

Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture analysis

EINAR E. NIELSEN,* MICHAEL M. HANSEN,* DANIEL E. RUZZANTE,*† DORTE MELDRUP* and PETER GRØNKJÆR‡

*Department of Inland Fisheries, Danish Institute for Fisheries Research, DK-8600 Silkeborg, Denmark, †Department of Marine Ecology, University of Aarhus, DK-8200 Aarhus, Denmark

Abstract

The study of hybrid zones is central to our understanding of the genetic basis of reproductive isolation and speciation, yet very little is known about the extent and significance of hybrid zones in marine fishes. We examined the population structure of cod in the transition area between the North Sea and the Baltic Sea employing nine microsatellite loci. Genetic differentiation between the North Sea sample and the rest increased along a transect to the Baltic proper, with a large increase in level of differentiation occurring in the Western Baltic area. Our objective was to determine whether this pattern was caused purely by varying degrees of mechanical mixing of North Sea and Baltic Sea cod or by interbreeding and formation of a hybrid swarm. Simulation studies revealed that traditional Hardy–Weinberg analysis did not have sufficient power for detection of a Wahlund effect. However, using a model-based clustering method for individual admixture analysis, we were able to demonstrate the existence of intermediate genotypes in all samples from the transition area. Accordingly, our data were explained best by a model of a hybrid swarm flanked by pure nonadmixed populations in the North Sea and the Baltic Sea proper. Significant correlation of gene identities across loci (gametic phase disequilibrium) was found only in a sample from the Western Baltic, suggesting this area as the centre of the apparent hybrid zone. A hybrid zone for cod in the ecotone between the high-saline North Sea and the low-saline Baltic Sea is discussed in relation to its possible origin and maintenance, and in relation to a classical study of haemoglobin variation in cod from the Baltic Sea/Danish Belt Sea, suggesting mixing of two divergent populations without interbreeding.

Keywords: Atlantic cod, Baltic Sea, hybrid-zone, individual admixture analysis, marine fish, microsatellite DNA

Received 26 September 2002; revision received 13 January 2003; accepted 24 January 2003

Introduction

Hybrid zones are defined as areas where genetically distinct groups of individuals interact and result in at least some offspring of mixed ancestry. Pure populations of the two genetically distinct groups are found outside the zone of interaction (Harrison 1990). Hybrid zones have been the focus of much attention during the last couple of decades

resulting in a number of extensive reviews of hybrid zone theory and examples of their occurrence (e.g. Endler 1977; Barton & Hewitt 1985; Harrison 1990; Harrison 1993). Basically, two scenarios for hybrid zone formation have been suggested. Primary hybrid zones are formed between populations diverged in sympatry or parapatry by differential environmental selection. Secondary hybrid zones are formed between formerly allopatric populations that come into secondary contact. Similarly, there are basically two types of models explaining how stable hybrid zones are maintained. Dispersal independent models include those where hybrids are thought to be more fit than both parental types within an ecotone, 'the

Correspondence: Einar E. Nielsen. Fax: + 45 89 213150, E-mail: een@dfu.min.dk

†Present address: Department of Biology, Dalhousie University, Halifax, NS, B3H 4J1, Canada

bounded hybrid superiority' model (Moore 1977). Dispersal/selection models concern situations where a stable cline is primarily maintained by selection against intermediate genotypes (Barton & Hewitt 1985). The latter type, termed frequently a 'tension zone', is often characterized by gametic phase disequilibrium due to dispersal of parental gene combinations into the centre of the zone (Barton & Hewitt 1985).

Hybrid zones in terrestrial environments have been the subject of many detailed studies (see, e.g. Harrison 1993 and references therein), whereas studies in marine organisms have been much more limited, both in numbers and detail (Gardner 1997). This also holds for marine fishes. Although several examples of identification of interspecific hybrids are available, such as hybrids between species of flatfish (Sick *et al.* 1963) and redfish (Roques *et al.* 2001), the field of intraspecific hybridization in marine fishes remains largely unexplored. This can, to a large extent, be attributed to the inherent difficulties of studying the structure of marine fish populations and their interactions using population genetic methods. In general, the level of neutral genetic differentiation among populations of 'classical' marine fishes (i.e. with wide distributions, large population sizes, high fecundity, pelagic eggs and larvae) has been found to be low compared to freshwater and anadromous fish (Gyllensten 1985; Ward *et al.* 1994; Waples 1998). Further, traditional genetic markers such as allozymes tend to lack the variability necessary to detect the expected small genetic differences. Microsatellites and other hyper-variable genetic markers have opened possibilities for genetic studies of marine fish with supposedly low levels of genetic differentiation (O'Connell & Wright 1997; Carvalho & Hauser 1998). Recently, a number of associated new statistical tools for individual-based analysis have been developed (Cornuet *et al.* 1999; Pritchard *et al.* 2000; see also Hansen *et al.* 2001), which allow the precise assignment of individuals to a set of baseline populations, as well as the estimation of the genetic contribution of baseline populations to the genotype of an individual, i.e. individual admixture proportions (Pritchard *et al.* 2000; see also Beaumont *et al.* 2001; Hansen 2002). Consequently, as the study of population structure in marine fishes employing more powerful molecular markers is still in its infancy, relatively little is known of population structure of marine fishes in general and of interactions between populations such as mixing and hybridization in particular.

A noteworthy exception to this general rule is Atlantic cod (*Gadus morhua*). Cod has been the subject of population genetic studies for several decades and genetic differentiation has been documented at a number of hierarchical levels, such as trans-Atlantic (Mork *et al.* 1985; Pogson *et al.* 1995) and more limited geographical regions such as the Northwest Atlantic (Ruzzante *et al.* 1998) and the Northeast Atlantic (Hutchinson *et al.* 2001; Nielsen *et al.* 2001a).

Further, evidence of interactions between populations in terms of 'mechanical' mixing of populations (Ruzzante *et al.* 2000) has also been described.

One of the earliest and probably best-known population genetic studies on cod is that of haemoglobin differences between cod in the Baltic and the Danish Belt Sea (Sick 1965). Cod is found throughout this region, which is characterized by a steep salinity gradient from near oceanic salinities in the North Sea to the brackish Baltic Sea. The observation of a heterozygote deficiency in the area around the island of Bornholm, where the two populations presumably meet, has been used as an illustrative example of the 'Wahlund principle'. The main conclusion by Sick (1965) was that there were two well-defined cod populations (Belt Sea and Baltic Sea) with a narrow zone of pure 'mechanical' mixing. However, these findings were challenged later by Moth-Poulsen (1982). Examining variation at three allozyme loci Moth-Poulsen (1982) advocated the presence of four distinct populations in the transition area, based on significant differentiation and no deficiency of heterozygotes within samples. Instead, the data indicated a gradual transition/cline between North Sea and Baltic Sea cod, i.e. a potential intraspecific hybrid zone.

Here we examine the genetic population structure of cod in the transition area between the North Sea and the Baltic Sea employing highly variable microsatellite DNA markers, with particular emphasis on estimating both population level and individual level admixture proportions. Throughout the study we use the term 'population level admixture proportions' to describe the relative genetic contribution of the parental populations (in this case the North Sea and the Baltic Sea proper populations) to each of the transition zone samples (Kattegat, Belt Sea, The Sound and Western Baltic), regardless of whether or not the samples consisted of 'pure' mechanically mixed individuals or admixed individuals. The term 'individual level admixture proportion' is used to denote how much of an individual's genome is derived from one or the other of the parental populations, i.e. the extent to which an individual is admixed. First, we estimate the levels of genetic divergence among samples collected along a transect from the North Sea to the Baltic Sea proper. Second, we estimate the most probable number of populations present in the area using a newly developed clustering method and estimate *population level* admixture proportions for each sample. Third, we investigate whether the observed pattern of gradual genetic transition is caused most probably by mechanical mixing of cod from two divergent populations or by interbreeding and formation of a hybrid swarm of intermediate genotypes. For this purpose we use the same clustering method used to estimate population level admixture, but here we employ it to estimate individual level admixture. Fourth, we also evaluate the power of traditional analysis (HW) to detect heterozygote deficiencies



Fig. 1 Map of sea areas around Denmark, showing sampling localities of cod.

(Wahlund effect) that can be created by mechanical mixture of noninterbreeding populations. This is conducted by simulating 'mechanically' mixed samples of pure North Sea and Baltic Sea cod. Fifth and finally, we evaluate whether the hybrid zone is consistent with hybrid inferiority or superiority (a tension zone) by testing for gametic phase disequilibrium but at the same time taking the statistical power of this approach into account.

Materials and methods

Areas studied

The Kattegat, the Danish Belt Sea, The Sound and the Western Baltic form the transition area between the high-

saline North Sea (30–34‰) and the low-saline Baltic proper (<8‰) (Fig. 1). Cod are found throughout this area. Spawning is somewhat diffuse in the transition area, both with respect to time and space (Peter Munk, Danish Institute for Fisheries Research pers. com.). In contrast, successful spawning in the Baltic proper is confined to the deep basins (Vallin *et al.* 1999). Samples were collected from 1996 to 1999 from the transition area as well as from the North Sea and the Baltic proper (see Table 1 and Fig. 1). Samples (gill tissue stored in ethanol) were collected close to the estimated spawning time for each sea area and consisted of adult cod primarily in maturing or spawning stages (Table 1).

Microsatellite analysis

DNA was extracted from gill tissue using a proteinase K/chex method (Estoup *et al.* 1996). We employed the following nine highly variable microsatellite loci: Gmo 1, Gmo 2, Gmo 132 (Brooker *et al.* 1994), Gmo 120 (Ruzzante *et al.* 1996a), Gmo 141 (Ruzzante *et al.* 1996b), Gmo 8, Gmo 19, Gmo 34 and Gmo 37 (Miller *et al.* 2000). Standard polymerase chain reaction (PCR) reagents were used and the microsatellites were run subsequently on an ALFexpress automated sequencer according to the manufacturer's recommendations.

Statistical analysis

We used the program FSTAT (Goudet 1995) to test for deviations from HW equilibrium, calculate allelic richness

Table 1 Summary statistics for cod (*Gadus morhua* L.) samples collected showing geographical sampling locality and position (mean), year and month of sampling, proportion of maturing (stages 3–5) and mature individuals (stages 6–7) and number (*n*) of samples collected

Geographic locality	Sample no.	Position (mean)	Year	Month	Proportion of maturing and mature (%)	No. of individuals
North Sea	1A	55.17° N–03.39° E	1996	Feb./March	18, 68	82
	1B	57.10° N–08.20° E	1999	Feb./March	32, 46	76
Kattegat	2A	57.15° N–11.35° E	1996	Feb./March	85, 10	50
	2B	57.05° N–11.05° E	1997	Feb./March	75, 11	40
Belt Sea	3A	55.11° N–10.28° E	1996	Feb./March	38, 50	88
	3B	54.27° N–11.23° E	1997	Feb./March	52, 43	40
The Sound	4	55.57° N–12.39° E	1997	Feb./March	40, 30	50
Western Baltic	5	54.53° N–13.33° E	1996	Feb./March	49, 2	59
Eastern Baltic, Bornholm basin	6A	55.19° N–15.54° E	1996	April	63, 24	74
	6B	54.51° N–15.28° E	1997	April	65, 35	80
Eastern Baltic, Gdansk basin	7A	54.57° N–19.04° E	1997	May/June	12, 88	46
	7B	54.56° N–19.11° E	1998	May/June	37, 63	102
Eastern Baltic, Gotland basin	8A	56.14° N–19.09° E	1997	May/June	9, 91	34
	8B	56.56° N–19.50° E	1998	May/June	33, 67	49

(El Mousadik & Petit 1996), estimate pairwise F_{ST} s between all samples (following Weir & Cockerham 1984) and for testing their significance. Initially all tests were performed on samples from individual years; however, estimates of genetic differentiation between years within the transition area were extremely small (F_{ST} ranged from 0 to 0.0005) and insignificant (P -values between 0.44 and 0.65). Subsequently, samples from different years from the same or nearby locations were pooled to allow higher statistical power for the following analysis. Population level admixture proportions and their standard deviations were estimated using the least-squares method by Robert & Hiorns (1965) as implemented in the program ADMIX 1_0 (Bertorelle & Excoffier 1998). We did not, however, use the coalescence-based approach included in ADMIX, because it has been shown to have much larger variance than allele frequency-based methods when differentiation is low (Bertorelle & Excoffier 1998) such as observed in marine fishes. The program STRUCTURE (Pritchard *et al.* 2000) was used, first to estimate the most likely number of populations represented by the samples, and next to estimate individual admixture proportions, i.e. the estimated proportion of an individual's genotype originating from one or the other of the parental populations. STRUCTURE is a model-based Bayesian, Markov chain Monte Carlo approach that clusters individuals to minimize Hardy-Weinberg disequilibrium and gametic phase disequilibrium between loci within groups. The number of populations represented in our samples was estimated by pooling all samples and calculating the probability of the data, assuming that they originated from one to four populations in the study area, as described in Pritchard *et al.* (2000). Individual admixture proportions (q) and their 90% posterior probability intervals were calculated by assuming an admixture model (i.e. allowing the genetic composition of individuals to be a mixture from different populations). The model was forced to consider the individuals from the baseline samples (North Sea and Baltic proper) as 'pure', i.e. nonadmixed individuals. The samples 1A and 1B ($n = 158$) were chosen to represent the 'pure' North Sea population while the samples 6A and 6B from the Bornholm Basin were chosen to represent the 'pure' Baltic population. This two-population model was chosen based on the most likely number of populations as estimated by STRUCTURE (see above) and the apparent low degree of population structure within the North Sea (Hutchinson *et al.* 2001) and the Baltic Sea (this study). The rationale for only using Bornholm Basin samples as a baseline was primarily that the majority of successful spawning in the Baltic takes place there at present, but also to avoid introducing potential bias by having different baseline sample sizes. As North Sea and Baltic Sea cod were not strongly differentiated genetically, allele frequencies in populations were assumed to be correlated. The expected distribu-

tions of individual admixture proportions under a pure mechanical mixing model were generated by creating samples of equal size to the samples from the transition area, consisting of simulated 'pure' North Sea and Baltic Sea genotypes (genotypes generated by randomly drawing alleles from the allele frequency distribution of each 'pure' sample). Genotypes from each 'pure' population were included in proportions equal to the estimated population admixture proportions based on the total sample. Under a hybrid swarm model, the expected distributions of individual admixture proportions were generated by creating samples of simulated genotypes produced by drawing alleles randomly from the observed allele frequency distribution for each sample in the transition area. This procedure is equal to a situation of random mating within the samples (hybridization). We used a Kolmogorov-Smirnov two-sample test to evaluate differences between the true observed distribution of individual admixture proportions and the distributions of simulated individuals under the two models. We also investigated whether there were differences in the distribution of individual admixture proportions between the observed and the simulated pure individuals from the baseline samples, i.e. a test for simulation bias. This was performed by comparing the distribution of individual admixture proportions of 40 randomly chosen individuals (20 chosen and removed from each baseline sample) with that of 40 simulated pure individuals. The simulated samples under the mechanical mixture model were further used to investigate our statistical power to detect Hardy-Weinberg and gametic phase disequilibrium. The simulated individuals were created by the program HYBRIDLAB (<http://www.evalife.dk/applications/hybridlab/> Nielsen *et al.* 2001b). Correlations of gene identities across loci (gametic phase disequilibrium) within samples were estimated using the program Estim1.1 (Vitalis & Couvet 2001).

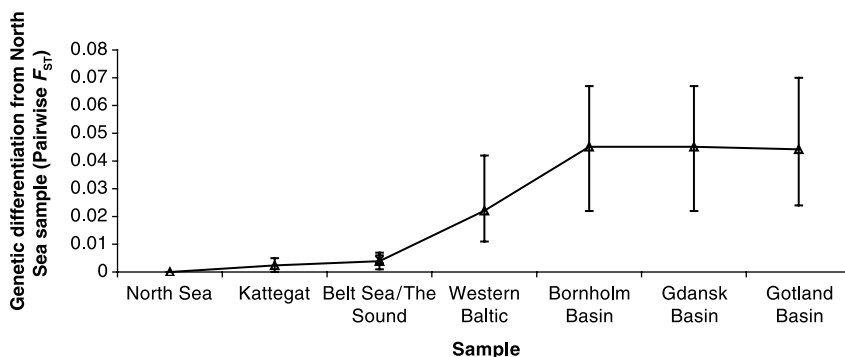
Results

Genetic differentiation among samples

The majority of tests for pairwise genetic differentiation were significant (Table 2). There was no significant differentiation among samples from the Kattegat, the Belt Sea and the Sound, between the North Sea and the Sound and among samples from the Baltic proper (Bornholm basin, Gotland basin and Gdansk basin). Genetic differentiation between the North Sea sample and the other samples increased gradually along a transect from the North Sea to the Baltic proper (Fig. 2). However, genetic differentiation between the North Sea sample and the Kattegat, Belt Sea and Sound samples was relatively small. The largest leap in level of differentiation occurred between the Belt Sea and Western Baltic samples. Genetic

Table 2 Estimates of pairwise genetic differentiation (F_{ST}) among cod samples, 95% confidence intervals (below diagonal) and P -values (above diagonal). Significance levels were adjusted according to the sequential Bonferroni method (Rice 1989)

Sampling locality	North Sea	Kattegat	Belt Sea	The Sound	Western Baltic	Bornholm basin	Gdansk basin	Gotland basin
North Sea	—	0.00036*	< 0.00003***	0.00414 NS	< 0.00003***	< 0.00003***	< 0.00003***	< 0.00003***
Kattegat	0.0024	—	0.45658 NS	0.40083 NS	< 0.00003***	< 0.00003***	< 0.00003***	< 0.00003***
	0.000–0.005							
Belt Sea	0.0039	0.0000	—	0.41208 NS	< 0.00003***	< 0.00003***	< 0.00003***	< 0.00003***
	0.001–0.007	0.000–0.001						
The Sound	0.0044	0.0000	0.0000	—	0.00078*	< 0.00003***	< 0.00003***	< 0.00003***
	0.003–0.006	0.000–0.001	0.000–0.001					
Western Baltic	0.0221	0.0131	0.0097	0.0071	—	0.0134 NS	0.47228 NS	0.67544 NS
	0.011–0.032	0.005–0.021	0.004–0.015	0.002–0.011				
Bornholm basin	0.0451	0.0353	0.0275	0.0251	0.0051	—	0.16003 NS	0.05608 NS
	0.022–0.077	0.017–0.062	0.013–0.048	0.011–0.046	0.000–0.015			
Gdansk basin	0.0451	0.0332	0.0275	0.0234	0.0040	0.0011	—	0.78544 NS
	0.025–0.073	0.018–0.056	0.016–0.044	0.013–0.040	0.000–0.012	0.000–0.003		
Gotland basin	0.0442	0.0317	0.0255	0.0214	0.0047	0.0018	0.0002	—
	0.024–0.071	0.016–0.054	0.014–0.041	0.011–0.039	0.000–0.012	0.000–0.004	0.000–0.002	

**Fig. 2** Genetic differentiation (pairwise F_{ST} s) between the North Sea sample and all other samples. Samples are included following a geographical transect going from the North Sea to the Gotland Basin in the Baltic proper.

variability in terms of allelic richness decreased going from the North Sea to the Baltic (see Appendix) with the average expected number of alleles in a sample of 50 individuals decreasing from 19 to 15, respectively.

Estimate of the number of populations and population level admixture

The model-based clustering method revealed that the most probable number of populations present in our total sample was two (posterior probability almost 100%). Estimated population-level admixture proportions indicated population admixture in all samples from the transition area (Table 3).

Only one of the 72 individual tests for deviations from HW proportions (see Appendix) was significant after Bonferroni correction for multiple testing (Rice 1989). However, the power to detect heterozygote deficiencies caused by population mixture was low, as illustrated by the lack of significant outcomes for any locus in the simulated

Table 3 Estimates of population level admixture in the transition area using the program ADMIX 1–0 (Bertorelle & Excoffier 1998). Given is the estimated contribution from the North Sea cod population and the estimated number of North Sea and Baltic cod individuals in the sample assuming a pure mechanical mixing model

Geographic locality	North Sea cod admixture proportion (SD)	Expected number of North Sea/Baltic cod, assuming a pure mechanical mixing model
Kattegat	0.85 (0.05)	76/14
Belt Sea	0.72 (0.04)	92/36
The Sound	0.67 (0.05)	33/17
Western Baltic	0.26 (0.06)	15/44

samples (see Appendix) consisting of varying mixtures of 'pure' North Sea and Baltic individuals (see input proportions in Table 3). Individual admixture analysis (Fig. 3) indicated that a large proportion of individuals in the

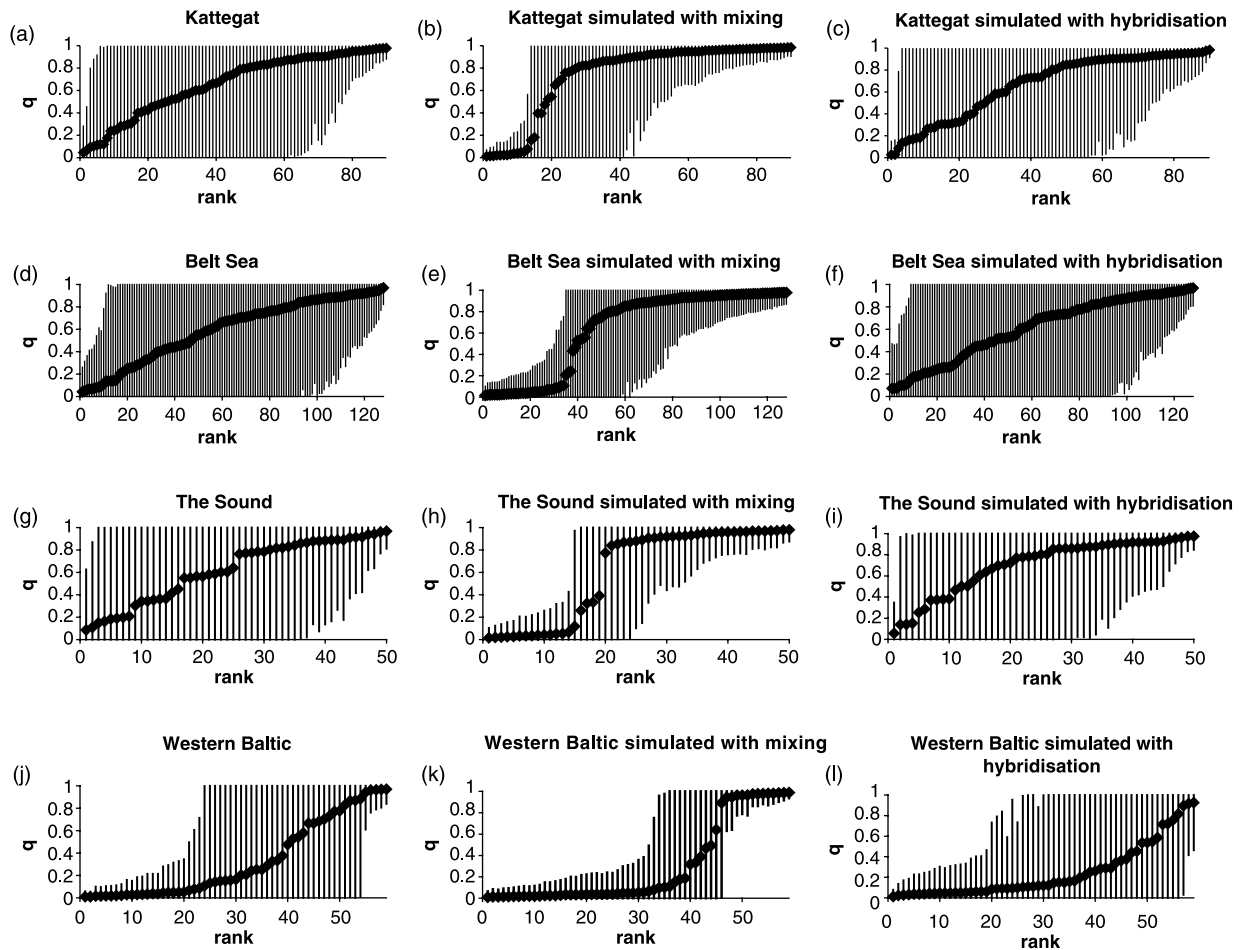


Fig. 3 Distribution of q (admixture proportion) values and 90% posterior probability intervals among individuals. The values have been ranked starting from 'pure' Baltic cod (a value of zero) to 'pure' North Sea cod (a value of one). (a, b, c) True and simulated samples from Kattegat; (d, e, f) true and simulated samples from the Belt Sea; (g, h, i) true and simulated samples from The Sound; (j, k, l) true and simulated samples from the Western Baltic. See text for details.

transition area samples consisted of admixed individuals (a, d, g, j). The distribution of individual admixture proportions was not significantly different from a model of a hybrid swarm (random mating c, f, i, l) (P -values 0.759, 0.964, 0.270 and 0.499) for any of the samples, but inconsistent with a model of pure mechanical mixing (b, e, h) (P -values < 0.000, < 0.000, 0.006) except for the Western Baltic (k) (P -value 0.174). For all samples the expected distributions under the two models were significantly different (P -values 0.003, < 0.000, 0.040, 0.008). No differences in distribution of individual admixture proportions between real and simulated pure individuals were detected (Fig. 4, P -value 0.913).

Estimates of correlation of gene identities at pairs of loci (gametic phase disequilibrium) revealed only a significant outcome for the Western Baltic sample (Table 4). Similarly, gametic phase disequilibrium was detected only for the simulated mixed sample from the Western Baltic (see above), illustrating low detection power.

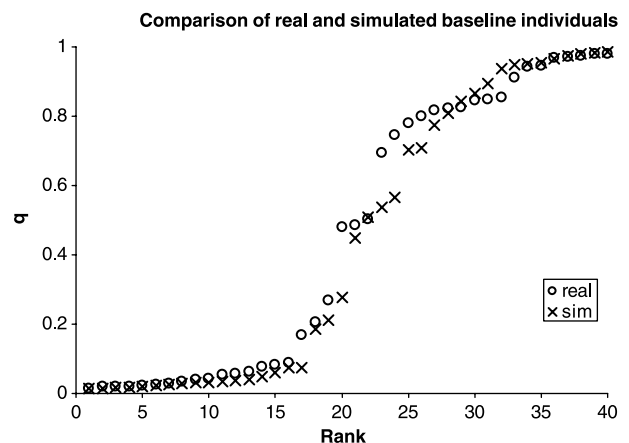


Fig. 4 Distribution of q -values for 40 unmarked baseline individuals (20 from the North Sea and 20 from the Baltic Sea). The values have been ranked starting from 'pure' Baltic cod (a value of zero) to 'pure' North Sea cod (a value of one). See text for further explanation.

	Correlation (S)	95% CI
Geographical samples		
North Sea	-0.000980	-0.000518-0.000362
Kattegat	0.000255	-0.000569-0.001296
Belt Sea	-0.000161	-0.000629-0.000376
The Sound	0.000212	-0.001485-0.001375
Western Baltic		
Eastern Baltic, Bornholm basin	0.001805	0.000168-0.003616
Eastern Baltic, Gdansk basin	0.000651	-0.000652-0.002286
Eastern Baltic, Gdansk basin	-0.000336	-0.001942-0.000869
Eastern Baltic, Gotland basin	-0.000443	-0.001942-0.001093
Simulated samples		
Kattegat	-0.000998	-0.002969-0.000621
Belt Sea	0.000185	-0.000768-0.000802
The Sound	0.000275	-0.001227-0.001543
Western Baltic	0.002824	0.000711-0.006934

Table 4 Correlations of gene identities across loci ('gametic phase disequilibrium') and associated 95% confidence intervals, estimated using the program Estim1.1 (Vitalis & Couvet 2001). The simulated samples consist of varying 'mechanical' mixtures of simulated pure North Sea and Baltic Sea multilocus genotypes equivalent to the estimated population level admixture proportions (see Table 3)

Discussion

Evidence of a hybrid zone

The results suggest a hybrid zone for Atlantic cod in the transition area between the North Sea and the Baltic proper. Intermediate allele frequencies are probably not created by mechanical mixing of fish from the two 'pure' populations, but more probably by the occurrence of admixed (hybrid) individuals. Furthermore, it is not likely that the observed pattern is merely a result of isolation by distance in a continuous population caused by restricted movement of individuals in relation to population size, as relatively large genetic changes are found over short distances compared to differences observed within the North Sea (Hutchinson *et al.* 2001) and within the Baltic proper (this study). Spawning areas are, except for the Baltic proper, diffuse and continuous making the occurrence of several discrete populations within the area unlikely. This is in line with our findings of the highest probability of the samples representing only two populations. In summation, the pattern is consistent with the definition by Harrison (1990), i.e. individuals of mixed ancestry are found in the zone of interaction and pure forms are found outside. Potential biases, such as the pooling of samples and the fact that not all individuals in our samples were mature, are not likely to have flawed these conclusions, because they do not point to an alternative explanation for our main conclusion, i.e. the occurrence of intermediate genotypes in the transition area. On the contrary, if some of the immature individuals found in our samples were individuals from either the North Sea or the Baltic proper on feeding migration before returning to their native sea area for spawning, more 'pure' parental genotypes would be observed in the distribution of individual admixture proportions in the transition area increasing the probability

of detecting departures from the curve expected under a hybrid swarm. The potential presence of migrating individuals could perhaps explain the lack of significant differences between the expected distribution of individual admixture proportions from a mechanical mixing model and the observed distribution for the Western Baltic. Because only about half the individuals of this sample are mature, the inclusion of a number of pure Baltic Sea cod cannot be ruled out.

We also investigated whether the use of highly variable loci could have biased our results by making 'pure' North Sea or Baltic Sea individuals in the transition samples appear admixed due to the presence of private alleles, i.e. alleles not found in the baseline samples due to the high allelic richness. Such alleles would not be found among the simulated pure individuals (as these would contain only alleles present in the baseline populations), making them appear less admixed. The 'acid test' for this hypothesis was then to see if unlabelled (not included in the baseline) pure individuals from the baseline populations had individual admixture proportions similar to simulated pure individuals or, alternatively, appeared admixed similar to individuals in the intermediate samples. If the latter was the case, then the allelic richness was seriously biasing the results. In other words, was the expected S-shaped distribution under a pure mechanical mixing model still apparent when using real instead of simulated fish? There was no significant difference between simulated and pure individuals and the pure individuals did not appear more admixed (see Fig. 4). Therefore, we do not expect any large bias from using highly variable loci. Similarly, we do not expect that the difference in allelic richness between baselines have biased our results, i.e. that individuals from the most variable population tend to be misclassified more often. When the data are actually informative about ancestry as in our case (F_{ST} between baseline samples 0.045), the

number of misclassified individuals (e.g. individuals that appear admixed) is expected to be low. This is also supported by the data in Fig. 4, where the simulated and real North Sea individuals appeared only marginally more admixed than pure and simulated Baltic Sea individuals.

Our results contrast the findings of Sick (1965). We found what appears to be a hybrid-swarm where he found two well-defined populations mixing around the island of Bornholm. This discrepancy could have several causes. First, Sick employed a genetic marker (haemoglobin) with low variability and probably subject to temperature selection (Karpov & Novikov 1980). Differential selection could provide an explanation for the higher level of differentiation and a sharper cline for the HB 1 system observed by Sick (1965) compared to (presumed) neutral microsatellites in this study. This scenario does not explain directly the Wahlund effect observed by Sick. His observations are congruent with a number of tagging studies (Otterlind 1985), demonstrating mixing in certain areas at certain times in the Belt Sea/Baltic region. Because our study has focused mainly on mature fish at spawning time, we have probably (and intentionally) avoided picking up a high number of these migrants. Finally, the possibility exists that the zone is not temporally stable, i.e. the level of mixing or the location of the zone is variable. The salinity of the Baltic is highly variable determined by periodic large inflows of high-saline water from the Kattegat. Population size of Baltic cod is very dependent on these inflows, which determine the size of the area suitable for reproduction. Similarly, it could change the distribution of populations and the hybrid swarm.

Type of hybrid zone

To infer 'process from pattern' (Endler 1977) and determine whether this apparent hybrid zone is of primary or secondary origin based on current structure is fraught with difficulties. Just as allopatric speciation is often regarded as most common, secondary hybrid zones are thought to occur most frequently (Barton & Hewitt 1985). However, both scenarios are possible and linked to the origin of Baltic cod. The geological history of the Baltic Sea is relatively well known. After the last glaciation, the Baltic Sea was a freshwater lake (the 'Ancylus Lake') until 7100 BP when a broad passage was created to the Kattegat, resulting in a relatively high salinity. This 'Littorina Sea' was invaded by marine fauna probably also including cod. The Littorina Period lasted for 3000 years. Since then, the salinity of the Baltic has been decreasing due to a decline in sea-level narrowing the passage to the Kattegat combined with a large freshwater discharge (Maagaard & Rheinheimer 1974). Therefore, either an already distinct 'Baltic cod' population living outside the Baltic (as suggested by Sick 1965) colonized the Baltic Sea giving rise to a secondary hybrid

zone at present, or colonization was by North Sea cod, with subsequent genetic differentiation caused by differential selection with respect to salinity or other environmental variables in sympatry or parapatry, resulting in a primary hybrid zone. Local adaptation to low salinity has been demonstrated for Baltic cod (see Vallin *et al.* 1999 and references therein), however, whether the Baltic cod adapted *in situ* or *ex situ* remains unresolved. Examples of adaptations to a low-saline environment from other marine organisms inhabiting the Baltic Sea provide no clear answers, because examples of both types (*in situ*: Gammarus & Kolding 1985; *ex situ*: Mytilus *et al.* 1991) are found. Sick, on the other hand, suggested that the Baltic cod originated from Northeastern arctic cod from the Barents Sea. This hypothesis is not supported by recent studies employing microsatellites (Nielsen *et al.* 2001a) that demonstrated a closer relationship between Baltic and North Sea cod than between Baltic cod and cod from the Barents Sea.

Similarly, it is difficult to evaluate the forces responsible for maintenance of the hybrid zone, i.e. whether this apparent hybrid zone is maintained by hybrid inferiority or superiority. This is particularly the case in intraspecific studies such as this where genetic differentiation between pure forms is small, resulting in low power for detection of gametic phase disequilibrium, even when it is expected to be largest, i.e. by mixing of pure nonadmixed individuals as illustrated by the simulation studies. However, with the low statistical power in mind, we still find some evidence of gametic phase disequilibrium in the central part of the transition area (Western Baltic). Therefore, the most probable explanation is that the zone is maintained by hybrid inferiority, i.e. a selection migration balance (tension zone, Barton & Hewitt 1985). However, this could also be caused by the inclusion of a number of pure Baltic individuals in the sample. In order to shed more light on how this hybrid zone is maintained, studies of fitness of parentals, natural and artificial hybrids, both under laboratory and natural conditions must be conducted. The occurrence of a hybrid zone for cod in the Belt Sea/Baltic Sea is not unique, as hybrid zones have been described for this region also for mussels and crustaceans (see Gardner 1997 for examples). Similarly, recent microsatellite studies of another classical marine fish, the turbot (*Scophthalmus maximus* L.) from the same region indicates a similar population structure (Nielsen *et al.* in prep.), with well-defined pure populations outside the zone of interaction and a fast gradual transition in the Belt Sea/Baltic Sea. Hybrid zones between populations might also be a common phenomenon for marine fishes in other areas. However, for many species further advances are impeded by poor knowledge of population structure in general. Further, as illustrated by this study, careful evaluation of statistical power has to be performed before discarding population mixture as an

explanation for significant genetic differentiation. Such problems have to be overcome before detailed studies of interactions between populations can be conducted and hybrid zones in marine fishes receive due attention. This will substantially improve our understanding of the evolutionary processes and ultimately speciation in the sea.

Acknowledgements

We thank Cathrin Schmidt for her assistance in the laboratory, Stig Pedersen for help with the illustrations, Anders Koed for assistance with the statistical analysis, Jonathan Pritchard and Dorte Bekkevold for inspiring comments and the Danish Ministry of Food, Agriculture and Fisheries for financial support.

References

- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.
- Beaumont M, Barratt EM, Gottelli D *et al.* (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology*, **10**, 319–336.
- Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, **15**, 1298–1311.
- Brooker AL, Cook D, Bentzen P, Wright JM, Doyle RW (1994) Organization of microsatellites differs between mammals and cold-water teleost fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1959–1966.
- Carvalho GR, Hauser L (1998) Advances in the molecular analysis of fish population structure. *Italian Journal of Zoology*, **65**, 21–33.
- Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.
- Endler JA (1977) *Natural Selection in the Wild*. Princeton University Press, Princeton.
- Estoup A, Largiadere CR, Perrot E, Chourrou D (1996) Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Molecular Marine Biology and Biotechnology*, **5**, 295–298.
- Gardner JPA (1997) Hybridization in the sea. In: *Advances in Marine Biology*, vol. 31 (eds Blaxter JHS, Southward AJ), pp. 2–65. Academic Press, New York.
- Goudet J (1995) FSTAT (vers. 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Gyllenstein U (1985) The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous and freshwater species. *Journal of Fish Biology*, **26**, 691–699.
- Hansen MM (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, **11**, 1003–1015.
- Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers: Methods and applications. *Fish and Fisheries*, **2**, 93–112.
- Harrison RG (1990) Hybrid zones: windows on the evolutionary process. *Oxford Surveys in Evolutionary Biology*, **7**, 69–128.
- Harrison RG (1993) *Hybrid Zones and the Evolutionary Process*. Oxford University Press, Oxford.
- Hutchinson WF, Carvalho GR, Rogers SI (2001) Marked genetic structuring in localised spawning populations of cod (*Gadus morhua*) within the North Sea and adjoining waters as revealed by microsatellites. *Marine Ecology Progress Series*, **223**, 251–260.
- Karpov AK, Novikov GG (1980) Hemoglobin alloforms in cod, *Gadus morhua* (Gadiformes, Gadidae), their functional characteristics and occurrence in populations. *Journal of Ichthyology*, **20**, 45–49.
- Kolding S (1985) Genetic adaptation to local habitats and speciation processes within the genus *Gammarus* (Amphipoda: Crustacea). *Marine Biology*, **89**, 249–255.
- Maagaard L, Rheinheimer G (1974) *Meereskunde der Ostsee*. Springer Verlag, Berlin.
- Miller KM, Le KD, Beacham TD (2000) Development of tri- and tetranucleotide repeat microsatellite loci in Atlantic cod (*Gadus morhua*). *Molecular Ecology*, **9**, 238–239.
- Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology*, **52**, 23–277.
- Mork J, Ryman N, Ståhl G, Utter F, Sundnes G (1985) Genetic variation in Atlantic cod (*Gadus morhua*) throughout its range. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1580–1587.
- Moth-Poulsen T (1982) Genetic variation of cod from the Danish Sound: Interrelations of stocks from adjacent waters. *ICES CM 1982/G: 46*.
- Nielsen EE, Hansen MM, Bach L (2001b) Looking for a needle in a haystack. Discovery of indigenous salmon in heavily stocked populations. *Conservation Genetics*, **2**, 219–232.
- Nielsen EE, Hansen MM, Schmidt C, Meldrup D, Grønkjær P (2001a) Determining the population of origin of individual cod in the Northeast Atlantic. *Nature*, **413**, 272.
- O'Connell M, Wright JM (1997) Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries*, **7**, 331–363.
- Otterlind G (1985) Cod migration and transplantation experiments in the Baltic. *Journal of Applied Ichthyology*, **1**, 3–16.
- Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**, 375–385.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.
- Roberts DF, Hiorns RW (1965) Methods of analysis of the genetic composition of a hybrid population. *Human Biology*, **37**, 38–43.
- Roques S, Sevigny J, Bernatchez L (2001) Evidence for broadscale introgressive hybridization between two redfish (genus *Sebastes*) in the North-west Atlantic: a rare marine example. *Molecular Ecology*, **10**, 149–165.
- Ruzzante DE, Taggart CT, Cook C (1996b) Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 2695–2705.
- Ruzzante DE, Taggart CT, Cook D (1998) A nuclear DNA basis for shelf- and bank-scale population structure in Northwest

- Atlantic cod (*Gadus morhua*): Labrador to Georges Bank. *Molecular Ecology*, **7**, 1663–1680.
- Ruzzante DE, Taggart CT, Cook C, Goddard S (1996a) Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 634–645.
- Ruzzante DE, Taggart CT, Lang S, Cook D (2000) Mixed-stock analysis of Atlantic cod near the Gulf of St. Lawrence based on microsatellite DNA. *Ecological Applications*, **10**, 1090–1109.
- Sick K (1965) Haemoglobin polymorphism of cod in the Baltic and the Danish Belt Sea. *Hereditas*, **54**, 49–73.
- Sick K, Frydenberg O, Nielsen JT (1963) Haemoglobin patterns of plaice, flounder and their natural and artificial hybrids. *Nature, London*, **198**, 411–412.
- 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.
- Vallin L, Nissling A, Westin L (1999) Potential factors influencing reproductive success of Baltic cod, *Gadus morhua*: a review. *Ambio*, **28**, 92–99.
- Vitalis R, Couvet D (2001) ESTIM 1.0: a computer program to infer population parameters from one and two-locus gene identity probabilities. *Molecular Ecology Notes*, **1**, 354–356.
- Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 439–450.
- Ward RD, Woodward M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater and anadromous fishes. *Journal of Fish Biology*, **44**, 213–232.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

This study represents one of the major common interests of the authors, i.e. population structure and evolution of marine fishes. The work presented here is part of a series of ongoing studies at DIFRES focusing on the genetic population structure and evolution of marine fishes in the Northeast Atlantic in general and in particular on species inhabiting the interface between the North Sea and the Baltic Sea. More information on the DIFRES population genetics group can be found at: www.dfu.min.dk/jffi. Daniel Ruzzante has since moved to the Department of Biology at Dalhousie University in Canada (www.dal.ca/~biology2), where he continues to focus his research on the population genetics of marine, anadromous and freshwater fishes. Peter Gronkjaer is currently in the Department of Marine Ecology at the University of Aarhus (www.biology.au.dk/marine.eco/), where he focuses his research on marine and freshwater fish recruitment.

Appendix

Summary of basic genetic data for cod samples from eight sampling locations and four simulated samples (see text for details). For each of the nine microsatellite loci analysed, the number of individuals scored (n), number of alleles (A), expected (H_E) and observed (H_O) heterozygosity and tests for deviations from Hardy–Weinberg (HW) proportions (P -value and level of significance), allelic richness ($r(50)$) and number of ‘private’ alleles for the transition zone samples (alleles not present in the baseline samples) are given

Sample	Locus									
	Variable	Gmo 1	Gmo 2	Gmo 120	Gmo 132	Gmo 141	Gmo 8	Gmo 19	Gmo 34	Gmo 37
1 North Sea	n	158	158	157	158	158	158	158	158	158
	A	9	16	41	30	50	39	26	8	16
	H_E	0.179	0.857	0.957	0.919	0.972	0.914	0.898	0.639	0.846
	H_O	0.190	0.816	0.936	0.937	0.930	0.854	0.816	0.658	0.757
	HW	1.0000	0.0883	0.1222	0.8352	0.0056	0.0068	0.0012	0.7969	0.0025
	$r(50)$	5.43	11.57	30.15	22.04	37.32	24.09	20.37	7.82	11.78
2 Kattegat	n	90	90	89	90	90	90	90	90	89
	A	9	14	30	28	45	29	24	7	13
	H_E	0.189	0.835	0.948	0.913	0.971	0.897	0.917	0.648	0.838
	H_O	0.189	0.744	0.876	0.956	0.944	0.844	0.889	0.633	0.843
	HW	0.6315	0.0142	0.0068	0.9667	0.1198	0.0605	0.2105	0.3975	0.6062
	$r(50)$	6.89	12.17	25.91	22.90	38.02	23.17	21.25	6.79	11.23
3 Belt Sea	n	128	128	128	128	128	128	128	128	128
	A	6	16	34	29	43	27	23	8	14
	H_E	0.164	0.833	0.953	0.914	0.967	0.863	0.925	0.615	0.841
	H_O	0.164	0.695	0.953	0.875	0.945	0.859	0.875	0.602	0.781
	HW	0.6290	0.0006*	0.5636	0.0840	0.1302	0.5173	0.0340	0.3444	0.0383
	$r(50)$	5.34	11.60	28.19	21.33	34.63	19.44	20.31	6.79	11.48
4 The Sound	n	50	50	50	50	50	50	50	50	50
	A	5	10	28	21	36	16	23	7	11
	H_E	0.206	0.844	0.938	0.906	0.970	0.855	0.935	0.568	0.851
	H_O	0.180	0.800	0.940	0.860	0.940	0.840	0.900	0.540	0.840
	HW	0.1710	0.2370	0.6136	0.1741	0.1957	0.4395	0.1963	0.3327	0.4815
	$r(50)$	5	10	28	21	36	16	23	7	11
5 Western Baltic	n	59	59	59	59	59	59	59	59	59
	A	7	12	26	22	30	19	20	7	10
	H_E	0.177	0.836	0.945	0.875	0.909	0.753	0.905	0.580	0.815
	H_O	0.153	0.763	1.000	0.831	0.831	0.780	0.763	0.576	0.729
	HW	0.0957	0.0660	1.000	0.1778	0.0247	0.7981	0.0012	0.5395	0.0617
	$r(50)$	6.51	11.54	24.58	20.77	28.59	17.62	19.30	6.69	9.80
6 Eastern Baltic, Bornholm basin	n	154	154	152	153	153	154	154	154	154
	A	4	12	32	21	39	25	22	5	13
	H_E	0.057	0.833	0.948	0.861	0.894	0.529	0.895	0.420	0.781
	H_O	0.058	0.779	0.914	0.876	0.876	0.532	0.890	0.435	0.753
	HW	1.000	0.0496	0.0475	0.7284	0.2290	0.6253	0.4580	0.7944	0.2222
	$r(50)$	2.95	10.04	25.84	16.27	29.37	16.05	17.83	4.54	9.213
7 Eastern Baltic, Gdansk basin	n	148	148	148	148	148	148	148	148	148
	A	4	13	34	23	42	21	21	5	12
	H_E	0.092	0.821	0.939	0.584	0.896	0.565	0.896	0.394	0.804
	H_O	0.074	0.824	0.885	0.851	0.892	0.541	0.865	0.399	0.736
	HW	0.0191	0.5852	0.0068	0.5062	0.4846	0.1722	0.1253	0.6142	0.0247
	$r(50)$	3.25	10.02	24.17	15.98	30.25	14.94	16.77	4.56	9.63

Appendix *Continued*

Sample	Locus									
	Variable	Gmo 1	Gmo 2	Gmo 120	Gmo 132	Gmo 141	Gmo 8	Gmo 19	Gmo 34	Gmo 37
8 Eastern Baltic, Gotland basin	<i>n</i>	83	83	83	83	83	83	83	83	83
	A	4	10	32	18	34	25	19	5	9
	H_E	0.094	0.789	0.940	0.823	0.902	0.556	0.892	0.462	0.778
	H_O	0.096	0.771	0.904	0.843	0.904	0.542	0.952	0.410	0.831
	HW	1.000	0.4019	0.1278	0.7346	0.6148	0.3698	0.9883	0.0870	0.9235
	$r(50)$ 'private'	3.44	9.14	26.36	15.63	29.98	19.28	17.29	4.60	8.03
Sim 2 Kattegat	<i>n</i>	90	90	90	90	90	90	90	90	90
	A	5	11	32	27	43	24	23	8	13
	H_E	0.186	0.864	0.951	0.923	0.973	0.871	0.915	0.605	0.843
	H_O	0.178	0.844	0.922	0.922	0.967	0.811	0.911	0.600	0.844
	HW	0.3708	0.2847	0.1597	0.5736	0.4542	0.0569	0.5347	0.5972	0.511
Sim 3 Belt Sea	<i>n</i>	128	128	128	128	128	128	128	128	128
	A	5	13	34	28	43	24	23	8	13
	H_E	0.182	0.847	0.954	0.917	0.969	0.826	0.921	0.605	0.830
	H_O	0.180	0.820	0.930	0.930	0.953	0.758	0.906	0.609	0.844
	HW	0.5306	0.5528	0.1250	0.7778	0.2111	0.0083	0.5347	0.5972	0.5111
Sim 4 The Sound n	50	50	50	50	50	50	50	50	50	
	A	4	10	25	25	40	21	20	8	11
	H_E	0.235	0.835	0.949	0.915	0.974	0.820	0.913	0.574	0.842
	H_O	0.220	0.881	0.920	0.900	0.940	0.740	0.880	0.520	0.860
	HW	0.4056	0.3056	0.2764	0.4764	0.1847	0.0653	0.3069	0.2514	0.6361
Sim 5 Western Baltic	<i>n</i>	59	59	59	59	59	59	59	59	59
	A	3	11	26	28	35	20	18	8	11
	H_E	0.173	0.852	0.953	0.892	0.932	0.621	0.904	0.476	0.798
	H_O	0.153	0.870	0.949	0.949	0.915	0.559	0.864	0.441	0.814
	HW	0.2736	0.8389	0.5653	0.9750	0.3944	0.0542	0.2264	0.2875	0.6361