Naturalis Repository

Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve Macoma balthica

P.C. Luttikhuizen, J. Drent, K.T.C.A. Peijnenburg (Katja), H.W. van der Veer,
K. Johannesson

Downloaded from:
https://doi.org/10.1111%2Fj.1365-294X.2012.05586.x

Article 25fa Dutch Copyright Act (DCA) - End User Rights

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with consent from the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available following a reasonable period after the work was first published, provided that reference is made to the source of the first publication of the work.

This publication is distributed under the Naturalis Biodiversity Center 'Taverne implementation' programme. In this programme, research output of Naturalis researchers and collection managers that complies with the legal requirements of Article 25fa of the Dutch Copyright Act is distributed online and free of barriers in the Naturalis institutional repository. Research output is distributed six months after its first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and copyrights owner(s) of this work. Any use of the publication other than authorized under this license or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the department of Collection Information know, stating your reasons. In case of a legitimate complaint, Collection Information will make the material inaccessible. Please contact us through email: collectie.informatie@naturalis.nl. We will contact you as soon as possible.

MOLECULAR ECOLOGY

Molecular Ecology (2012) 21, 3048–3061 doi: 10.1111/j.1365-294X.2012.05586.x

Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*

P. C. LUTTIKHUIZEN,*† J. DRENT,* K. T. C. A. PEIJNENBURG,‡§ H. W. VAN DER VEER* and K. JOHANNESSON†

**Department Marine Ecology, Royal Netherlands Institute for Sea Research, PO Box 59, 1790AB Den Burg, The Netherlands, †Department of Marine Ecology, Tjärnö, University of Gothenburg, SE 452 96, Strömstad, Sweden, ‡Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands, §Netherlands Centre for Biodiversity Naturalis, PO Box 9517, 2300RA Leiden, The Netherlands*

Abstract

The role of natural selection in speciation has received increasing attention and support in recent years. Different types of approaches have been developed that can detect genomic regions influenced by selection. Here, we address the question whether two highly different methods—F_{ST} outlier analysis and admixture analysis—detect largely the same set of non-neutral genomic elements or, instead, complementary sets. We study genetic architecture in a natural secondary contact zone where extensive admixture occurs. The marine bivalves *Macoma balthica rubra* and *M. b. balthica* descend from two independent trans-Arctic invasions of the north Atlantic and hybridize extensively where they meet, for example in the Kattegat–Danish Straits–Baltic Sea region. The Kattegat–Danish Straits region forms a steep salinity cline and is the only entrance to the recently (ca. 8000 years ago) established brackish water basin the Baltic Sea. Salinity along the contact zone drops from 30‰ (Skagerrak, *M.b.rubra*) to 3‰ (Baltic, *M.b.balthica*). Both outlier analysis and genomic clines analysis suggest that large parts of the genome are influenced by non-neutral effects. Contrasting samples from well outside the hybrid zone, outlier analysis detects 16 of 84 amplified fragment length polymorphism markers as significant F_{ST} outliers. Genomic clines analysis detects 31 of 84 markers as non-neutral inside the hybrid zone. Remarkably, only three markers are detected by both methods. We conclude that the two methods together identify a suite of markers that are under the influence of non-neutral effects.

Keywords: admixture mapping, ecological speciation, F_{ST} outliers, genomic clines, population genomics

Received 18 March 2011; revision received 17 February 2012; accepted 1 March 2012

Introduction

Speciation with gene flow has in recent years rapidly gained recognition as a result of theoretical as well as empirical developments (Schluter 2009; Via 2009). This has stimulated the study of the relative importance of isolation and selection in the speciation process (Coyne & Orr 2004). Much is now known about the facilitating conditions for the origination and maintenance of new species under the pressure of gene flow, and empirical examples continue to accumulate (Van Doorn *et al.* 2001; Bolnick & Fitzpatrick 2007; Schluter 2009). However, our understanding of the genetic architecture of partial reproductive isolation is still limited (Schluter & Conte 2009; Bernatchez *et al.* 2010; Wolf *et al.* 2010). Zones of introgressive hybridization in particular are of interest for studying aspects of reproductive isolation, because there, the incipient taxa keep up their identity while species differences are simultaneously maintained

Correspondence: Pieternella C. Luttikhuizen, Fax: +31(0)222319674; E-mail: luttik@nioz.nl

© 2012 Blackwell Publishing Ltd

and broken up (Harrison 1990; Payseur 2010). Marine hybrid zones are especially elusive in this respect because of the potential for large dispersal distances; if hybridization, introgression and dispersal are so common, what is it that keeps the species apart? Answering this question begins with finding regions in their genomes that are influenced by other processes than neutral ones alone. Different types of approaches are now available to detect such genomic regions. Here, we address the question whether two highly different methods (F_{ST} outlier analysis and admixture analysis) detect largely the same set of non-neutral genomic elements or, instead, complementary sets.

The basic premise of population genomics is that allele frequencies at neutral loci follow a consistent distribution while at loci under selection, or linked to selected loci, allele frequencies can be found to deviate from that distribution (Butlin 2010). Outlier approaches begin by producing a hypothetical distribution of differentiation statistics (usually F_{ST}) using the same overall levels of variability and differentiation as measured in an empirical data set and continue by determining which of the empirical loci are significantly more (or less, though the feasibility of this is debated) differentiated than the hypothetical loci (e.g. Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Bonhomme *et al.* 2010). While these outlier loci are generally interpreted to be under environmental selection, they may also stem from endogenous effects related to genetic incompatibilities (Bierne *et al.* 2011). F_{ST}-based comparisons have proven very successful for detecting genomic regions under selection in populations in divergence up until the early speciation stage (e.g. Wilding *et al.* 2001; Egan *et al.* 2008; Bradbury *et al.* 2010; Mehner *et al.* 2010). The approach has also been used to study a secondary contact zone empirically and theoretically by Murray & Hare (2006). They showed that the mean and variance of F_{ST} distributions are elevated in secondary contact populations, and their decay may take hundreds of generations. However, the F_{ST} distribution among loci in secondary contact zones is not distinguishable from that in intraspecific comparisons. This thus enables the detection of F_{ST} outlier loci in secondary contact zones (Murray & Hare 2006; Colbeck *et al.* 2011).

An alternative method for detecting genomic regions affected by genetic incompatibilities or selection has been developed more recently and is specifically tailored for admixed populations such as exist in secondary contact zones: admixture mapping or genomic cline analysis (Buerkle & Lexer 2008). In secondary contact zones, genomes with different evolutionary histories meet and interact. This interaction leaves signatures in the genomic architecture that can be used to make inferences about the functionality of particular regions of the genome. Admixture is a term used for the mixing of genomes of divergent parental taxa; it is a broader version of the term 'hybrids' that includes cases that would not normally fall under hybridization such as divergent populations that are fully interfertile. Admixture mapping is being applied increasingly widely to human populations (Hoggart *et al.* 2004; Seldin 2007; Rosenberg *et al.* 2010). The method also seems highly relevant for natural populations of other taxa (Rieseberg & Buerkle 2002; Buerkle & Lexer 2008) and is now gradually being explored (Lexer *et al.* 2007; Nolte *et al.* 2009; De Carvalho *et al.* 2010; Teeter *et al.* 2010; Gagnaire *et al.* 2011). The method includes techniques that are particularly interesting for taxa that are impossible or difficult to culture, and thus are not suited for the production of physical genetic linkage maps.

Genomic clines analysis is a form of admixture mapping that examines the behaviour of individual loci against genome-wide status of introgression (Buerkle & Lexer 2008; Gompert & Buerkle 2009, 2010). Specifically, it tests a null model of neutrality in which genotype probabilities at focal loci do not deviate significantly from what may be expected based on genome-wide hybrid status. Loci that emerge as deviant are believed to be linked to genomic regions involved in reproductive isolation (Buerkle & Lexer 2008; Teeter *et al.* 2010). Alternatively, they might be loci that introgress more or less readily than neutral loci due to selective effects within the hybrid zone (Bierne *et al.* 2011). Genomic clines analysis promises to be a strong improvement of the hitherto frequently used approach of geographic clines comparisons among loci.

The Kattegat–Danish Straits transition zone, linking the northeast Atlantic with the Baltic Sea—is ideally suited for studying evolution in action (e.g. Bekkevold *et al.* 2005; Gaggiotti *et al.* 2009). It represents a strong environmental gradient associated with ample opportunity for local adaptation, and the environment on the Baltic side is of a very young age as a habitat for marine taxa (8000 years) (Ignatius *et al.* 1981). The environmental gradient consists primarily of a salinity gradient running from fully saline in the Atlantic to brackish in the Baltic proper and almost freshwater in the northern Gulf of Bothnia (BG). Traditionally, flora and fauna of the Baltic Sea were viewed as recent extensions from the eastern Atlantic. However, that this view is too simplistic is now abundantly evident. The Baltic Sea not only houses, for the majority of taxa, populations that have evolved differently as a result of isolation and bottlenecks (Johannesson & André 2006), but even produced the most recently evolved marine species known to date (the brown macroalga *Fucus radicans*, originated ca. 400 years ago, Pereyra *et al.* 2009).

© 2012 Blackwell Publishing Ltd

A special class of Baltic taxa is formed by the two bivalves, *Macoma balthica balthica* (*M.b.b.*) and *Mytilus trossulus* (*M.t.*). Atlantic *Mytilus edulis* and *Macoma balthica rubra* (*M.b.r.*) meet Baltic *M.t.* and *M.b.b.*, respectively, at the entrance to the Baltic Sea, where they each hybridize and show introgressive hybridization (Väinölä & Hvilsom 1991; Riginos *et al.* 2002; Nikula *et al.* 2008; Stuckas *et al.* 2009; Väinölä & Strelkov 2011). Both taxa have dispersive, free-swimming larval stages, implying high potential for gene flow. The similar nature of the two unrelated taxa gives added value to the study system. Here, we present genome scan data for *Macoma* spp. across this hybrid zone with the aim of detecting genomic regions involved in maintaining species identity in the face of gene flow and introgression.

The genus *Macoma*, comprised of infaunal (i.e. burrowing in soft sediments) bivalves, originated in the Pacific Ocean, and some of its species have spread to the north Atlantic via the Arctic following the opening of the Bering Strait around 3.5 Ma as part of what is known as the trans-Arctic biotic interchange (Durham & MacNeil 1967; Vermeij 1991). *Macoma balthica* invaded the north Atlantic in this way at least three times (Väinölä & Varvio 1989; Luttikhuizen *et al.* 2003a; Väinölä 2003; Nikula *et al.* 2007). The Baltic Sea basin is inhabited by descendants of the most recent, probably Holocene, invasion and the northeast Atlantic coasts by the descendants of an earlier invasion, some 2 Ma (Luttikhuizen *et al.* 2003a; Väinölä 2003; Nikula *et al.* 2007). Where these two groups (now referred to as *Macoma balthica balthica* and *Macoma balthica rubra*, respectively) meet, they show evidence of past and/or ongoing hybridization, such as at the entrance of the White Sea, the Barents Sea (Strelkov *et al.* 2007) and in the Kattegat–Baltic area (Nikula *et al.* 2008). Nikula *et al.* (2008) analysed introgression by population-wise geographical cline analyses of 10 allozyme loci in the Kattegat–Baltic area, uncovering a complex pattern of introgression that differs between northern and southern coasts as well as between deeper and shallower waters.

Here, we show the results of genome scans using amplified fragment length polymorphisms (AFLPs) that enable both F_{ST} outlier analysis and individual-based genomic clines analysis. A denser mapping of the genome (*M. balthica* has 18 chromosomes; Luttikhuizen & Pijnacker 2002) will allow gaining insight into genetic architecture of partial reproductive isolation and other traits, and the possibility of screening many more loci with AFLPs than with allozymes may provide a better view of the behaviour of the genome as a whole. Allozymes are not infrequently under the influence of natural selection (Place & Powers 1979; Hilbish & Koehn 1985; Johannesson *et al.* 1995; Riginos *et al.* 2002), and the previous studies of the Baltic *Macoma* hybrid zone specifically chose to examine allozyme loci that showed the greatest differentiation between the main evolutionary lineages (Väinölä 2003; Nikula *et al.* 2008). By comparing F_{ST} outlier detection and genomic clines analysis on genomic scans of *Macoma* populations inside and outside the Kattegat–Baltic Sea area, we explore the possibilities of finding genomic regions that are under the influence of selection and may be involved in preserving species status in spite of homogenizing forces such as introgressive hybridization. Specifically, we test the hypothesis that these methods detect complementary rather than the same sets of non-neutral genomic elements.

Material and methods

Samples

A total of 644 *Macoma* spp. were collected at 21 locations between May and August 2007 (Table 1, Fig. 1). Sampling was done by sieving sediment that was brought up by a corer in shallow water, by a grab operated from a vessel in deeper water or by sweeping with a broom in the intertidal. Sampling was done with high spatial density inside the hybrid zone, while baseline samples were taken in the Wadden Sea (WS) and the Gulf of Bothnia (BG). Entire individuals were stored in pure ethanol until further processing. A piece of siphon or gill tissue was used for DNA extraction.

AFLP genotyping

Tissue samples were digested, and DNA was extracted using a CTAB-based protocol. In brief, tissue was digested in 800 μL extraction buffer containing 2 μL β-mercaptoethanol and 6 units proteinase K, and DNA was extracted twice using chloroform isoamylalcohol, followed by isopropanol precipitation. Finally, the DNA pellet was washed twice with ethanol and resuspended in 50 μL 10% TE. DNA yields typically ranged from 100 to 300 ng/μL. DNA extracts were transferred to 96-well plates. AFLP procedures followed Vos *et al.* (1995) using the Plant Mapping Kit (Applied Biosystems) to produce DNA fingerprints for each of five selective primer combinations: *Eco*RI-AGC/*Mse*I-CAC, *Eco*RI-AAG/*Mse*I-CAC, *Eco*RI-AGC/*Mse*I-CAG, *Eco*RI-AAG/*Mse*I-CAG and *Eco*RI-ACA/*Mse*I-CAG. Restriction was carried out in a 10 μL total volume reaction containing 1× enzyme buffer, 1 U MseI (Invitrogen), 5 U EcoRI (Invitrogen) and 2 μL 1:10 dH_2O-diluted DNA, incubated for 2 h at 37 °C and terminated for 10 min at 65 °C. Ligation was done in a 20-μL volume containing the entire restriction product plus 1 U T4 DNA ligase (Invitrogen), 0.5× T4 enzyme buffer, 1 μL *Mse*I adaptor

© 2012 Blackwell Publishing Ltd

Table 1 Sampling locations for *Macoma balthica*, sampling number (*N*), sampling depth, local salinity and geographical coordinates

Sample	Location	*N*	Depth	Salinity	N	E
WS	Wadden Sea	50	Intertidal	28.5‰	53°27.785′	06°47.938′
GS	Getterön	24	Shallow	18.1‰	57°07.347′	12°14.136′
BS	Båstad	28	Shallow	14.3‰	56°26.067′	12°52.207′
RW	Öresund	28	12 m	30‰	56°02.153′	12°41.231′
KH	Öresund	32	11.5 m	23‰	56°00.550′	12°42.027′
BB	Barsebäck	28	Shallow	9.4‰	55°46.342′	12°55.528′
LM	Lommabukten	27	Shallow	3.2‰	55°41.291′	13°03.077′
FK	Falsterbokanal	29	Shallow	8.5‰	55°23.920′	12°56.877′
WB	Western Baltic	28	44 m	13.3‰	54°53.333′	13°06.388′
HH	Hörte Hamn	30	Shallow	4.1‰	55°23.185′	13°32.755′
BO	Bökenäs	30	Shallow	6.2‰	56°09.421′	15°23.231′
KS	Kalmar	30	15 m	6.8‰	56°37.832′	16°22.635′
JO	Jogersö	30	Shallow	6.3‰	58°39.902′	17°03.189′
UT	Utterviken	30	23 m	6.2‰	58°50.451′	17°33.168′
GR	Gråskar	26	40.5 m	6.3‰	58°47.550′	17°39.551′
KF	Krabbfjärden	28	47.1 m	7.1‰	58°45.938′	17°42.147′
HV	Hällsviken	28	22 m	6.3‰	58°49.731′	17°31.620′
VN	Viksnäs	28	Shallow	6.0‰	58°51.594′	17°30.803′
AS	Askö	30	Shallow	6.0‰	58°48.998′	17°39.913′
VA	Väddö	30	Shallow	5.3‰	59°55.210′	18°54.850′
BG	Gulf of Bothnia	50	24.0 m	4.4‰	63°29.376′	19°48.224′

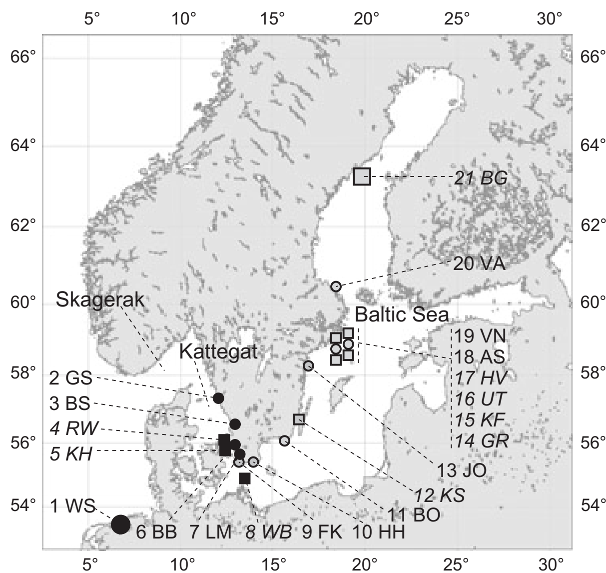

Fig. 1 Map of study area showing sampling locations for *Macoma balthica* in Kattegat–Baltic hybrid zone. Black symbols: western part of hybrid zone; grey symbols: eastern part of hybrid zone; round symbols: intertidal or shallow water locations; square symbols: deep water locations; larger symbols at WS and BG indicate reference samples located furthest outside hybrid zone. Locations are numbered in geographical order along the hybrid zone from Atlantic into Baltic.

and 1 μL *Eco*RI adaptor. The ligation reaction was incubated for 2 h at 37 °C, terminated at 65 °C for 10 min and then diluted 10× in 10% TE. Preamplification was carried out in a total volume of 20 μL containing 1× PCR buffer, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP, 0.5 μL *Eco*RI-A primer, 0.5 μL *Mse*I-C primer, 0.4 U Taq DNA polymerase (TaKaRa, Japan) and 4 μL 10× diluted restriction-ligation product. Preamplification PCR consisted of 72 °C for 2 min; 22 cycles of 94 °C for 20 s, 56 °C for 30 s and 72 °C for 2 min; followed by a final 60 °C for 30 min. Preamplification products were subsequently diluted 20× in dH_2O. Selective amplification reactions were carried out in a total volume of 10 μL and contained 2.5 mM $MgCl_2$, 1× PCR buffer, 0.25 mM of each dNTP, 1 μM *Eco*RI selective primer, 5 μM *Mse*I selective primer, 0.04 U Taq DNA polymerase and 1.5 μL diluted preamplification template DNA. Selective amplification PCR cycling consisted of an initial 94 °C for 2 min, then 10 times a first round of cycles, 20 times a second cycle, followed by a final 60 °C for 30 min. The first selective PCR round of cycle was a step-down procedure of 94 °C for 20 s, 30 s step-down from 66 to 56 °C and 72 °C for 2 min. The second round of cycles consisted of 94 °C for 20 s, 56 °C for 30 s and 72 °C for 2 min. The *Eco*RI selective primers were fluorescently labelled, and the amplicons were visualized on a Beckman Coulter CEQ8000 capillary electrophoresis system alongside a 400 size standard (Beckman Coulter).

DNA extractions were organized on 96 well plates grouped per sampling location, and the locations randomly distributed over the plates. This layout was kept during restriction, ligation, preamplification, selective amplification and capillary electrophoresis. Reactions

© 2012 Blackwell Publishing Ltd

were carried out simultaneously per four plates, and all genotyping was completed within 1 month. This procedure was followed to minimize systematic genotyping variation. For all peaks, fragment length and peak height were estimated using the manufacturer's software and exported without calling threshold. Binning and noise filtering were carried out subsequently. Peaks were binned into fragment size classes on the basis of estimated fragment size and proximity to other bins: bins were retained when modes fell within 1 bp and if neighbouring mode frequency distributions did not overlap. Initially, across the five primer pairs, 317 bins were scored in the 60–400 bp range. Peak heights were normalized across runs within primer pairs. Subsequent noise filtering and locus selection broadly followed Whitlock *et al.* (2008). The AFLP genotyping procedure was replicated partially for 39 DNA extracts (using a subset of the five primer pairs), yielding a total of 2924 band presence/band absence duplicate genotyping assays. Genotyping error was minimized by varying the phenotyping calling threshold and filtering loci according to noise content. Optimal phenotype calling threshold was 700 rfu (normalized relative fluorescence units; note that this is the value after normalization) and optimal locus selection threshold 20% (see Whitlock *et al.* 2008). The final data set thus obtained consisted of 90 loci scored with an estimated replicability of 95.7%, an error rate typical of the AFLP technique (Bonin *et al.* 2007).

The overall variation in the data was visualized per primer pair in multidimensional scaling (MDS) plots constructed using Euclidean distances (PRIMER version 5.2.9, Clarke & Gorley 2003) between fingerprints of individuals consisting of the AFLP band presence/absence data.

Outlier test for selection

Loci were tested for natural selection based on the distribution of differentiation statistic F_{ST}, contrasting the two extreme populations WS and BG. Under a neutral model of evolution, F_{ST} is expected to relate in a predictable way to locus-specific heterozygosity. The coalescent approach of Beaumont & Nichols (1996) generates joint heterozygosity and F_{ST} distributions under simple genetic null models, which are robust to a range of population structures, demographic histories and mutation rates (Beaumont & Nichols 1996; Murray & Hare 2006). We obtained a null distribution using a two-deme model of neutral differentiation with migration, generating 50 000 loci sampled for $n = 50$ individuals per deme (identical to the sample size of the samples WS and BG being compared). Global critical frequency of the most common allele was set to 0.99, which resulted in the exclusion of six of 90 loci in the combined WS and BG subset of data. The population mutation parameter $N_e\mu$ was assumed to be 0.088 as estimated from a mean homozygosity among polymorphic AFLP loci of 0.740. Mean homozygosity was obtained by taking the average of the homozygosities in the reference samples (0.714 for BG and 0.765 for WS). The mutation parameter was then estimated from the relationship theta = 4 $N_e\mu = (1 - H)/H$, where H equals homozygosity, derived by Kimura (1968). The implementation of the approach used was the simulation program Dfdist (Beaumont & Nichols 1996). This method was chosen because, as one of the earliest ones developed, it is one of the most widely applied ones; for alternative approaches, see, for example, Beaumont & Balding (2004), Foll & Gaggiotti (2008) and Bonhomme *et al.* (2010). Comparisons of the performances of these methods can be found in, for example, Perez-Figueroa *et al.* (2010), Nunes *et al.* (2011) and Buckley *et al.* (2012).

False discovery rate correction (FDR) of *P*-values followed Benjamini & Hochberg (1995), which was chosen because it controls the expected proportion of falsely rejected positives rather than the family-wise error rate. This leads to increased power (Benjamini & Hochberg 1995) and therefore is a less conservative procedure than other Bonferroni methods. One aim of this work was to compare methods, and therefore, the same FDR method was used for both the outlier and genomic clines analyses.

Genomic clines analysis

Genome-wide admixture status of individual bivalves in the Kattegat–Baltic admixture zone—was estimated as a maximum likelihood estimate of hybrid index *h* (Buerkle 2005). Excess admixture of individual loci was evaluated following the genomic clines method of Gompert & Buerkle (2009), implemented in the R-script 'Introgress' (Gompert & Buerkle 2010). Excess admixture is defined as the extent to which alleles or genotypes are introgressed into a foreign genetic background or, alternatively, are found in their native genetic background (Buerkle & Lexer 2008). A genomic cline is the regression of observed single-locus genotype frequencies against those expected based on genome-wide hybrid index data. The hybrid index used here is based on a maximum likelihood approach and is applicable to dominant markers that need not exhibit fixed differences between the parental taxa. The likelihood function is determined based on an individual's unknown genotype and the marker frequencies in both parental taxa (for further details see Buerkle 2005). In 'Introgress', multinomial regression is used to estimate the likelihood of the model given the data—in this case, the

© 2012 Blackwell Publishing Ltd

presence or absence data of dominant AFLP markers. The null model was produced using a parametric procedure (applicable in case not all markers examined exhibit fixed differences; Gompert & Buerkle 2009) and consisted of neutral introgression; that is, all loci are introgressed to the same extent as the genome as a whole. Significance of deviations was evaluated by accounting for multiple comparisons following FDR correction (Benjamini & Hochberg 1995), similarly as with the outlier detection analysis.

Results

AFLP data

Details on the levels of polymorphism detected with the five AFLP primer pairs used are given in Table 2.

Figure 2 shows a representative example of a multidimensional scaling (MDS) plot as a visualization of the structure of the AFLP data. Depicted are data for primer pair *Eco*RI-AAG/*Mse*I-CAC with all 74 bins retained. MDS plots for the remaining four primer pairs were highly comparable (not shown). Genome scans for individuals collected at reference locations are clearly separated with no appreciable overlap (top panel of Fig. 2). Genome scans inside the hybrid zone are scattered throughout the multidimensional space (other panels of Fig. 2) and, while individuals of many locations occupy a particular region within that space (e.g. sample Gråskar, GR), some are disseminated to smaller or greater extent (e.g. sample Hörte Hamn, HH).

Outlier detection

Trimmed weighted mean F_{ST} for 84 loci (Weir & Cockerham 1984) was 0.10, and this mean was used to produce the null expectations for F_{ST} and heterozygosity. Contrasting the WS and the BG samples, 16 of the 84 loci examined were deemed to be outliers compared with the simulated null distributions (false discovery rate correction of P-values followed Benjamini & Hochberg (1995)) (Fig. 3). The distribution of P-values among loci was such that approximately half (59%) was larger than 0.5 for the 69 neutral loci, and no trend could be discerned between the proportion of loci with P-values >0.5 and heterozygosity, indicating good model fit.

Genomic clines

Hybrid index was evenly distributed across individuals in the hybrid zone (Table 3; Fig. 4; note that this graph does not include reference samples WS and BG). The even hybrid index distribution is necessary for carrying out the genome clines analysis, because if all hybrid genomic backgrounds are represented in the analysis, excess admixture can be estimated most accurately (Buerkle & Lexer 2008). An elevated portion of pure genomes of both types was encountered inside the hybrid zone, indicating influx from both sides: out of a total of 544 individuals, 21 (3.9%) pure 'Atlantic' genomes were found, ten of which among the western hybrids and a further eight at locations FK and HH, that is, westernmost within the eastern part of the hybrid zone, and 63 (11.6%) pure 'Baltic' genomes were found, 61 of which among the eastern hybrids. Most notable are the remaining five pure individuals, two of which are pure 'Baltic' found at LM in the Öresund, and three pure 'Atlantic' found at KS, KF and VA in the Baltic proper.

Genomic clines analysis shows that 53 of 84 polymorphic AFLP loci conform to a neutral model of introgression (Table 4), meaning that the probability to observe band presence for a particular marker can be predicted well from the hybrid status of an individual's genome as a whole. The remaining 31 loci significantly deviate from this model and can be classified (see Gompert & Buerkle 2009; Nolte *et al.* 2009) into three subgroups: positively selected loci (N = 18), negatively selected loci (N = 12) and one epistatic locus (C187). Positively selected loci have higher band presence frequencies than might be expected from their genomic background (e.g. locus A244 in Fig. 5). Similarly, negatively selected

Table 2 Levels of variation detected using five amplified fragment length polymorphism (AFLP) primer pairs. F_{ST} estimates of sample differentiation are given for the comparison of reference samples WS and BG, while P_1 is estimated for all 21 samples

Primer pair	N_{loci}	Mean F_{ST}	F_{ST} range	Mean P_1	P_1 range
*Eco*RI-AGC/*Mse*I-CAC	21	0.125	−0.037 to 0.56	0.57	0.031–0.98
*Eco*RI-AAG/*Mse*I-CAC	11	0.102	0.0022 to 0.40	0.76	0.32–0.95
*Eco*RI-AGC/*Mse*I-CAG	23	0.142	−0.017 to 0.52	0.49	0.083–0.98
*Eco*RI-AAG/*Mse*I-CAG	23	0.176	−0.019 to 0.71	0.70	0.26–0.98
*Eco*RI-ACA/*Mse*I-CAG	6	0.118	−0.012 to 0.55	0.54	0.15–0.96

F_{ST} = differentiation statistic as obtained during outlier detection analysis (see Materials and methods), N_{loci} = number of loci, P_1 = frequency of band presence. Ranges give minimum and maximum values among samples.

© 2012 Blackwell Publishing Ltd

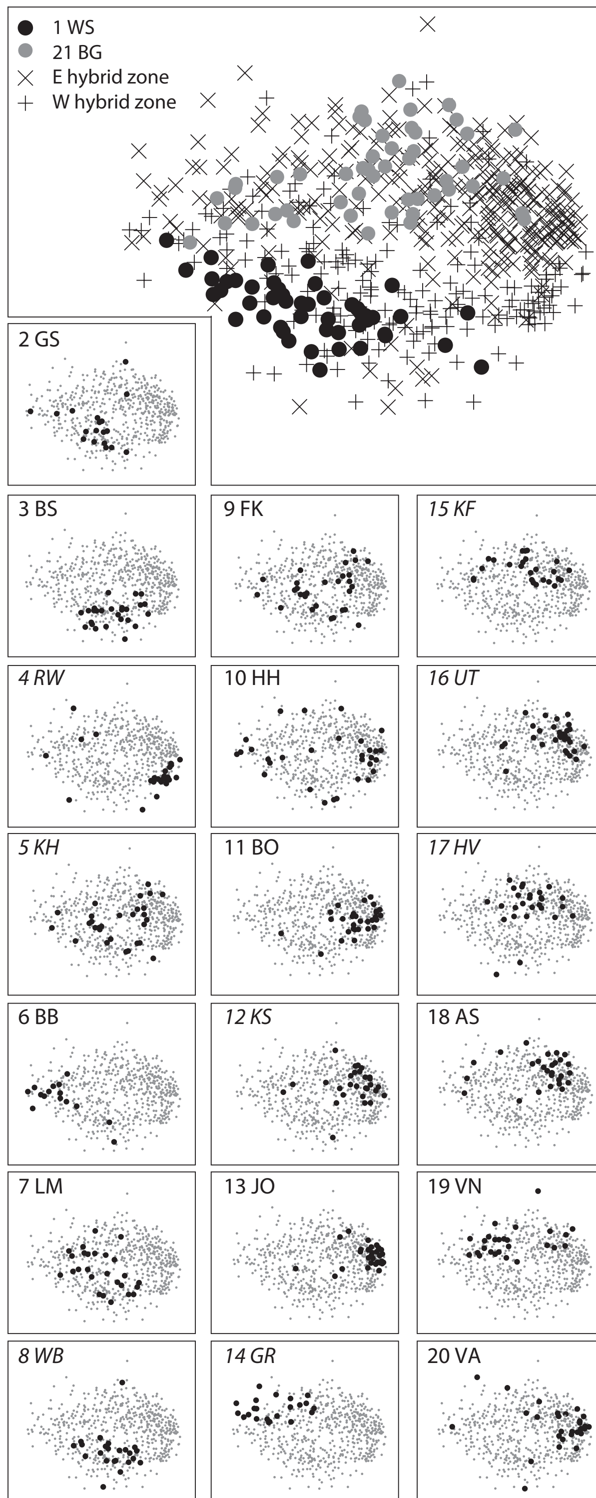

Fig. 2 Multidimensional scaling plots of variation among individual genome scans of *Macoma balthica* in Kattegat–Baltic hybrid zone. Data for one of the five primer pairs (*Eco*RI-AAG/*Mse*I-CAC) shown, with all 74 bins retained. Top panel shows reference samples (black symbols: Wadden Sea; grey symbols: Gulf of Bothnia, Baltic Sea) among inner hybrid zone samples (plusses: western locations; crosses: eastern locations; corresponding to black and grey symbols in Fig. 1, respectively). Remaining panels display genome variation along hybrid zone structured according to sampling location; every panel shows individuals for a particular sampling location (large dots) among all other individuals (small dots). 'W hybrid zone' = individuals in western part of hybrid zone (locations numbered 2–8); 'E hybrid zone' = individuals in eastern part of hybrid zone (locations numbered 9–20).

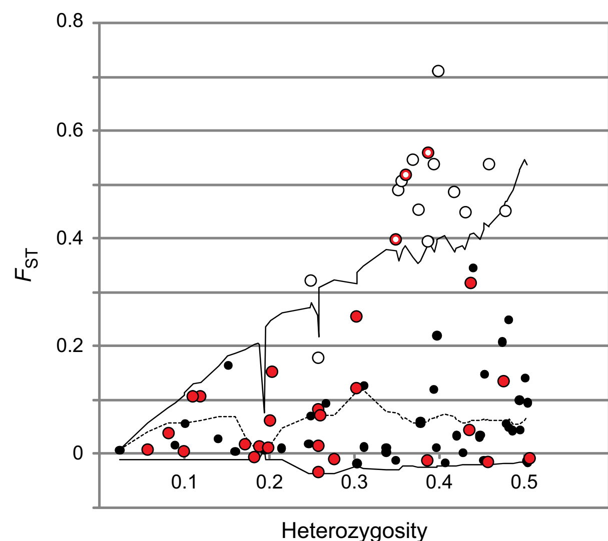

Fig. 3 Results of outlier analyses for *Macoma balthica* across Kattegat–Baltic hybrid zone. Reference samples WS and BG were contrasted to find loci that are significant outliers on the basis of simulated neutral expectations in relation to heterozygosity. Lines delimit 95% quantile of neutral expectations (broken line: median); dots are empirical data for 84 polymorphic AFLP markers. Symbols with white centres ($N = 16$ markers) are significant F_{ST} outlier loci. Shaded symbols indicate loci ($N = 31$ markers) that behave non-neutrally in a genomic sense inside the hybrid zone as concluded from genomic clines analyses (for further details, see text). Black symbols ($N = 37$) are loci that are outliers in neither analysis, and shaded symbols with white centres ($N = 3$) are outliers in both analyses.

loci have lowered frequencies compared with genome-wide status (e.g. locus C139 in Fig. 5), while, for epistatic loci, marker frequency and genomic background show evidence of an interaction (locus C187 in Fig. 5).

Three loci (A146, A244 and C139) emerged as significant from both F_{ST} outlier and genomic clines analysis (open red symbols in Fig. 3). In addition, there are 13 loci that are significantly more differentiated between the reference samples than might be expected from a neutral model of evolution while behaving neutrally within hybrid genomic backgrounds (Table 4, open black symbols in Fig. 3). Conversely, out of the 31 loci

© 2012 Blackwell Publishing Ltd

Table 3 Hybrid index (h) with standard deviation (SD) along the hybrid zone between the Atlantic Ocean and the Baltic Sea in *Macoma balthica*. Numbers and abbreviations of sampling locations correspond to Table 1 and Fig. 2.

Location		h_{mean}	SD
2	GS	0.35	0.23
3	BS	0.34	0.20
4	RW	0.38	0.22
5	KH	0.39	0.18
6	BB	0.33	0.24
7	LM	0.55	0.28
8	WB	0.18	0.15
9	FK	0.32	0.22
10	HH	0.35	0.32
11	BO	0.70	0.20
12	KS	0.79	0.24
13	JO	0.78	0.20
14	GR	0.68	0.25
15	KF	0.53	0.27
16	UT	0.74	0.20
17	HV	0.61	0.27
18	AS	0.83	0.26
19	VN	0.88	0.12
20	VA	0.57	0.29

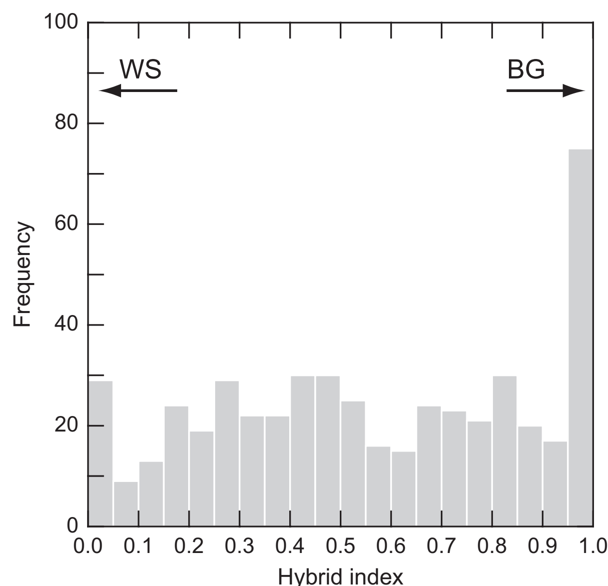

Fig. 4 Frequency distribution of hybrid index in Kattegat–Baltic *Macoma balthica* hybrid zone. Hybrid index is maximum likelihood estimate following Buerkle (2005) in relation to reference locations Wadden Sea (WS; $h = 0$) and Gulf of Bothnia (BG; $h = 1$).

that behave non-neutrally in hybrids with respect to genomic composition, there are 28 that are not significantly differentiated between reference populations WS and BG (closed red symbols in Fig. 3).

Discussion

This study set out to test the hypothesis that two methods for finding non-neutral genomic elements identify complementary rather than similar sets of markers. The data presented here corroborate the hypothesis. The set of AFLP markers identified as non-neutral by F_{ST} outlier analysis is a largely different set than the set detected by genomic clines analysis. Apparently, the two methods detect non-neutral genomic elements in different ways, which implies that they may be detecting different classes of traits. Speciation studies could thus benefit from using both approaches to identify genome regions of interest, not only to increase the number of target markers but also because the range of trait types covered may be broadened.

The only other study available to date that has applied both admixture analysis and F_{ST} outlier analysis to genome scan data is on eels *Anguilla marmorata* (Gagnaire *et al.* 2011). In one of their two population comparisons, the study also finds several markers (four of eight) that appear as deviant from genomic clines analysis but not as outliers from F_{ST} outlier analysis. However, in their other comparison, all three deviant markers are also significant outlier loci. Although their study was not set up to directly compare both methods, the difference between the methods may be less extreme than in our study on the bivalve *Macoma balthica*.

The genomic clines analysis presented here shows that large portions of the genome behave non-neutrally inside this marine bivalve hybrid zone. As much as 38% of the AFLP markers do not conform to a neutral mixing model of hybrid genotypes (Table 4). Conversely, assuming the markers are randomly scattered across the chromosomes, it also implies that large parts of the genome are admixed. Furthermore, genomic differentiation appears to be genetically organized in a complex way, as evidenced from the diversity in models of selection indicated by the genomic clines analysis (Fig. 5, Fig. S1, Table 4). F_{ST} outlier analysis on two reference samples collected outside the contact zone identifies 19% of AFLP markers as being significantly more differentiated than expected on the basis of a neutral F_{ST} distribution. The overlap between the sets detected by both methods consists of <4% of all markers.

The fact that there is little overlap between the sets of markers detected by the two methods suggests that different types of genomic elements are identified. Genomic clines analysis uses samples consisting of hybrids (both first and later generation hybrids as well as backcrosses) and may thus be biased towards finding markers linked to loci involved in partial reproductive isolation (i.e. endogenous effects). Outlier detection

© 2012 Blackwell Publishing Ltd

Table 4 Non-neutrality in the *Macoma balthica* genome as evidenced from outlier and genomic clines analyses (significance of deviance at study-wide $P < 0.05$ level with false discovery rate correction). Right column gives AFLP markers whose frequency is significantly correlated with salinity. Markers printed in bold are significantly deviant in both outlier and genomic clines analysis. Positive and negative selection indicate higher and lower prevalence, respectively, of band presence in the hybrid zone than expected on the basis of genome-wide admixture status. Epistasis is used to signify an interaction between genome admixture status and band prevalence

	F_{ST} outlier analysis	Genomic clines analysis			Salinity
Neutral	68 markers	53 markers			57 markers
Deviant	Differential selection 16 markers:	Positive selection 18 markers:	Negative selection 12 markers:	Epistasis 1 marker:	Correlated 27 markers:
	A140	A80	A83	C187	A117
	A146	A87	A120		A122
	A176	A106	**A146**		A140
	A244	**A244**	A318		A146
	B60	B79	C136		A176
	C78	C128	**C139**		B60
	C119	C130	C158		B122
	C139	C231	C195		B147
	C217	C301	C204		C67
	C244	D68	D90		C139
	D86	D77	D140		C158
	D102	D80	D236		C162
	D116	D112			C187
	D121	D143			C244
	D151	D197			C301
	E254	D255			D72
		E89			D102
		E219			D116
					D121
					D123
					D129
					D140
					D151
					D255
					E76
					E215
					E219

studies differentiated populations and may be more prone to detect genomic elements involved in local adaptation (due to exogenous effects). It is probably, however, as pointed out recently by Bierne *et al.* (2011), that in contact zones both reproductive isolation and local adaptation play important roles. After all, the environment is not constant throughout many hybrid zones—it certainly is not for the particular zone studied here with its strong salinity gradient. Moreover, it must be taken into account that the local environment differs between where the 'pure' populations reside and where the hybrid zone is. That means that the genomic clines analysis might be detecting local adaptation genes for different selective effects. In summary, to conclude that the markers detected by outlier analysis are linked to adaptation genes and those detected by genomic clines analysis are linked to reproductive isolation genes is not possible. However, there might be a difference between the methods in terms of a relative bias towards the detection of markers under the influence of endogenous vs. exogenous effects. More studies are needed to test this.

Relatively strong and complex genomic differentiation is a view that may seem at odds with the earlier conclusion drawn from allozyme data of a hybrid swarm with amalgamation of the previously isolated genomes (Strelkov *et al.* 2007; Nikula *et al.* 2008). However, Nikula *et al.* (2008) also noted genotypic disequilibria in the steepest part of the cline and concluded that larval mixing and limited interbreeding must play a role in that area, in spite of the amalgamation they observed in the inner hybrid swarm. The isolation is corroborated by our genomic clines analysis, but we cannot evaluate the amalgamation because the dominant nature

© 2012 Blackwell Publishing Ltd

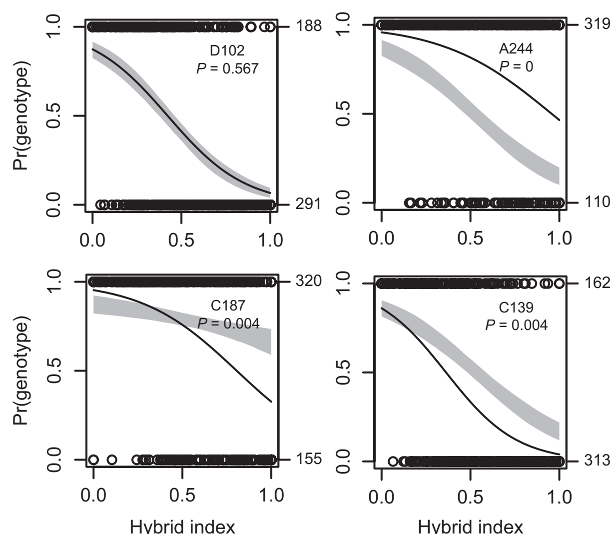

Fig. 5 Examples of fitted genomic cline curves for *Macoma balthica* Kattegat–Baltic hybrid zone. Hybrid index is genome-wide hybrid status of individuals and ranges from 0 on the western side (defined as Wadden Sea) to 1 on the eastern side (defined as Gulf of Bothnia). *Y*-axis indicates the probability of observing one of the two possible genotypes (band presence/band absence) of the focal locus. To the right of each graph is indicated the number of observations per genotype class. Each circle represents an individual observation. Neutral model means that an individual locus behaves in accordance with genome-wide hybrid status across individuals. Grey areas indicate neutral models, and black lines are lines fitted to the data. D102: example of a locus adhering to neutral model; A244: example of positively selected locus, that is, AFLP band is more often present against genomic background than expected from neutral mixing; C187: example of locus following epistasis pattern, that is, elevated band presence at low hybrid index, switching to decreased band frequency at high hybrid index; C139: example of negatively selected locus.

of the AFLP markers precludes the analyses of genotypic disequilibria. Recent whole genome sequences of two incipient malaria mosquito species revealed similar unexpectedly high levels of genetically complex differentiation (Lawniczak *et al.* 2010), at odds with earlier gene-based data suggesting extensive amalgamation (Turner *et al*. 2005). The presently emerging view of the porous nature of genomes (Feder & Nosil 2010) is exemplified in *M. balthica* as in other taxa (e.g. Noor *et al.* 2007): the not yet completely isolated genomes are open to introgression in spite of considerable and genomically widespread isolation.

Spatial studies on genetic architecture are needed both from cases of direct ecological divergence and from zones of secondary contact. The two *M. balthica* lineages that hybridize in the Kattegat–Baltic zone–probably split at least 2–3.5 Ma (Luttikhuizen *et al.* 2003a; Nikula *et al*. 2007). They are both of Pacific origin and are trans-Arctic invaders of the north Atlantic. The first to colonize the Atlantic following the opening of the Bering Strait was *M.b.r.*, which now inhabits the North Sea and European Atlantic coasts, and the second, *M.b.b.*, arrived very recently (post-Pleistocene) and is now found in the Baltic and White Seas (Strelkov *et al.* 2007; Nikula *et al.* 2008). The secondary contact of these lineages is thus recent compared to their time of isolation. This study shows that genomes in secondary contact may be open to introgression but reproductive isolation can nevertheless be genetically pervasive and that more candidate markers can be identified when combining outlier analysis with genomic clines analysis. Whether this reproductive isolation stems from local adaptation or from endogenous incompatibility or both can only be unequivocally established using other approaches such as experiments (Bierne *et al.* 2011). Nevertheless, the suite of markers detected with both methods probably encompasses a broader array of types of effects than either method on its own.

The genomes of Baltic *Macoma* and North Sea *Macoma* as a whole seem to be less differentiated than their functional genes only; genomic differentiation between the subspecies was here detected for 19% of AFLP markers, with an average F_{ST} over all markers of 0.10, and this degree of differentiation is lower than what has been reported earlier for allozymes and mitochondrial DNA (Luttikhuizen *et al.* 2003a; Väinölä 2003; Nikula *et al*. 2008). The discrepancy is particularly striking considering the fact that our Baltic Sea reference sample was obtained from deep into the BG in a most extreme environment (e.g. a salinity of 4‰). This suggests that non-neutral processes, either differential selection or genetic incompatibilities, play an important role in maintaining species identity across the hybrid zone (Nosil *et al.* 2009; Butlin 2010). The observation that expressed genes may show higher differentiation than the bulk of the genome (though the initial choice of allozymes to study was influenced by their level of differentiation to begin with) is paralleled in the hybridization between the blue mussels *Mytilus edulis* and *Mytilus trossulus* in the Kattegat–Baltic zone (Riginos *et al.* 2002; Riginos & Cunningham 2005), and hence, *Macoma* and *Mytilus* form valuable independent replications of secondary cline formation in this area. Note, however, that in blue mussels this is not an all-or-nothing situation, as noncoding DNA can also be differentiated (Bierne *et al.* 2003). Including comparisons with the Pacific region of origin of Baltic *Macoma* would improve the capacity to disentangle the effects of selection from the effects of random drift that has taken place because isolation of Pacific and Baltic *Macoma*.

As the contact zone is most marked in the area where the salinity gradient shows its steepest cline, salinity

© 2012 Blackwell Publishing Ltd

has been invoked as a possible direct selective agent involved in maintaining the hybrid zone. This should then be reflected in the genomes studied here. Salinity does indeed significantly correlate with many of the AFLP markers analysed in this study (Table 4). When applying the same procedure to account for multiple comparisons as we did for outlier detection and genomic clines analysis (FDR correction; Benjamini & Hochberg 1995), 27 of the 84 markers display a significant correlation between salinity and marker frequency. Of these, 13 are positive and 14 negative correlations; nine are also significant in the outlier analysis and eight in the genomic clines analysis. It must of course be noted that this is not a valid analysis per se, because even if salinity were not a direct selective agent, significant correlations would also be expected simply as a result of colinearity of hybrid zone and habitat. In spite of this, it is striking that marker D102, one of the F_{ST} outlier loci, is most strongly correlated with salinity of all markers ($\rho = -0.799$; $P = 0.00001$). Possibly, this marker resides in a genome region containing a locus that is differentially selected in the different salinity regimes on either side of the transition zone.

To reduce the problem of colinearity between hybrid zone and habitat, we may inspect the correlations between markers and salinity in a subset of samples spanning the geographical region where the salinity cline is steepest, that is, samples three to nine (see Fig. 1). Again, using FDR correction, we find that only five of the 84 markers are significantly correlated with salinity. Of these markers, three (A126, A244 and C217) do not correlated with salinity when all samples are considered. However, the two most strongly correlated markers D102 ($\rho = -0.869$; $P = 0.01106$) and D255 ($\rho = 0.912$; $P = 0.00424$) were also the most strongly correlated in the all-sample comparison. This stresses the likelihood of the genomic region where marker D102 (and possibly also marker D255, though it is not significant in the outlier analysis) resides as containing genetic elements involved in salinity adaptations.

The low-to-intermediate level of overall differentiation observed is relevant for the active discussion on genomic islands of speciation. Theory has shown that differentiated 'islands' in the genome, embedded within an otherwise-undifferentiated genome, may originate by divergence hitchhiking during speciation when multiple loci are under divergent selection (Feder & Nosil 2010). Empirical evidence in this respect is mixed, with some recently diverged species showing genomic islands (or even 'continents') and others one or a few localized divergent regions (e.g. Rogers & Bernatchez 2007; Mäkinen *et al.* 2008; Via & West 2008; Strasburg *et al.* 2009; Lawniczak *et al.* 2010; Michel *et al.* 2010). Furthermore, many loci under weak selection may together contribute to differentiation, and therefore, outlier analysis may not be particularly relevant to indicate when selection is involved or not—that is, loci under weaker selection are not detected by outlier analyses.

The complexity of local population structure is evidenced by the high levels of differentiation among sites within regions (W or E) and even between geographically close samples; see, for example sample 3 vs. 4, 4 vs. 5 and 13 vs. 14 in Fig. 2. This level of small-scale differentiation is consistent with the growing insight that marine species may experience low effective dispersal rates in spite of their large potential for gene flow (e.g. Luttikhuizen *et al.* 2003a,b, 2008; Derycke *et al.* 2005; Peijnenburg *et al.* 2006; Chen & Hare 2008; Larsson *et al.* 2010; Yebra *et al.* 2011; for reviews, see Hauser & Carvalho 2008; Hellberg 2009). Alternatively, complex dispersal patterns combined with recruitment heterogeneity may have produced temporal patterns of so-called chaotic genetic patchiness (Johnson & Black 1982; Muths *et al.* 2009; Hogan *et al.* 2010). The observation of apparently pure genotypes on the 'wrong' side of the hybrid zone, then, poses the question whether these are direct dispersers or local hybrids with an accidental extreme genotype. When examining the frequency distributions of the hybrid indices per location (data not shown), it can be seen that variance in hybrid index within locations is very high, which is consistent with both local extreme hybrids and complex dispersal patterns. The latter could be tested using oceanographic modelling.

Conclusion

Genomic clines analysis in this secondary contact zone has proven a useful addition to finding F_{ST} outliers because the two methods detected some matching but mostly additional non-neutral genomic regions. It will be of particular interest to combine the thus obtained markers with whole genome sequences of the two hybridizing taxa to map the genomic regions here suggested to be involved in their partial reproductive isolation. However, to disentangle the exogenous (selection) and endogenous (genetic incompatibilities) causes of partial reproductive isolation, controlled field and/or laboratory experiments involving artificial crosses are indispensable.

Acknowledgements

We are grateful to J. Albertsson, J. van der Meer and J. Frankowski for sending samples BG (JA and JM) and WB (JF) and to C.C.P. Luttikhuizen for help with field work. Skippers of three research vessels Kronan of Kalmar (S. Tobiasson), Sabella (T. Lindblom) and Limanda (E. Eriksson and M. Karls-

© 2012 Blackwell Publishing Ltd

son) helped to collect the subtidal samples, and molecular laboratory technicians B. Jönsson and A. Bol provided invaluable expertise. The manuscript greatly benefited from the comments of several anonymous reviewers. This work was funded by ESF Conservation Genetics grant 1152 and ALW-Meervoud grant 2011 to PCL and the BONUS project (Baltic Organisations Network for funding Science EEIG) to KJ.

References

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619–1626.

Bekkevold D, André C, Dahlgren TG *et al.* (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution*, **59**, 2656–2668.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate – A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, **57**, 289–300.

Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 1783–1800.

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. *Molecular Ecology*, **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. *Molecular Ecology*, **20**, 2044–2072.

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology Evolution and Systematics*, **38**, 459–487.

Bonhomme M, Chevalet C, Servin B *et al*. (2010) Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society Series B, Biological Sciences*, **277**, 3725–3734.

Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, **21**, 267–280.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684–687.

Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology and Evolution*, **23**, 686–694.

Butlin RK (2010) Population genomics and speciation. *Genetica*, **138**, 409–418.

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, *Acartia tonsa*. *Molecular Ecology*, **17**, 1451–1468.

Clarke KR, Gorley RN (2003) *PRIMER 5 (Plymouth Routines in Multivariate Ecological Research)*. Plymouth Marine Laboratory, Plymouth, UK.

Colbeck GJ, Turgeon J, Sirois P, Dodson JJ (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, **20**, 1976–1987.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer, Sunderland, Massachusetts.

De Carvalho D, Ingvarsson PK, Joseph J *et al.* (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L), a widespread forest tree. *Molecular Ecology*, **19**, 1638–1650.

Derycke S, Remerie T, Vierstraete A *et al.* (2005) Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*. *Marine Ecology Progress Series*, **300**, 91–103.

Durham JW, MacNeil FS (1967) Cenozoic migrations of marine invertebrates through the Bering Strait region. In: *The Bering Land Bridge* (ed. Hopkins DM), pp. 326–349. Stanford University Press, Stanford, California.

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, **62**, 1162–1181.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gaggiotti OE, Bekkevold D, Jorgensen HBH *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

Gagnaire PA, Minegishi Y, Zenboudji S *et al.* (2011) Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla marmorata*. *Evolution*, **65**, 3413–3427.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, **7**, 69–128.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.

Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics*, **40**, 291–310.

© 2012 Blackwell Publishing Ltd

Hilbish TJ, Koehn RK (1985) The physiological basis of natural selection at the LAP locus. *Evolution*, **39**, 1302–1317.
Hogan JD, Thiessen RJ, Heath DD (2010) Variability in connectivity indicated by chaotic patchiness within and among populations of a marine fish. *Marine Ecology Progress Series*, **417**, 263–U289.
Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM (2004) Design and analysis of admixture mapping studies. *American Journal of Human Genetics*, **74**, 965–978.
Ignatius H, Axberg S, Niemistö L, Winterhalter B (1981) Quaternary geology of the Baltic Sea. In: *The Baltic Sea* (ed. Voipio AV), pp. 54–104. Elsevier, Amsterdam.
Johannesson K, André C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, **15**, 2013–2029.
Johannesson K, Johannesson B, Lundgren U (1995) Strong natural selection causes microscale allozyme variation in a marine snail. *Proceedings of the National Academy of Sciences, USA*, **92**, 2602–2606.
Johnson MS, Black R (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology*, **70**, 157–164.
Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research*, **11**, 247.
Larsson LC, Laikre L, André C, Dahlgren TG, Ryman N (2010) Temporally stable genetic structure of heavily exploited Atlantic herring (*Clupea harengus*) in Swedish waters. *Heredity*, **104**, 40–51.
Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–514.
Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 74–84.
Luttikhuizen PC, Pijnacker LP (2002) Mosaic haploid-diploid embryos and polyspermy in the tellinid bivalve *Macoma balthica*. *Genome*, **45**, 59–62.
Luttikhuizen PC, Drent J, Baker AJ (2003a) Disjunct distribution of highly diverged mitochondrial lineage clade and population subdivision in a marine bivalve with pelagic larval dispersal. *Molecular Ecology*, **12**, 2215–2229.
Luttikhuizen PC, Drent J, Delden Wvan, Piersma T (2003b) Spatially structured genetic variation in a broadcast-spawning bivalve: quantitative versus molecular traits. *Journal of Evolutionary Biology*, **16**, 260–272.
Luttikhuizen PC, Campos J, van Bleijswijk J, Peijnenburg KTCA, van der Veer HW (2008) Phylogeography of the common shrimp, *Crangon crangon* (L) across its distribution range. *Molecular Phylogenetics and Evolution*, **46**, 1015–1030.
Mäkinen HS, Cano M, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Molecular Ecology*, **17**, 3565–3582.
Mehner T, Pohlmann K, Elkin C, Monaghan MT, Nitz B, Freyhof J (2010) Genetic population structure of sympatric and allopatric populations of Baltic ciscoes (*Coregonus albula* complex, Teleostei, Coregonidae). *BMC Evolutionary Biology*, **10**, 14.
Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences, USA*, **107**, 9724–9729.
Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, **15**, 4229–4242.
Muths D, Jollivet D, Gentil F, Davoult D (2009) Large-scale genetic patchiness among NE Atlantic populations of the brittle star *Ophiothrix fragilis*. *Aquatic Biology*, **5**, 117–132.
Nikula R, Strelkov P, Väinölä R (2007) Diversity and trans-Arctic invasion history of mitochondrial lineages in the North Atlantic *Macoma balthica* complex (Bivalvia : Tellinidae). *Evolution*, **61**, 928–941.
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). *Molecular Ecology*, **17**, 1505–1522.
Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.
Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007) Divergence between the *Drosophila pseudoobscura* and *D-persimilis* genome sequences in relation to chromosomal inversions. *Genetics*, **177**, 1417–1428.
Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
Nunes VL, Beaumont MA, Butlin RK, Paulo OS (2011) Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient. *Molecular Ecology*, **20**, 193–205.
Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.
Peijnenburg KTCA, Fauvelot C, Breeuwer JAJ, Menken SBJ (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology*, **15**, 3319–3338.
Pereyra RT, Bergstrom L, Kautsky L, Johannesson K (2009) Rapid speciation in a newly opened postglacial marine environment, the Baltic Sea. *BMC Evolutionary Biology*, **9**, 9.
Perez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.
Place AR, Powers DA (1979) Genetic variation and relative catalytic efficiencies: lactate dehydrogenase B allozymes of *Fundulus heteroclitus*. *Proceedings of the National Academy of Sciences, USA*, **76**, 2354–2358.
Rieseberg LH, Buerkle CA (2002) Genetic mapping in hybrid zones. *The American Naturalist*, **159**, S36–S50.
Riginos C, Cunningham CW (2005) Local adaptation and species segregation in two mussel (*Mytilus edulis* x *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, **14**, 381–400.
Riginos C, Sukhdeo K, Cunningham CW (2002) Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution*, **19**, 347–351.

© 2012 Blackwell Publishing Ltd

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonas* sp Salmonidae) species pairs. *Molecular Biology and Evolution*, **24**, 1423–1438.

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M (2010) Genome-wide association studies in diverse populations. *Nature Reviews Genetics*, **11**, 356–366.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences, USA*, **106**, 9955–9962.

Seldin MF (2007) Admixture mapping as a tool in gene discovery. *Current Opinion in Genetics & Development*, **17**, 177–181.

Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Molecular Biology and Evolution*, **26**, 1341–1355.

Strelkov P, Nikula R, Väinölä R (2007) *Macoma balthica* in the White and Barents Seas: properties of a widespread marine hybrid swarm (Mollusca : Bivalvia). *Molecular Ecology*, **16**, 4110–4127.

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

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PloS Biology*, **3**, 1572–1578.

Väinölä R (2003) Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. *Marine Biology*, **143**, 935–946.

Väinölä R, Hvilsom MM (1991) Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). *Biological Journal of the Linnean Society*, **43**, 127–148.

Väinölä R, Strelkov P (2011) *Mytilus trossulus* in Northern Europe. *Marine Biology*, **158**, 817–833.

Väinölä R, Varvio S-L (1989) Biosystematics of *Macoma balthica* in northwestern Europe, In: *23rd European Marine Biology Symposium, Olsen and Olsen* (eds Ryland JS, Tyler PA), pp. 309–316.

Van Doorn GS, Luttikhuizen PC, Weissing FJ (2001) Sexual selection at the protein level drives the extraordinary divergence of sex-related genes during sympatric speciation. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **268**, 2155–2161.

Vermeij GJ (1991) Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology*, **17**, 281–307.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences, USA*, **106**, 9939–9946.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Vos P, Hogers R, Bleeker M *et al*. (1995) AFLP – a new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, **8**, 725–735.

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

Wolf JBW, Lindell J, Backstrom N (2010) Speciation genetics: current status and evolving approaches. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 1717–1733.

Yebra L, Bonnet D, Harris RP, Lindeque PK, Peijnenburg K (2011) Barriers in the pelagic: population structuring of *Calanus helgolandicus* and *C. euxinus* in European waters. *Marine Ecology Progress Series*, **428**, 135–149.

Data accessibility

AFLP data per location and per individual: DRYAD entry doi:10.5061/dryad.70np2513.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Fitted genomic cline curves for *Macoma balthica* Kattegat–Baltic hybrid zone.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Blackwell Publishing Ltd