics 163: 239–243

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564–567

Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr Protoc Hum Genet 60:2.12.1–12.12.18

Gorroochurn P, Hodge SE, Heiman GA, Greenberg DA (2007) A unified approach for quantifying, testing and correcting population stratification in case-control association studies. Hum Hered 64:149–159

Gosset CC, Bierne N (2013) Differential introgression from

a sister species explains high F_{ST} outlier loci within a mussel species. J Evol Biol 26:14–26

Hoarau G, Rijnsdorp AD, Van Der Veer HW, Stam WT, Olsen JL (2002) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. Mol Ecol 11:1165–1176

Johannesson K, André C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol Ecol 15:2013–2029

Kijewski T, Zbawicka M, Väinölä R, Wenne R (2006) Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis.* Mar Biol 149:1371–1385

Kruuk LEB, Baird SJE, Gale K, Barton NH (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics 153: 1959–1971

Luttikhuizen PC, Drent J, Peijnenburg KTCA, Van Der Veer HW, Johannesson K (2012) Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica.* Mol Ecol 21:3048–3061

McDonald JH (1994) Detecting neutral selection by comparing geographic variation in protein and DNA polymorphisms. In: Golding B (ed) Non-neutral evolution: theories and molecular data. Chapman & Hall, New York, NY, p 88–100

McDonald JH, Seed HR, Koehn RK (1991) Allozyme and morphometric characters of three species of *Mytilus* in the northern and southern hemispheres. Mar Biol 111: 323–335

Morin PA, Luikart G, Wayne RK (2004) SNP in ecology, evolution and conservation. Trends Ecol Evol 19:208–216

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. Genome Res 15:1566–1575

Nikula R, Strelkov P, Väinölä R (2008) A broad transition zone between an inner Baltic hybrid swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). Mol Ecol 17:1505–1522

Pariset L, Joost S, Ajmone Marsan P, Valentini A (2009) Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. BMC Genet 10:7

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol Ecol Resour 10:806–820

Quesada H, Wenne R, Skibinski DOF (1995) Differential introgression of mitochondrial DNA across species boundaries within the marine mussel genus *Mytilus.* Proc R Soc Lond 262:51–56

Quintela M, Berlin S, Wang B, Hoglund J (2010) Genetic diversity and differentiation among *Lagopus lagopus* populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. Mol Ecol 19: 2380–2393

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Riginos C, Cunningham CW (2005) Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. Mol Ecol 14:381–400

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Schlötterer C (2002) Towards a molecular characterization of adaptation in local populations. Curr Opin Genet Dev 12:683–687

Staden R, Judge DP, Bonfield JK (2001) Sequence assembly and finishing methods. In: Baxevanis AD, OuelletteBFF (eds) Bioinformatics: a practical guide to the analysis of genes and proteins, 2nd edn. John Wiley & Sons, New York, NY, p 303–322

Stuckas H, Stoof K, Quesada H, Tiedemannet R (2009) Evolutionary implications of discordant clines across the Baltic *Mytilus* hybrid zone (*Mytilus edulis* and *Mytilus trossulus*). Heredity 103:146–156

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). Biol J Linn Soc 43:127–148

Väinölä R, Strelkov P (2011) *Mytilus trossulus* in Northern Europe. Mar Biol 158:817–833

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34:275–305

Wennerström L, Laikre L, Ryman N, Utter FM and others (2013) Genetic biodiversity in the Baltic Sea: species-specific patterns challenge management. Biodivers Conserv 22:3045–3065

Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF (2010) SNP identification, verification, and utility for population genetics in a non-model genus. BMC Genet 11:32

Zbawicka M, Drywa A, Śmietanka B, Wenne R (2012) Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. Mar Biol 159:1347–1362

Zillén L, Conley DJ, Andrén T, Andrén E, Björck S (2008) Past occurrences of hypoxia in the Baltic Sea and the role of climate variability, environmental change, and human impact. Earth Sci Rev 91:77–92

Editorial responsibility: Judith Grassle, New Brunswick, New Jersey, USA

Submitted: March 20, 2013; Accepted: March 31, 2014
Proofs received from author(s): May 26, 2014