ICES Journal of Marine Science

ICES Journal of Marine Science (2017), 74(8), 2135–2147. doi:10.1093/icesjms/fsx046

Original Article

Genetic analysis of goldsinny wrasse reveals evolutionary insights into population connectivity and potential evidence of inadvertent translocation via aquaculture

Eeva Jansson[1]*, María Quintela[1], Geir Dahle[1], Jon Albretsen[2], Halvor Knutsen[2,3], Carl André[4], Åsa Strand[4], Stein Mortensen[1], John B. Taggart[5], Egil Karlsbakk[1,6], Bjørn Olav Kvamme[1], and Kevin A. Glover[1,6]

[1]*Institute of Marine Research, PO Box 1870 Nordnes, 5817 Bergen, Norway*
[2]*Institute of Marine Research Flødevigen, 4817 His, Norway*
[3]*Centre for Coastal Research, University of Agder, 4604 Kristiansand, Norway*
[4]*Department of Marine Sciences-Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden*
[5]*School of Natural Sciences, Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, UK*
[6]*Institute of Biology, University of Bergen, PO Box 7803, 5020 Bergen, Norway*

**Corresponding author: tel: +47 55 23 85 00; fax: +47 55 23 85 31; e-mail: eeva.jansson@imr.no*

Jansson, E., Quintela, M., Dahle, G., Albretsen, J., Knutsen, H., André, C., Strand, Å., Mortensen, S., Taggart, J. B., Karlsbakk, E., Kvamme, B. O., and Glover, K. A. 2017. Genetic analysis of goldsinny wrasse reveals evolutionary insights into population connectivity and potential evidence of inadvertent translocation via aquaculture. – ICES Journal of Marine Science, 74: 2135–2147.

Received 14 December 2016; revised 27 February 2017; accepted 3 March 2017; advance access publication 18 April 2017.

The salmon industry is heavily dependent on wrasse for delousing infected fish. The goldsinny wrasse is numerically the most important, and each year, millions are harvested from the wild and transported large distances into fish farms. Population genetic knowledge is required to sustainably exploit this species. Here, 1051 goldsinny wrasses from 16 locations across Scandinavia, the British Isles, and Spain were genotyped with 14 microsatellite and 36 SNP markers. Within-population genetic diversity decreased towards north, and a genetic break was observed across the North Sea. Samples from Northern Norway differed from rest of the Scandinavian samples, and samples from the British Isles differed from the Spanish ones. Within Scandinavia, isolation-by-distance was detected. Observed genetic patterns fitted well with expectations derived from oceanographic drift simulations. A sample from mid-Norway deviated from these patterns however, and was genetically very similar to southern Scandinavian samples. We conclude that the population structure of this species is primarily determined by the opposing evolutionary forces of passive drift, limited adult migration and spawning-site fidelity, whereas the deviation in isolation-by-distance observed in mid-Norway is potentially caused by inadvertent translocations of wrasse from southern Scandinavia via current aquaculture practise. Inclusion of outlier loci gave greater resolution, suggesting that diversifying selection may also affect population structuring among goldsinny wrasses.

Keywords: cleaner fish, Ctenolabrus rupestris, escapees, genetic population structure, microsatellite, particle simulation, SNP.

Introduction

Population genetic patterns are shaped by a complex interplay of historical events, species-specific traits, ecological processes, geographical features (e.g. Bradbury *et al.*, 2008; Eldon *et al.*, 2016), and to an ever-increasing degree, anthropogenic impact (Micheli *et al.*, 2013; Henriques *et al.*, 2016). Knowledge of these patterns and the processes underlying them are of vital importance for the sustainable exploitation of populations, and the conservation of

species (Hauser and Carvalho, 2008; Allendorf *et al.*, 2010; Dudgeon *et al.*, 2012). Within marine fisheries, there is a concern that failing to take population genetic structure into consideration can lead to an unsustainable harvest, loss of genetic variation, ecosystem disturbance, and ultimately (local) population extinction (Ciannelli *et al.*, 2013).

Marine populations are often very large, with typically, high dispersal potential, and the environments they live in offer few absolute physical barriers to hinder migration (e.g. Hauser and Carvalho, 2008). When populations are well-connected, wide-scale genetic homogeneity is to be expected (Waples and Gaggiotti, 2006; Lowe and Allendorf, 2010)—a phenomenon often reported in genetic studies of marine organisms (e.g. Cassista and Hart, 2007; Côté *et al.*, 2013; Deagle *et al.*, 2015). However, there is emerging evidence that panmixia might be more of an exception than a rule even in the marine realm. It has been shown that: (i) hydrographic and biogeographical boundaries often create detectable genetic breaks or barriers (e.g. Sá-Pinto *et al.*, 2012; Blanco Gonzalez *et al.*, 2016), (ii) local adaptation is often observed (e.g. Berg *et al.*, 2015; Jorde *et al.*, 2015), and (iii) very abundant species may show genetic sub-structuring (e.g. Benestan *et al.*, 2015; Blanco-Bercial and Bucklin, 2016; Eldon *et al.*, 2016). Moreover, very small genetic differences can reflect biologically meaningful divergence (e.g. Purcell *et al.*, 2006; Hemmer-Hansen *et al.*, 2007; Knutsen *et al.*, 2011), and seemingly very similar species may show largely contradicting genetic patterns (e.g. Severance and Karl, 2006; DeFaveri *et al.*, 2012).

Wrasses (*Labridae*) are a large family of marine fish with over 500 described species worldwide. Within the North Atlantic, six species are present: cuckoo (*Labrus mixtus*), scale-rayed (*Acantholabrus palloni*), ballan (*Labrus bergylta*), corkwing (*Symphollus melops*), goldsinny (*Ctenolabrus rupestris*), and rock cook (*Centrolabrus exoletus*). Most of these species are small inshore reef-dwellers, and traditionally, have not been of economic interest nor exploited at large scale (Darwall *et al.*, 1992). However, due to the recent high demand for cleaner-fish to remove parasitic sea lice (*Lepeophtheirus salmonis*) from farmed Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*), ballan, corkwing, goldsinny and rock cook wrasse are all now extensively harvested from the wild (Skiftesvik *et al.*, 2014, 2015).

The use of cleaner fish within the aquaculture industry first started in Norway in 1988 and in different parts of the British Isles in 1989–1990 (Bjordal, 1988; Darwall *et al.*, 1992). In 1997, 3.5 million wild-caught wrasses were used in Norway (most of these being goldsinnies; Gjøsæter 2002); however, their use decreased in the period 1998–2005 due to the increasing reliance of the industry on chemotherapeutants for delousing farmed salmon. When the salmon louse started to develop resistance to delousing agents (Nilsen, 2008; see also Besnier *et al.*, 2014), the demand for cleaner fish skyrocketed and intensive capture of wild wrasses resumed around 2007 (Skiftesvik *et al.*, 2014, 2015). Currently, ∼20 million wrasses are caught annually in Norway (Norwegian Directorate of Fisheries; www.fiskeridir.no).

With demand outstripping local supply of cleaner fish, it is common for wild wrasses in southern Norway and the adjoining Swedish coast (Gjøsæter 2002; Svåsand *et al.* 2016), to be transported over large distances (often ≥1000 km), and released into fish farms in mid-Norway. Some of these wrasses escape from sea-cages (Woll *et al.*, 2013), and once the salmon production cycle has ended, surviving wrasse may also be released into the surrounding sea. Thus, through current aquaculture practice, millions of wrasses are harvested and translocated great distances each year. Furthermore, despite dissimilar life-history strategies and population ecology (Darwall *et al.*, 1992; Skiftesvik *et al.*, 2015), identical or very similar fishery restrictions have applied to all species of wrasse since 2011 in Norway. Another significant drawback in the management of wild wrasses is the lack of relevant population genetic knowledge of the individual species (but see Sundt and Jørstad, 1998; D'Arcy *et al.*, 2013; Blanco Gonzalez *et al.*, 2016).

The goldsinny wrasse is the smallest of the wrasses used as cleaner-fish (<18 cm), and has the widest Atlantic distribution from Morocco to ∼68° north in Norway (Pollard, 2010). Abundance is temperature-dependent (Darwall *et al.*, 1992), and population densities are much lower near the northern edge of the distribution (Sundt and Jørstad, 1998). Together with corkwing, goldsinnies are the most numerous wrasse species in Norway and Sweden (Skiftesvik *et al.*, 2014) but there are large regional differences in their abundance (Gjøsæter, 2002; Skiftesvik *et al.*, 2015). Male goldsinny wrasses occupy small (∼2 m^2) permanent territories, which they defend during the reproductive season between April and September (Hilldén, 1984; Darwall *et al.*, 1992). Unlike other wrasses in the Northeast Atlantic, they do not build nests for reproduction or show parental care, but spawn pelagically. Most of the eggs sink to the bottom near-by, but it has been estimated that ∼10% of the eggs float (Hilldén, 1984), and thus may be transported by currents.

High abundance and fecundity (∼20 000 eggs/year/female) suggest that goldsinny wrasse could be somewhat resilient to exploitation (Darwall *et al.*, 1992). Furthermore, pelagic eggs could promote population connectivity over larger areas (compared with other wrasses that have demersal eggs; Skiftesvik *et al.*, 2014) and buffer against local fishing pressure. On the other hand, the slow growth of this species (4–5 years to reach the minimum commercial size of 11 cm; Skiftesvik *et al.* 2014) combined with the high breeding-site philopatry (Hilldén, 1984) indicates that goldsinnies may be sensitive to overexploitation. The only population genetic study of goldsinny wrasses conducted so far was from the 1990s and using a limited number of allozyme markers. These studies reported significant differences between samples collected from southern and mid-Norway (Sundt and Jørstad, 1998), and also between inner fjord and coastal samples (Sundt and Jørstad, 1993).

Given the present exploitation of goldsinny wrasse for aquaculture, through extensive harvest in some regions, and inadvertent translocation to other areas, there is a pressing need to characterize the population structure of this species. Here, we used newly developed microsatellite and SNP markers (Jansson *et al.*, 2016) to genotype over 1000 individuals from 16 locations along the north-eastern Atlantic coast. To our knowledge, this is the first genetic study of this species that includes samples from outside Norway, and also the first study post 1990s, since when the exploitation and translocation of goldsinny wrasses has increased sharply. We combined oceanographic modelling of pelagic life stages with genetic patterns to study the importance of the species (passive) dispersal ability.

Material and methods

Sampling and genotyping

In total, 1051 goldsinny wrasses were collected from 16 locations (Figure 1, Table 1) along the species' North Atlantic distribution

range: Norway (six sites, $N_{tot} = 386$), south-western Sweden (five sites, $N_{tot} = 372$), British Isles (three sites, $N_{tot} = 173$), and Galicia, north-west Spain (two sites, $N_{tot} = 118$). Samples from Scandinavia were collected in June–August 2014 (except for the GOT/VAR samples which were collected in June 2015), from the British Isles in June–August 2015, and from Spain during January–February in 2016. Fish were caught in coastal waters using fyke nets, pots (in Scandinavia and UK), and octopus traps (in Spain). All samples used were collected in compliance with EU Directive 2010/63/EU, and the national legislations in each country. Fish were killed upon catch and samples were taken immediately or killed and whole fish stored frozen until sampling in laboratory facilities.

Genomic DNA was extracted from fin clips stored in absolute ethanol using the Qiagen DNeasy Blood & Tissue Kit. Samples were genotyped using 17 microsatellite and 48 nuclear SNP markers developed for this species (Jansson *et al.*, 2016). Amplification conditions were identical to those described in Jansson *et al.* (2016). Genotyping success for each locus and individual was monitored: a cut off value of ≥60% successful amplification (for all loci combined and for SNP and microsatellite loci separately) was used to accept or reject any locus or individual from further analyses.

Genetic analyses

Microsatellite loci were screened for null alleles, large allele drop outs and potential scoring errors with the software MICRO-CHECKER (v.2.2.3; van Oosterhout *et al.*, 2004). The frequency of detected null allele(s) was estimated with maximum likelihood method using the EM algorithm of Dempster *et al.* (1977) implemented in the software Genepop (v.4.3; Rousset, 2008). In addition, to evaluate the effect of inclusion of possible null allele(s) containing loci on population differentiation estimates, the software FreeNA (Chapuis and Estoup, 2007) was used. This method gives uncorrected and corrected F_{ST} values. Confidence intervals (95%) of null frequencies were based on 1000 bootstraps.

To test whether loci deviated from neutrality, outlier analyses were conducted for microsatellite and SNP datasets separately with LOSITAN (Antao *et al.*, 2008) and BayeScan (v.2.1; Foll and Gaggiotti, 2008). To avoid overrepresentation of Scandinavian samples in these tests, a subsample of 400 individuals was used (100 individuals from each area; Table 1). LOSITAN was run with the following settings: 50 000 simulations, 95% confidence interval, forced mean F_{ST}, and with a 0.05 false discovery rate. A stepwise mutation model was used for the microsatellite dataset, whereas for SNPs the infinite model was used. Default parameter setting was used for the BayeScan run (prior odds 10, samples size 5000, thinning interval 10 000, pilot runs 20, pilot run length 5000, and additional burn-in 50 000), and the decision whether the locus was under selection was based on the magnitude of Bayes Factor (BF) as suggested by Jeffreys [1961; a log10(BF) >0.5 "substantial" evidence for selection, 1.5–2.0 "very strong" and >2.0 "decisive"]. The outlier tests were repeated three times for each marker type to check for consistency.

Genepop v.4.3 (Raymond & Rousset, 1995; Rousset, 2008) was used in exact tests for locus, population-wise and global Hardy–Weinberg expectations (HWE). Tests were based on the Markov chain method with 10 000 dememorizations, 20 batches, and 5000 iterations per batch. Global HWE tests across loci and populations were performed with Fisher's method. Possible linkage

Figure 1. Sampling locations. Norwegian sites ($N = 6$) are marked with black squares, Swedish sites ($N = 5$) with empty circles, British Isles sites ($N = 3$) with filled grey circles, and Spanish sites ($N = 2$) with stars. Sampling location abbreviations are as given in Table 1.

(LD) between all locus pairs in each population and over all populations was also tested with Genepop using the same MCMC settings as above.

Genetic diversity indices; expected/observed heterozygosity (H_e/H_o), inbreeding coefficient (F_{IS}), number of alleles (A), and the number of effective alleles (N_E; for SNPs) were calculated with GenAlEx 6.5 (Peakall and Smouse, 2006, 2012). To test whether the obtained F_{IS} values deviated significantly from zero, corresponding 95% confidence intervals were calculated with software GENETIX (v. 4.05.2; Belkhir *et al.*, 2004) based on 500 bootstraps. FSTAT (v.2.9.3; Goudet, 2001) was used to calculate allelic richness (A_R) for microsatellite loci and to compare genetic diversity (measured as allelic richness, and observed and expected diversity of microsatellite loci) between different areas (Table 1). Probability values for comparisons were obtained from 500 permutations.

Pairwise genetic differentiation between all populations (F_{ST}; Nei, 1977) was calculated using GenAlEx 6.5. Probability for each F_{ST} was calculated based on 9999 permutations. Because two types of markers were used in parallel and produced highly concordant results, no correction for within-population diversity (see Meirmans and Hedrick, 2011) was employed. To investigate spatial population genetic patterns further, two different individual-based clustering approaches were employed: a

Table 1. Summary information on goldsinny wrasse samples including sampling location, used abbreviation, area (N_SCA for Northern Scandinavia, SCA for Scandinavia, BRI for British Isles, and GAL for Galicia), approximate geographical position (Lat = Latitude, Long = Longitude), mean surface temperatures for January and July, and number of samples (*N*).

			Geographic location		Mean temperature (°C)		
Sampling location	**Abbreviation**	**Area**	**Lat**	**Long**	**January**	**July**	***N***
Stefjorden (Tysfjord), Norway	STE	N_SCA	68.219 N	16.407 E	4.6	10.8	30
Bodø, Norway	BOD	N_SCA	67.443 N	14.667 E	4.6	13.8	49
Flatanger, Norway	FLA	SCA	64.514 N	10.711 E	6.5	13.6	81
Bergen, Norway	BER	SCA	60.426 N	5.294 E	6.2	14.4	32
Flødevigen, Norway	FLO	SCA	58.874 N	8.779 E	4.7	17.2	80
Hvaler, Norway	HVA	SCA	59.045 N	10.932 E	4.4	17.5	100
Koster Island (Strömstad), Sweden	KOS	SCA	58.874 N	11.006 E	3.9	17.8	50
Lysekil, Sweden	LYS	SCA	58.275 N	11.415 E	4.0	17.7	100
Hälsö, Sweden	HAL	SCA	57.737 N	11.632 E	3.2	17.9	50
Gothenburg, Sweden	GOT	SCA	57.649 N	11.845 E	3.2	17.9	94
Varberg, Sweden	VAR	SCA	57.102 N	12.238 E	2.2	18.3	94
Isle of Mull, Scotland UK	SCO	BRI	56.431 N	6.184 W	8.2	13.6	50
Weymouth, South England UK	SEN	BRI	50.574 N	2.447 W	9.7	15.2	63
Mulroy Bay, Ireland	IRE	BRI	55.148 N	7.685 W	9.9	14.1	60
A Coruña, Galicia North, Spain	GAL1/GAL_N	GAL	43.378 N	8.474 W	13.6	17.4	55
Aldán, Galicia South, Spain	GAL2/GAL_S	GAL	42.444 N	8.891 W	14.1	17.4	63

Bayesian method using the software STRUCTURE (v.2.3.4; Pritchard *et al.*, 2000; Falush *et al.*, 2003), and discriminant analysis of principal components (DAPC; Jombart *et al.*, 2010) implemented in the ADEGENET package (v.1.4-2; Jombart, 2008; Jombart and Ahmed, 2011) in R (version 3.2.2; R Core Team, 2015). To assess the most likely number of subpopulations (*K*), ten independent Structure runs for fixed *K* values from 1 to 5 were performed (no larger values of *K* were tested based on results from preceding short test runs; data not shown). The combined dataset including both classes of markers were used, and each run consisted of 1 000 000 MCMC replicates after an initial burn-in of 100 000 (enough to reach convergence). An admixture model was chosen, and the allele frequencies were assumed correlated. Runs were performed for the whole dataset ($N = 1051$), as well as for Scandinavian ($n = 758$) and non-Scandinavian ($n = 293$) samples separately. Due to detected weak genetic differentiation within and outside Scandinavia (see Results), sampling locations were given as *a priori* for the separate runs (for inference of weak population structure, see Hubisz *et al.*, 2009). To assess the most likely number of clusters, the output from each run was analysed using the Evanno method (Evanno *et al.*, 2005) implemented in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Software CLUMPP (Jakobsson and Rosenberg, 2007) was used to average individual membership coefficients (*Q*) across the runs using the *LargeKGreedy* algorithm and G' pairwise similarity statistics.

Next, the DAPC approach was employed. As this method does not make any assumptions regarding population genetic models, it may be more effective for identifying hierarchical structures and genetic clines (Jombart *et al.*, 2010). DAPC was performed for the whole dataset as well as for a dataset where outliers SNPs were removed. Successive *K*-means clustering was run with the "find.clusters" function with a maximum *K* set to 15. The value of BIC (Bayesian Information Criterion) decreased only subtly after $K = 2$–4 (Supplementary Figure S1) suggesting that the most likely number of clusters is within this range. Based on pairwise F_{ST} values and Structure results, the "dapc" function was executed using a grouping based on four main areas: North Scandinavia, Scandinavia, British Isles, and Spain (see Table 1) with 70 PC axes retained (explaining >80% of variation). This grouping was also used to test the power of re-assignment of individuals back to sampling localities. To evaluate the used grouping and to avoid over-fitting (i.e. using too many PCs), a cross-validation approach with 10% of the data as a test data set was used. Based on cross-validation, the number of PCs was reduced to 50. Re-assignment was repeated with the leave-one-out procedure in software GENECLASS2 (Piry *et al.*, 2004) using the same main areas as baseline populations (i.e. putative origins), Rannala and Mountain (1997) criterion for calculation, and a threshold of 0.05.

Instead of having clear hierarchical subunits, natural populations are often gradually differentiated in space due to limited dispersal (i.e. isolation-by-distance, IBD). This underlying pattern can lead to spurious results in cluster analyses (Frantz *et al.*, 2009; Meirmans, 2012). Geographic distance between approximate sampling locations (Table 1) was calculated as the shortest waterway distance, and the possible linear association between genetic and geographic distances was tested. First, a simple linear model was created, and if linear regression between parameters was confirmed, a Mantel test (Mantel, 1967) was performed in the software PaSSaGE (v.2; Rosenberg and Anderson, 2011) using 999 permutations.

The association between genetic structure and environment (temperature) was investigated using the spatial analysis method (SAM) described by Joost *et al.*, (2007). SAM calculates logistic regressions between all possible marker–environmental pairs and determines whether a model including an environmental variable is more informative than a model including only the constant. The effect of temperature was tested using the mean annual seawater surface temperature, its standard deviation, as well as January and July temperatures as explanatory factors (all measures were averaged across several years, and the website http://www.seatemperature.org was used as a source for all the variables; Table 1). A model was considered significant only if both G and Wald Beta 1 tests rejected the corresponding null hypothesis with

Table 2. Summary statistics of genetic variability within each sampling location.

	Microsatellite results (averaged over 14 loci) (N=1 032)				SNP results (averaged over 36/34[a] loci) (N=1 036)		
Sample location	A	A_R	H_e	F_{IS}	N_E	H_e	F_{IS}
STE	9.6	8.2	0.632	0.036	1.588/1.586	0.347/0.345	**0.131**/**0.102**
BOD	10.7	8.1	0.636	−0.006	1.587/1.594	0.349/0.351	0.034/0.008
FLA	12.0	8.4	0.655	−0.002	1.641/1.651	0.371/0.375	**0.062**/0.047
BER	10.5	8.6	0.655	0.011	1.628/1.636	0.363/0.365	0.050/0.030
FLO	12.1	8.4	0.653	0.031	1.647/1.652	0.373/0.375	0.043/0.025
HVA	13.1	8.5	0.651	−0.006	1.631/1.633	0.366/0.366	**0.046**/0.031
KOS	11.1	8.4	0.655	0.016	1.615/1.652	0.360/0.363	0.024/0.009
LYS	13.0	8.3	0.653	0.014	1.631/1.637	0.367/0.368	0.019/0.005
HAL	11.9	8.8	0.670	−0.004	1.639/1.642	0.369/0.370	0.053/0.032
GOT	12.8	8.4	0.651	0.005	1.649/1.653	0.373/0.374	**0.063**/**0.044**
VAR	13.0	8.5	0.679	**0.043**	1.606/1.617	0.356/0.359	**0.055**/0.037
SCO	12.1	9.1	0.679	0.014	1.650/1.643	0.370/0.366	**0.074**/0.055
SEN	11.1	8.9	0.667	−0.014	1.603/1.608	0.345/0.346	−**0.302**/−**0.312**
IRE	11.6	8.8	0.676	**0.062**	1.616/1.613	0.354/0.352	0.035/0.018
GAL_N	12.4	9.1	0.685	**0.041**	1.650/1.634	0.367/0.360	0.018/0.019
GAL_S	13.1	9.0	0.683	0.002	1.661/1.647	0.376/0.371	0.014/0.012
Mean	11.9	8.6	0.661	0.015	1.628/1.631	0.363/0.363	0.026/0.010

SNP results are given with and without two loci deviating from HWE.
[a]Two loci deviating from HWE removed.

the threshold of 1.50×10^{-5} after Bonferroni correction. Individuals with missing markers were purged due to the impossibility of computing the G test. The aforementioned analyses were restricted to the loci with a major allele frequency between 5 and 95% across the whole dataset.

Simulation of drift and connectivity among locations

Oceanographic drift modelling was used to predict population connectivity based on transport of pelagic eggs and larvae and to compare expected drift with the observed genetic connectivity patterns. The hydrodynamic model used is described in detail in Lien *et al.* (2014), and the particle-tracking algorithms applied are similar to the methods in Vikebø *et al.* (2010). The ocean current model used had a horizontal resolution of 4 km and applied 32 vertical, topography-following levels, and daily averaged model currents from 55 spawning seasons (1960–2014) was used as input to the trajectory model. Due to data availability restrictions, Spanish sites were excluded from this analysis. The same number of particles (1400) was released from each of the 14 locations, all representing slightly offshore/exposed locations due to limitations of the resolution in the ocean current model. The floats were released every tenth day during pre-defined spawning periods, so that Scandinavian samples up to Bergen area (Figure 1) had a time window from 31st of May to 10th of July, whereas for the rest of the samples the interval was set from 30th of June to 10th of August. Releases of floats followed a simple Gaussian distribution in time. An equal number of particles was released every meter between 1 and 7 m depth. Drift period was set to 25 d for all floats (Darwall *et al.*, 1992). The simulation was repeated over 55 spawning seasons, and connectivity matrices with standard deviations were constructed between locations. Connectivity patterns measured as expected passive drift between locations and observed genetic divergence were compared visually as well as with Mantel's test using 999 permutations with the software PaSSaGE.

Results

The final dataset consisted of 14 microsatellite and 36 SNP markers. Data validation steps are explained in detail in Supplementary Text File 1. Two of the SNP loci were identified as possible outliers (*Locus4688_92* and *Locus5704_64*), and thus the subsequent analyses were performed with and without them. All 1051 samples were included, but for the separate analyses of the SNP and microsatellite datasets, 1036 and 1032 samples were acceptable, respectively. The amount of genetic variation across loci was highly variable: gene diversity (H_e) range for microsatellite loci was from ∼0.10 to almost 0.95, and for SNPs from ∼0.08 to 0.50 (Supplementary Tables S1 and S2), whereas averaged H_e estimates across populations were rather similar ranging from 0.63 to 0.68 for microsatellites, and from 0.35 to 0.38 for SNPs (Table 2). There was a general trend towards (slightly) positive F_{IS} values, and a significant deficiency of heterozygosity was observed in three populations (VAR, IRE, and GAL_N) with microsatellites, and in another two with SNPs (STE and GOT).

Decreasing genetic diversity towards north was observed (Table 2). For the microsatellite markers, North Scandinavian populations (N_SCA; Table 1) had significantly lower heterozygosity than the rest of the Scandinavian samples (p-value for H_o: 0.018, for H_e: 0.034). When comparing all Scandinavian to all British Isles populations, significantly lower allelic richness ($p = 0.008$) and gene diversity ($p = 0.012$) were detected in Scandinavia (though p-value for H_o was non-significant 0.206). The same comparison between Scandinavian and Spanish samples gave an even stronger signal of reduced diversity (p-values of 0.002, 0.054 and 0.002 for allelic richness, observed heterozygosity and gene diversity, respectively).

Genetic differentiation and role of outliers

Overall, genetic divergence between populations was low to moderate (Table 3), with the highest pairwise F_{ST} values ∼0.05. However, some distinct genetic patterns were found irrespective the marker type used. First, Scandinavian populations were

Table 3. Pairwise genetic differentiation (F_{ST}) between goldsinny wrasse sampling locations.

	STE	BOD	FLA	BER	FLO	HVA	KOS	LYS	HAL	GOT	VAR	SCO	SEN	IRE	GAL_N	GAL_S
STE		0.0092	**0.0124**	0.0135	0.0194	0.0184	0.0202	0.0186	0.0227	0.0177	0.0217	0.0374	0.0417	0.0352	0.0453	0.0384
BOD	0.0072		0.0108	0.0147	0.0162	0.0162	0.0175	0.0147	0.0165	0.0147	0.0170	0.0345	0.0367	0.0303	0.0393	0.0330
FLA	0.0073	0.0043		0.0052	0.0042	0.0038	0.0047	0.0037	0.0059	0.0039	**0.0056**	0.0357	0.0413	0.0332	0.0424	0.0371
BER	0.0118	0.0091	0.0061		0.0082	0.0079	0.0094	0.0083	0.0066	0.0069	**0.0099**	0.0400	0.0472	0.0419	0.0512	0.0462
FLO	0.0079	0.0049	0.0029	0.0051		0.0027	0.0033	0.0038	0.0052	0.0029	0.0043	0.0403	0.0459	0.0380	0.0463	0.0414
HVA	0.0096	0.0074	0.0042	0.0055	**0.0048**		0.0033	0.0033	0.0040	0.0027	0.0033	0.0351	0.0408	0.0335	0.0432	0.0393
KOS	0.0093	0.0071	0.0043	0.0049	0.0038	0.0032		0.0054	0.0070	0.0042	0.0062	0.0426	0.0468	0.0383	0.0486	0.0435
LYS	**0.0087**	0.0071	0.0039	0.0048	0.0043	0.0028	0.0040		0.0033	0.0022	0.0030	0.0393	0.0456	0.0374	0.0474	0.0430
HAL	0.0098	0.0069	0.0039	0.0057	0.0042	0.0037	0.0045	0.0034		0.0035	0.0052	0.0386	0.0481	0.0404	0.0479	0.0444
GOT	**0.0094**	0.0076	0.0043	0.0046	0.0039	0.0028	0.0033	0.0031	0.0047		0.0043	0.0380	0.0436	0.0363	0.0448	0.0402
VAR	0.0072	0.0070	0.0036	0.0050	0.0028	**0.0040**	0.0040	0.0037	0.0044	0.0039		0.0381	0.0415	0.0348	0.0485	0.0440
SCO	0.0368	0.0283	0.0273	0.0330	0.0279	0.0267	0.0276	0.0293	0.0255	0.0296	0.0327		0.0121	0.0084	0.0140	0.0108
SEN	0.0397	0.0326	0.0306	0.0385	0.0319	0.0305	0.0306	0.0325	0.0298	0.0328	0.0364	0.0082		0.0064	0.0179	0.0159
IRE	0.0328	0.0266	0.0260	0.0305	0.0262	0.0242	0.0250	0.0275	0.0231	0.0275	0.0305	0.0061	0.0067		0.0133	0.0090
GAL_N	0.0386	0.0327	0.0301	0.0366	0.0322	0.0312	0.0316	0.0339	0.0286	0.0350	0.0364	0.0116	0.0115	0.0099		0.0056
GAL_S	0.0362	0.0300	0.0280	0.0336	0.0299	0.0274	0.0285	0.0304	0.0252	0.0314	0.0334	0.0083	0.0088	**0.0073**	0.0041	

The upper right corner show results for 36 SNP markers and bottom left corner for 14 microsatellite markers. Statistically significant results are underlined ($p < 0.05$), bolded ($p \leq 0.01$), or with grey background ($p \leq 0.001$). Probability for each value was calculated based on 9 999 permutations.

clearly differentiated from British Isles and Spanish populations ($F_{ST} \sim 0.02$–0.05). Moreover, Northern Scandinavian samples from Stefjorden and Bodø differed (mainly) from the rest of the Scandinavia ($F_{ST} \sim 0.005$–0.02), and Spanish samples from the British Isles samples ($F_{ST} \sim 0.01$–0.02). Interestingly, inclusion of the two outlier SNP loci clearly increased the resolution power within Scandinavia (showing larger differences between N_SCA vs. SCA; Table 3, cf. Supplementary Table S3) but at the same time led to lower discriminatory power on broader scale (i.e. comparisons of populations across the North Sea without outliers showed higher divergence).

The winter temperature across the geographic span of samples (Table 1) ranged between 2.2 (VAR) and 14.1 °C (GAL_S) and was found to be associated to patterns of genetic differentiation at 31 markers; nine microsatellite and 14 SNP loci, respectively (Supplementary Table S4a). Summer temperature, ranging between 10.8 (STE) and 18.3 °C (VAR), correlated with markers *Cru037_155*, *Locus5704_64_A*, and *Locus4263_1032_A*. Thus, only two markers: *Cru037_155* (microsatellite) and *Locus5704_64_A* (SNP) were found to correlate with temperature irrespective of the season. When restricting the data set to Scandinavia, no outliers were found for winter temperatures (ranging between 2.2 and 6.5 °C). However, summer temperatures were linked to one allele (nucleotide A) in *Locus5704_64*. This marker was also found to be associated to mean annual temperature and its standard deviation (Supplementary Table S4b). Interestingly, *Locus5704_64* was also indicated to be under directional selection by BayeScan and LOSITAN.

Sub-structuring and reassignment

Individual cluster analyses gave concordant results to those based on population differentiation: a DAPC plot (Figure 2) using all markers showed clear distinction between populations across the North Sea. Moreover, divergence between the Spanish and British Isles samples was evident, as well as between the Northern Scandinavian and rest of the Scandinavian samples (with geographically intermediate populations from Flatanger and Bergen located in midway on the plot). When DAPC analysis was performed without the outlier loci, no population sub-structuring was found within Scandinavia (Supplementary Figure S2), but distinction between the Spanish and British Isles populations became clearer.

The major dichotomy separating samples either side of the North Sea was also the main finding in the Structure analysis (Figure 3a; with $K = 2$, $\Delta K = 2851.3$), and represented the highest hierarchical level of population structuring. In separate runs for both groups, samples within and outside Scandinavia, $K = 3$ led to highest mean LnP(K) and ΔK values (Supplementary Figures S3b and c) suggesting three groups as the most plausible subdivision. However, inspection of the bar plots from these simulations (Figure 3b) revealed subtle and more gradual differences (with asymmetrical individual assignments) than distinct clustering. Among Scandinavian samples, individuals from Northern populations (STE/BOD) displayed differing admixture proportions. Spanish populations were very similar to each other and different from the British Isles populations. Within the British Isles, Scottish samples had somewhat differing admixture proportions compared with the Ireland and South England samples.

Re-assignment of individuals into their putative areas of origin (N-SCA, SCA, BRI, and GAL; Table 1) had a very good average

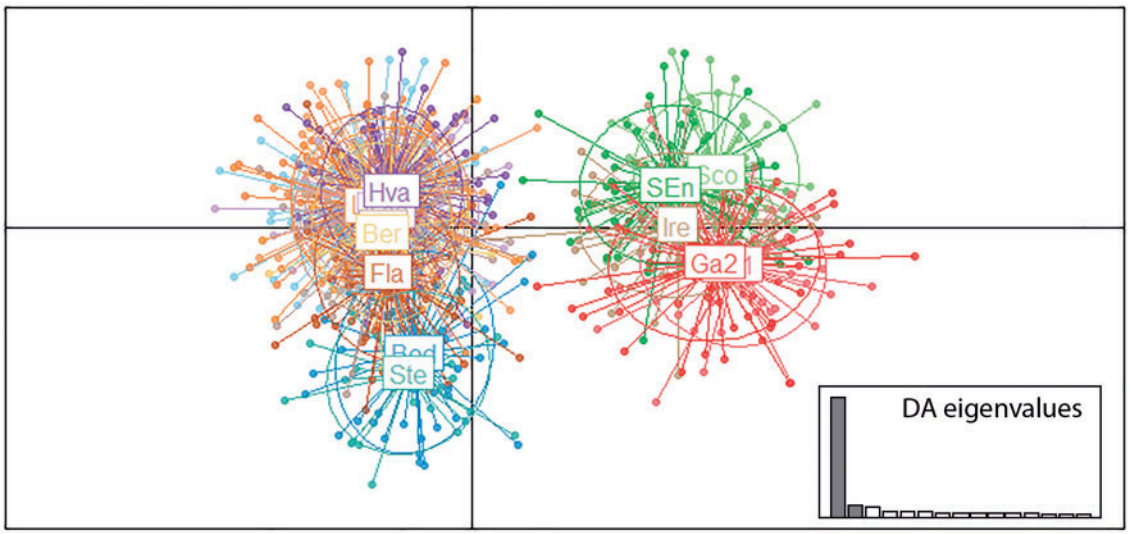

Figure 2. Discriminant analysis of principal components for goldsinny wrasse samples. Fifty markers were used including two outlier SNPs. Projected inertia % for the axes: PC1 = 5.08%, PC2 = 2.35%. All Scandinavian samples are grouped on the left, with northernmost populations (BOD and STE) separated along the second axis. Samples from British Isles (SCO, SEN and IRE) and Spain (GA1 and GA) cluster together on the right side. Corresponding DAPC plot without outlier loci is shown in Supplementary Figure S2.

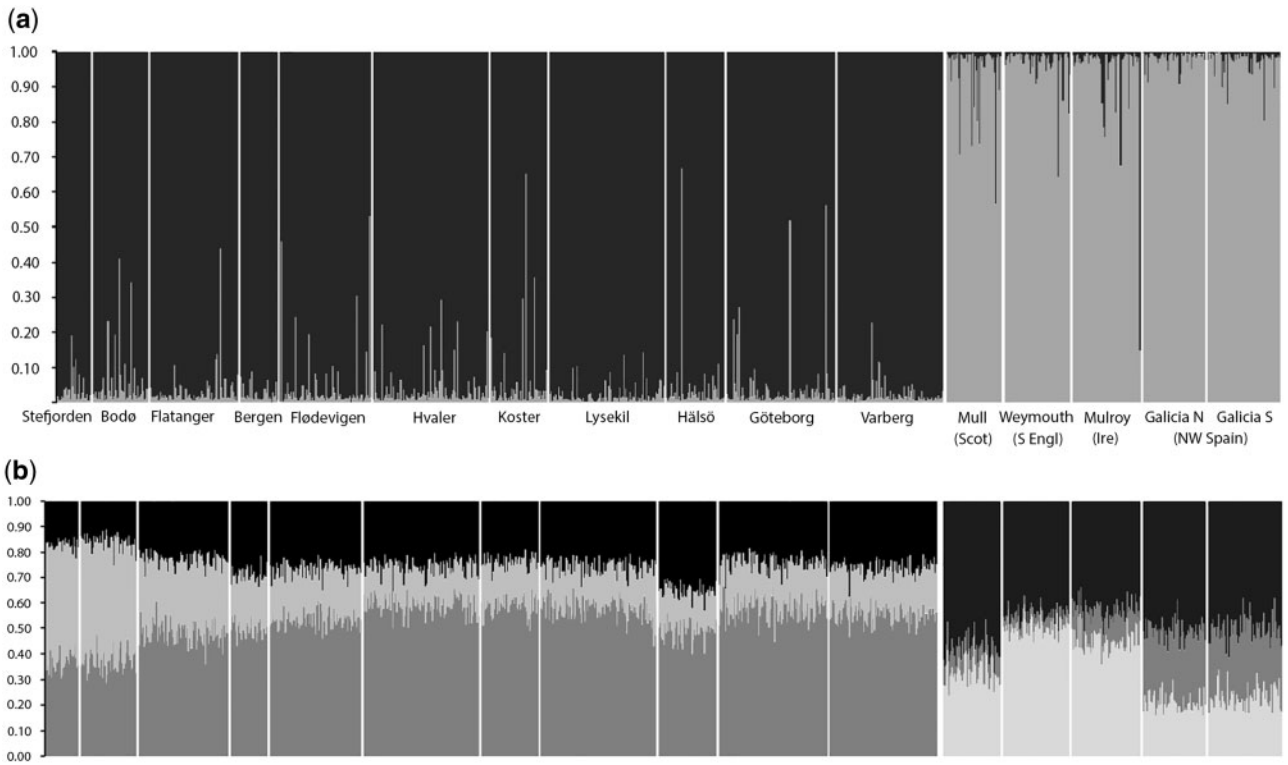

Figure 3. Bayesian clustering of goldsinny wrasse samples performed in STRUCTURE. All loci were used and results averaged over ten runs with CLUMPP. Each vertical bar represents one individual and its colour segments the probability to belong to different clusters. (a) Clustering for the whole dataset with the most supported $K = 2$. (b) Regional analyses with some substructure found; $K = 3$ was the most supported solution.

success rate ranging from 87.7 to 92.5% (Figure 4; see Supplementary Figure S4 for results of assignment on individual level). This indicates that genetic differences between the three main regions were large enough for robust genetic-assignment.

Comparison of genetic and waterway distances between sampling locations demonstrated that these parameters were correlated for both marker types (i.e. IBD, see Supplementary Figure S5). In addition to a general association between genetic and

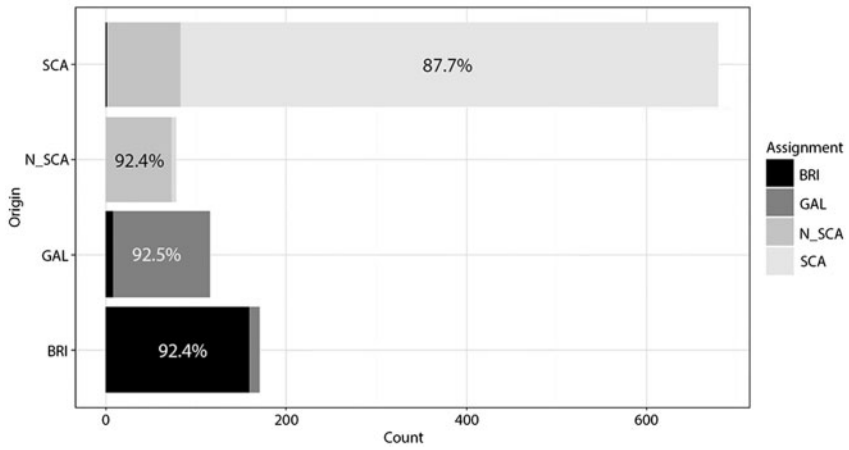

Figure 4. Re-assignment of individuals probabilities back to broader-scale sampling areas. Each bar represents samples from one area, British Isles (BRI), Galicia, Spain (GAL), Northern Scandinavia (N_SCA), and Scandinavia (SCA), whereas colour segments denote proportions where the individuals were assigned to with highest probability. Percentage shown in each bar is the proportion of correct assignments, i.e. to the same area where the samples originated from.

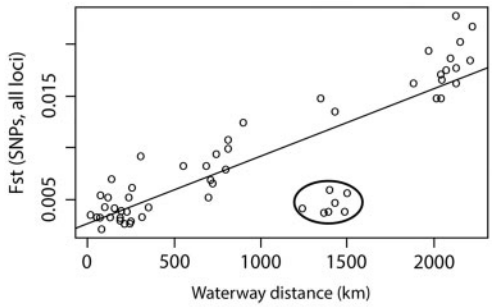

Figure 5. Isolation-by-distance within Scandinavia. Figure shows the correlation between waterway and genetic distances within the Scandinavian sampling locations using SNPs ($r = 0.841$, $p < 0.0001$). Comparisons between Flatanger and Southern Scandinavia showing lower than expected divergence are circled. Pairwise comparison excluding outlier SNPs and comparison using microsatellites are given in Supplementary Figure S6a and b.

geographic distances, there were also region-specific patterns. IBD tests showed a strong linear positive correlation between genetic and waterway distances in Scandinavia. A linear model displayed a good fit (data not shown) and for SNPs (Figure 5), oceanographic distance explained ~70% of the variation in genetic divergence ($p < 0.001$). Removal of outlier loci did not change the results (Supplementary Figure S6a). For microsatellites, a similar but less clear pattern was observed, and the explanatory power of the model was lower (~50%; Supplementary Figure S6b).

Seven population pairs within a distance of ~1200–1500 km of each other displayed distinctly low differentiation (Figure 5). All of these comparisons were between the sample collected from Flatanger in mid-Norway (Table 1), and all sampling sites in Southern Scandinavia (i.e. populations south of Bergen). To test whether genetic differentiation was significantly lower than expected, a new independent IBD model without those seven comparisons was calculated ($y = 0.002797 + 0.000007558x$; $r = 0.9577$; $p < 0.0001$). Based on the model, expected F_{ST}-values for each of the seven distances were calculated with 95% confidence intervals (data not shown). In all cases, the observed value fell clearly (two- to threefold) below the lower *CI* bound indicating a significant deviation for these seven data points.

Oceanic connectivity modelling

Oceanographic drift simulations (Figure 6; see also Supplementary Figures S7 and S8a and b) showed that a high degree of connectivity via transport of pelagic egg and larvae is to be expected within Southern Scandinavia in the Kattegat and the Skagerrak area. The main transport pathway from the Skagerrak is from the south toward north (Supplementary Figure S8a) along with the Norwegian Coastal Current (NCC; Supplementary Figure S9). The NCC is likely to contribute with some northward drift to the Bergen area, then further from Bergen to Flatanger, and from Flatanger to Bodø and Tysfjorden. The northernmost sampling site in Tysfjorden is likely to have a very high self-retention rate ($44 \pm 13\%$ of drifting particles do not leave the area), but this area will potentially also receive some inflowing particles from the Bodø area. It is noteworthy that no (direct) drift is expected between mid-Norway and south-Scandinavia, and that among the sampling sites from the British Isles, only a minute amount of drifters is likely to flow from Ireland to Scotland. Any drift between Scandinavia and the British Isles is also unlikely to happen.

The genetic relationships among the sampling locations were strongly in agreement with the results from the drift model. Because no connectivity between Scandinavia and the British Isles was expected (Supplementary Figure S8a), correlation tests between the observed genetic divergence and expected connectivity were restricted to Scandinavian samples only. Percentages of simulated unidirectional floats between sampling locations were combined (i.e. floats to/from between any specific location pair) into one matrix. Significant and from intermediate to rather strong negative correlation between the variables was confirmed (for SNPs: $Z = 4.38$, $r = -0.482$, $t = -2.27$, $p = 0.004$; for microsatellites: $Z = 143\,326.28$, $r = -0.653$, $t = -3.52$, $p = 0.001$).

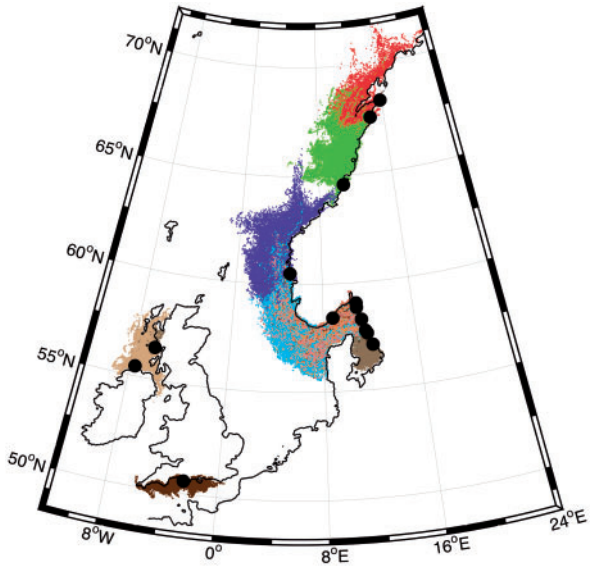

Figure 6. Modelled oceanographic drift of particles released near sampling locations. The black circles show the offshore release locations and the coloured clouds where the particles are expected to drift after 25 d. Drift was simulated during May–August in 1960–2014, and the averaged results over the years are shown in this figure. Drift results for each location separately are shown in Supplementary Figure S7. A connectivity matrix between locations is given in Supplementary Figure S8a and b.

Discussion

This is the first comprehensive study of population genetic structure in the goldsinny wrasse, a species heavily exploited in some regions through fishing, and translocated to other locations to serve as a cleaner-fish in salmon aquaculture. A large genetic-break was revealed across the North Sea. This suggests a lack of direct genetic connectivity between Scandinavian (Swedish and Norwegian) and other European (British Isles and Spain) populations. Within these two regions, further population structuring was observed, and a trend towards reduced genetic variation was observed in samples collected from the northern areas. Although the goldsinny wrasse displays potentially dispersive planktonic egg and larval stages, we conclude that restricted adult movement, limited larval dispersal, spawning site fidelity, as well as other potential mechanisms limit genetic exchange within this species. Furthermore, the unexpectedly high genetic similarity between the sample from Flatanger in mid-Norway, which is an aquaculture intense region where goldsinny and other wrasses are routinely transported to, and samples from southern Norway/Sweden where goldsinny and other wrasses are routinely harvested from and supplied to the aquaculture industry in mid-Norway, provides the first potential evidence of inadvertent mixing of genetically distinct stocks associated with the use of wrasse as cleaner fish in the aquaculture industry.

The major genetic break across the North Sea reported in this study for the goldsinny wrasse has previously been observed for both corkwing (Robalo *et al.*, 2011; Knutsen *et al.*, 2013) and ballan wrasses (D'Arcy *et al.*, 2013; Quintela *et al.*, 2016), indicating that despite pelagic life stages, large areas of open deep water (the North Sea) can act as effective dispersal barriers for these species. The observation of lower genetic diversity in the Scandinavian goldsinny populations compared with more southern populations is also consistent with the results of studies of ballan and corkwing wrasses. Both above-mentioned patterns are probably shaped by historical events, namely (re-)colonization of species when the last glacial maximum (~21 kb; Lambeck *et al.*, 2010) ended, and ice sheets covering the entire Scandinavia started to retreat quickly about 10–11 000 years before present. The following range shift towards north has left its traces on present-day population gene pools of various organisms via founder and bottleneck effects where only a limited number of individuals successfully colonized new areas (Hewitt, 2000; Coyer *et al.*, 2003; Mäkinen *et al.*, 2006), or in some cases, survived and spread from the few remaining ice-free areas (Parducci *et al.*, 2012; Lagerholm *et al.*, 2014).

Goldsinny wrasse displays a lower level of population-genetic divergence than another north-eastern Atlantic wrasse, corkwing. Although the measured F_{ST} across the North Sea was on average 0.031/0.041 for goldsinny wrasse (for microsatellites/SNPs, respectively; Table 3), the corresponding estimate for corkwing wrasse was four- to fivefold higher, 0.159 (using nine microsatellite loci; Knutsen *et al.*, 2013). Also, the reduction of genetic diversity in Scandinavia reported for corkwing (≥30% microsatellite variation lost compared with British Isles populations; Knutsen *et al.*, 2013), and ballan wrasse (150 alleles were found among 89 samples from Galicia, Spain vs. 115 among 241 samples from Norway; Quintela *et al.*, 2014, 2016) was much less pronounced in the case of goldsinny wrasse: mean gene diversity was only ~4% and allelic richness ~6% lower in Scandinavia compared with the British Isles samples (Table 2). These differences are likely due to the differences in breeding ecology between these species. Although other wrasse species spawn in nests and have benthic eggs, the goldsinny wrasse has planktonic eggs (Darwall *et al.*, 1992). Even though only a very small portion of these eggs would be flushed offshore (Hilldén, 1984) and carried away by currents, more effective dispersal and higher connectivity between (nearby) populations would be expected compared with the other wrasse species, which have stationary eggs and only larvae are pelagic. Parallel comparisons of fish species with differing duration of pelagic life stage have shown that species with longer pelagic stages generally show less population sub-structuring (e.g. Purcell *et al.*, 2006; Young *et al.*, 2015).

An extended pelagic phase can help to override unsuitable habitats, colonize new areas, and expand distribution area. The goldsinny wrasse inhabits inshore habitats with rocks and vegetation (Darwall *et al.*, 1992), whereas sandy habitats may not be able to hold viable wrasse populations (Knutsen *et al.*, 2013). Extensive sandy areas around the Jæaren and Lista in south-western Norway were recently suggested to act as a dispersal barrier for corkwing wrasse (Blanco Gonzalez *et al.*, 2016), separating western and southern Norwegian populations. This study did not include samples close to this area, but surrounding sampling points further away (BER/FLO; Table 1) showed low and non-significant divergence (F_{ST}=0.0051/0.0082; Table 3) suggesting that at least such rather short (~26 km; Blanco Gonzalez *et al.*, 2016) habitat discontinuities are insufficient to create genetic barriers between goldsinny wrasse populations.

Oceanographic drift models of passive dispersal have often proven relatively good predictors of (genetic) connectivity in marine fish (e.g. Coscia *et al.*, 2013; Knutsen *et al.*, 2013; Teacher *et al.*, 2013). This was also the case for the goldsinny wrasse in this study. Here, based on the simulated passive dispersal, a very high level of connectivity among sampling locations within southern

Scandinavia was expected. This was corroborated by the genetic data from both SNPs and microsatellites. Furthermore, the northernmost Scandinavian populations (BOD/STE) were genetically distinct, as predicted by the drift model, and mid-Norwegian samples (FLA/BER) were intermediate with some significant pairwise comparisons (Table 3). At least on a coarse coastal scale, the amount and direction of connectivity between goldsinny populations in Scandinavia is thus largely influenced by the Norwegian Coastal Current (Supplementary Figure S9), which has created the observed IBD pattern. On the contrary, even though there is a minor coastal flow around the British Isles, strong tides dominate the currents back and forth so that drifters are expected to spread more multi-directionally and not very far (Supplementary Figure S7). The drift model suggested some connectivity between Ireland and Scotland (but not between the other locations), which did not have significant genetic differentiation from each other. Small but significant differentiation between Southern English and Scottish samples was observed, but not between Irish and Southern English ones (Table 3). These somewhat contradictory results may be due to sampling gaps (see Selkoe and Toonen, 2011); with only three population samples collected from the British Isles, the true connectivity is likely underestimated if and when dispersal takes place predominantly between nearby locations in a stepping stone manner.

Historical events and passive drift are likely to have played a major role in shaping the observed population genetic structure among present-day goldsinny populations. However, the possibility of other forces being involved cannot be ruled out. First, it is possible that human-mediated gene flow via transport of goldsinnies to fish farms from south Scandinavia to West-Norway, which has been on-going for more than two decades (Sundt and Jørstad, 1998), may have decreased genetic divergence. Indeed, the level of genetic differentiation between Flatanger (one of primary recipient areas for translocations due to scarcity of wrasses locally) and southern Scandinavian sampling sites (i.e. source areas) was lower than expected (Figure 5), which indicates that this may have already occurred. However, because the general level of genetic differentiation was so low in Scandinavia, robust re-assignment that could give direct evidence of introgression was not feasible with the used marker set (except for distinguishing the northernmost samples; Figure 4).

Second, selection might also play a role shaping population genetic patterns of goldsinny wrasse. Two SNPs were detected as outliers and their inclusion clearly increased population-genetic resolution within Scandinavia. In addition, one SNP was correlated with some key temperature variables across the study region. Outlier loci have repeatedly come in useful to delineate marine population structures (e.g. Teacher *et al.*, 2013; Hemmer-Hansen *et al.*, 2014; Candy *et al.*, 2015), but their true biological significance can be hard to disentangle. For instance, if gene flow is reduced due to geographic distance but at the same time important environmental factor(s) (see Riginos *et al.*, 2016) forms a parallel gradient, consequent genetic patterns will be similar (Orsini *et al.*, 2013). "Allele surfing" during population expansion can also mimic positive selection patterns by creating allele frequency clines (Excoffier and Ray, 2008), and further complicate interpretation of detected genetic structures. We observed a congruent strong pattern of IBD ($r = 0.709$–0.841) within Scandinavia irrespective of the marker set used. Northernmost populations formed a separate genetic unit but to determine whether this is merely a matter of distance and neutral processes or also linked to adaptation to e.g. lower temperatures, needs further investigation.

It is noteworthy that, because sampling in this study was restricted to coastal areas only, possible additional genetic substructures, e.g. inside extensive and highly heterogeneous fjord systems within Norway would go undetected. In previous studies using allozymes, Sundt and Jørstad (1993, 1998) reported significant genetic differentiation of goldsinny wrasse within fjords. Regional genetic structuring has also been reported for corkwing wrasse (Blanco Gonzalez *et al.*, 2016): besides the above-mentioned major break due to the sandy area, a moderate IBD along the west coast and genetically fairly homogeneous southern population structure were detected. Similar observations from this study imply that this pattern—high homogeneity in south and gradual increase of genetic differences along the west coast—might be of more general phenomenon among Norwegian coastal fishes, and that the strength of this structuring would be determined by species-specific dispersal capabilities.

From a sustainable management point of view, the ongoing long-range aquaculture-related translocations of goldsinny wrasse from Sweden and southern Norway, to the west of Norway, may be questioned. First, transportation poses a threat of pathogen transmission between areas, and between wild and cultured fish (e.g. Treasurer, 2012; Wallace *et al.*, 2015). Second, transportation and subsequent (inadvertent) release enables gene flow between translocated and local populations, which can be detrimental. For instance, if fish stocks are locally adapted, maladapted genes can spread through introgression endangering the local populations (e.g. Laikre *et al.*, 2010). Third, local overexploitation may deplete source populations into a level where genetic stochasticity and risk of extinction increase considerably.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

This study was funded by the Norwegian Ministry for Trade, Industry and Fisheries. The Swedish Cultural Foundation in Finland (Svenska Kulturfonden) is acknowledged for personal grant to EJ. Additional support for people involved in this study was provided by the Swedish Research Council FORMAS, the GU Centre for marine evolutionary biology (Cemeb) and EU Interreg. We thank Lorenz Hauser and two anonymous reviewers for their constructive comments. We want to express our gratitude to Jacek Koszałka and Niklas Jansson for their help with figures, and to Lisbeth Sælemyr, Emil E. Høyesen, Ylva Fredricsson, Esra Hasan and Ann Cathrine B. Einen, Amy Callaghan, Catherine McManus, Per Andersen, Ole Ingar Paulsen, Eva Farestveit, Kim Halvorsen, Reine Andreasson, Kerstin Roysson, Rolf Sørensen, Jan Harald Haraldsen, Per Andersen and Nils Vestvik for providing fish or assisting with the sampling. We acknowledge the support of the Galician Council for Marine Affairs (Consellería do Mar, Xunta de Galicia), namely the cooperation of the fisherieś observers (L. Pérez Miser, J. Chapela Portela, C. Aguiar Couto, J.M. García Rozamontes, J.M. Garrido Vispo, J.M. Pérez Veres, J. García Romero, J. González Pérez), the coordinators (J. Molares Vila, F.J. Filgueira Rodríguez), the staff of the Monitoring Unit for Small-Scale Inshore Fisheries (UTPB) and R. Barreiro, B. Carro and C. Caramelo from the University of A Coruña, for their help to obtain the samples from Galicia.

References

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics, 11: 697–709.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., and Luikart, G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a *Fst*-outlier method. BMC Bioinformatics, 9: 323.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. 2004. GENETIX 4.0.5.2., Software under Windows™ for the Genetics of the Populations. Laboratory Genome, Populations, Interactions, University of Montpellier II, France.

Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S. et al. 2015. Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). Genome Biology and Evolution, 7: 1644–1663.

Besnier, F., Kent, M., Skern-Mauritzen, R., Lien, S., Malde, K., Edvardsen, R. B., Taylor, S. et al. 2014. Human-induced evolution caught in action: SNP-array reveals rapid amphi-atlantic spread of pesticide resistance in the salmon ectoparasite *Lepeophtheirus salmonis*. BMC Genomics, 15: 937.

Bjordal, Å. 1988. Cleaning symbiosis between wrasse (Labridae) and lice infested salmon (*Salmo salar*) in mariculture. International Council for the Exploration of the Sea, Mariculture Committee 188/F 17. 8 pp.

Blanco Gonzalez, E., Knutsen, H., and Jorde, P. E. 2016. Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. PLoS ONE, 11: e0163052.

Blanco-Bercial, L., and Bucklin, A. 2016. New view of population genetics of zooplankton: RADseq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*. Molecular Ecology, 25: 1566–1580.

Bradbury, I. A., Laurel, B., Snelgrove, P. V. R., Bentzen, P., and Campana, S. E. 2008. Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. Proceedings of the Royal Society of London. Series B, Biological sciences, 275: 1803–1809.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., and Bernatchez, L. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). Molecular Ecology, 24: 3299–3315.

Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A., and Narum, S. R. 2015. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. Molecular Ecology Resources, 15: 1421–1434.

Cassista, M. C., and Hart, M. W. 2007. Spatial and temporal genetic homogeneity in the Arctic surfclam (*Mactromeris polynyma*). Marine Biology, 152: 569–579.

Chapuis, M.-P., and Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution, 24: 621–631.

Ciannelli, L., Fisher, J. A. D., Skern-Mauritzen, M., Hunsicker, M. E., Hidalgo, M., Frank, K. T., and Bailey, K. M. 2013. Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review. Marine Ecology Progress Series, 480: 227–243.

Côté, C. L., Gagnaire, P.-A., Bourret, V., Verreault, G., Castonguay, M., and Bernatchez, L. 2013. Population genetics of the American eel (*Anguilla rostrata*): $F_{ST} = 0$ and North Atlantic Oscillation effects on demographic fluctuations of a panmictic species. Molecular Ecology, 22: 1763–1776.

Coyer, J. A., Peters, A. F., Stam, W. T., and Olsen, L. 2003. Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. Molecular Ecology, 12: 1817–1829.

Coscia, I., Vogiatzi, E., Kotoulas, G., Tsigenopoulos, C. S., and Mariani, S. 2013. Exploring neutral and adaptive processes in expanding populations of gilthead sea bream, *Sparus aurata* L., in the North-East Atlantic. Heredity, 108: 537–546.

Deagle, B. E., Faux, C., Kawaguchi, S., Meyer, B., and Jarman, S. N. 2015. Antarctic krill population genomics: apparent panmixia, but genome complexity and large population size muddy the water. Molecular Ecology, 24: 4943–4959.

D'Arcy, J., Mirimin, L., and FitzGerals, R. 2013. Phylogeographic structure of a protogynous hermaphrodite species, the ballan wrasse *Labrus bergylta*, in Ireland, Scotland, and Norway, using mitochondrial DNA sequence data. ICES Journal of Marine Science, 70: 685–693.

Darwall, W. R. T., Costello, M. J., Donnelly, R., and Lysaght, S. 1992. Implication of life-history strategies for a new wrasse fishery. Journal of Fish Biology, 41: 111–123.

DeFaveri, J., Shikano, T., Ghani, N. I. A., and Merilä, J. 2012. Contrasting population structures in two sympatric fishes in the Baltic Sea basin. Marine Biology, 159: 1659–1672.

Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B, 39: 1–38.

Dudgeon, C. L., Blower, D. C., Broderick, D., Giles, J. L., Holmes, B. J., Kashiwagi, T., Krück, N. C. et al. 2012. A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. Journal of Fish Biology, 80: 1789–1843.

Earl, D. A., and vonHoldt, B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4: 359–361.

Eldon, B., Riquet, F., Yearsley, J., Jollivet, D., and Broquet, T. 2016. Current hypotheses to explain genetic chaos under the sea. Current Zoology, 62: 551–566.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, 14: 2611–2620.

Excoffier, L., and Ray, N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends in Ecology and Evolution, 23: 347–351.

Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164: 1567–1587.

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

Frantz, A. C., Cellina, S., Krier, A., Schley, L., and Burke, T. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? Journal of Applied Ecology, 46: 493–505.

Gjøsæter, J. 2002. Distribution and density of goldsinny wrasse (*Ctenolabrus rupestris*) (Labridae) in the Risør and Arendal areas along the Norwegian Skagerrak coast. Sarsia, 87: 75–82.

Goudet, J. 2001. FSTAT. A Program to Estimate and Test Gene Diversities and Fixation Indices (version 2.9.3).

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

Hemmer-Hansen, J., Nielsen, E. E., Grønkjaer, P., and Loeschcke, V. 2007. Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). Molecular Ecology, 16: 3104–3118.

Hemmer-Hansen, J., Therkildsen, N. O., Meldrup, D., and Nielsen, E. E. 2014. Conserving marine biodiversity: insights from life-history trait candidate genes in Atlantic cod (*Gadus morhua*). Conservation Genetics, 15: 213–228.

Henriques, R., von der Heyden, S., Lipinski, M. R., du Toit, N., Kainge, P., Bloomer, P., and Matthee, C. A. 2016. Spatio-temporal

genetic structure and the effects of long-term fishing in two partially sympatric offshore demersal fishes. Molecular Ecology, 25: 5843–5861.

Hilldén, N.-O. 1984. Behavioural ecology of the Labrid fishes (Teleostei: Labridae) at Tjärnö on the Swedish west coast. Doctoral dissertation, University of Stockholm, Sweden.

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources, 9: 1322–1332.

Jakobsson, M., and Rosenberg, N. A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23: 1801–1806.

Jansson, E., Taggart, J., Wehner, S., Dahle, G., Quintela, M., Mortensen, S., Kvamme, B. O. et al. 2016. Development of SNP and microsatellite markers for goldsinny wrasse (*Ctenolabrus rupestris*) from ddRAD sequencing data. Conservation Genetics Resources, 8: 201–206.

Jombart, T. 2008. ADEGENET: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403–1405.

Jombart, T., and Ahmed, I. 2011. ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27: 3070–3071.

Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11: 94.

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., and Taberlet, P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Molecular Ecology, 16: 3955–3969.

Jorde, P. E., Søvik, G., Westgaard, J.-I., Albretsen, J., André, C., Hvingel, C., Johansen, T. et al. 2015. Genetically distinct populations of northern shrimp, *Pandalus borealis*, in the North Atlantic: adaptation to different temperatures as an isolation factor. Molecular Ecology, 24: 1742–1757.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., André, C., and Stenseth, N. C. 2011. Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Molecular Ecology, 20: 768–783.

Knutsen, H., Jorde, P. E., Gonzales, E. B., Robalo, J., Albretsen, J., and Almada, V. 2013. Climate change and genetic structure of leading edge and rear end populations in a northwards shifting marine fish species, the corkwing wrasse (*Symphodus melops*). PLoS ONE, 8: e6749.

Lagerholm, V. K., Sandoval-Castellanos, E., Ehrich, D., Abramson, N. I., Nadachowski, A., Kalthoff, D. C., Germonpré, M. et al. 2014. On the origin of the Norwegian lemming. Molecular Ecology, 23: 2060–2071.

Laikre, L., Schwartz, M. K., Waples, R. S., Ryman, N., and The GeM Working Group. 2010. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends in Ecology and Evolution, 25: 520–529.

Lambeck, K., Purcell, A., Zhao, J., and Svensson, N.-O. 2010. The Scandinavian Ice Sheet: from MIS 4 to the end of the Last Glacial Maximum. Boreas, 39: 410–435.

Lien, V. S., Gusdal, Y., and Vikebø, F. B. 2014. Along-shelf hydrographic anomalies in the Nordic Seas (1960–2011): locally generated or advective signals? Ocean Dynamics, 64: 1047–1059.

Lowe, W. H., and Allendorf, F. W. 2010. What can genetics tell us about population connectivity? Molecular Ecology, 19: 3038–3051.

Mäkinen, H. S., Cano, J. M., and Merilä, J. 2006. Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. Molecular Ecology, 15: 1519–1534.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research, 27: 209–220.

Meirmans, P. G. 2012. The trouble with isolation by distance. Molecular Ecology, 21: 2839–2846.

Meirmans, P. G., and Hedrick, P. W. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources, 11: 5–18.

Micheli, F., Halpern, B. S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., Lewison, R. et al. 2013. Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: assessing current pressures and opportunities. PLoS ONE, 8: e79889.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41: 225–233.

Nilsen, F. 2008. Påvisning av Emamectinresistente lakselus i Noreg. Norsk Fiskeoppdrett, 6: 58–60. (in Norwegian)

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., and De Meester, L. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Moleculer Ecology, 22: 5983–5999.

Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., Bennett, K. D. et al. 2012. Glacial survival of boreal trees in northern Scandinavia. Science, 335: 1083–1086.

Peakall, R., and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6: 288–295.

Peakall, R., and Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28: 2537–2539.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudoin, L., and Estoup, A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity, 95: 536–539.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, 155: 945–959.

Pollard, D. 2010. *Ctenolabrus rupestris*. *In* The IUCN Red List of Threatened Species 2010: e.T187751A8620934. http://dx.doi.org/10.2305/IUCN.UK.2010-4.RLTS.T187751A8620934.en (last accessed 25 October 2016).

Purcell, J. F. H., Cowen, R. K., Hughes, C. R., and Williams, D. A. 2006. Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. Proceedings of the Royal Society of London. Series B, Biological Sciences, 273: 1483–1490.

Quintela, M., Danielsen, E. A., Lopez, L., Barreiro, R., Svåsand, T., Knutsen, H., Skiftesvik, A. B. et al. 2016. Is the ballan wrasse (*Labrus bergylta*) two species? Genetic analysis reveals within-species divergence associated with plain and spotted morphotype frequencies. Integrative Zoology, 11: 162–172.

Quintela, M., Danielsen, E. A., Svåsand, T., Knutsen, H., Skiftesvik, A. B., and Glover, K. A. 2014. Isolation and characterization of twenty microsatellite loci for the ballan wrasse, *Labrus bergylta*. Conservation Genetics Resources, 6: 425–428.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rannala, B., and Mountain, J. L. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America, 94: 9197–9201.

Raymond, R., and Rousset, F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86: 248–249.

Riginos, C., Crandall, E. D., Liggins, L., Pongaerts, P., and Treml, E. A. 2016. Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. Current Zoology, 62: 581–601.

Rosenberg, M. S., and Anderson, C. D. 2011. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods in Ecology and Evolution, 2: 229–232.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources, 8: 103–106.

Robalo, J. I., Castilho, R., Francisco, S. M., Almada, F., Knutsen, H., Jorde, P. E., Pereira, A. M., and Almada, V. C. 2011. Northern refugia and recent expansion in the North Sea: the case of the wrasse *Symphodus melops* (Linnaeus, 1758). Ecology and Evolution, 2: 153–164.

Sá-Pinto, A., Branco, M. S., Alexandrino, P. B., Fontaine, M. C., and Baird, S. J. E. 2012. Barriers to gene flow in the marine environment: insights from two common intertidal limpet species of the Atlantic and Mediterranean. PLoS ONE, 7: e50330.

Selkoe, K. A., and Toonen, R. J. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Marine Ecology Progress Series, 436: 291–305.

Severance, E. G., and Karl, S. A. 2006. Contrasting population genetic structures of sympatric, mass-spawning Caribbean corals. Marine Biology, 150: 57–68.

Skiftesvik, A. B., Blom, G., Agnalt, A.-L., Durif, C. M. F., Browman, H. I., Bjelland, R. M., Harkestad, L. S. et al. 2014. Wrasse (Labridae) as cleaner fish in salmonid aquaculture – the Hardangerfjord as a case study. Marine Biology Research, 10: 289–300.

Skiftesvik, A. B., Durif, C. M. F., Bjelland, R. M., and Browman, H. I. 2015. Distribution and habitat preferences of five species of wrasse (Family Labridae) in a Norwegian fjord. ICES Journal of Marine Science, 73, 890–899.

Sundt, R. C., and Jørstad, K. E. 1993. Population genetic structure of wrasse used as cleanerfish in Atlantic salmon farming in Norway. ICES CM 1993/G: 30.

Sundt, R. C., and Jørstad, K. E. 1998. Genetic population structure of goldsinny wrasse, *Ctenolabrus rupestris* (L.), in Norway: implications for future management of parasite cleaners in the salmon farming industry. Fisheries Management and Ecology, 5: 291–302.

Svåsand, T., Karlsen, Ø., Kvamme, B. O., Stien, L. H., Taranger, G. L., and Boxaspen, K. K. (red.). 2016. Risikovurdering av norsk fiskeoppdrett 2016. Fisken og havet, særnr. 2-2016. (in Norwegian)

Teacher, A. G. F., André, C., Jonsson, P. R., and Merilä, J. 2013. Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. Evolutionary Applications, 6: 549–567.

Treasurer, J. W. 2012. Diseases of north European wrasse (Labridae) and possible interactions with cohabited farmed salmon, *Salmo salar* L. Journal of Fish Diseases, 35: 555–562.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4: 535–538.

Vikebø, F. B., Husebø, Å., Slotte, A., Stenevik, E. K., and Lien, V. S. 2010. Effect of hatching date, vertical distribution, and interannual variation in physical forcing on northward displacement and temperature conditions of Norwegian spring-spawning herring larvae. ICES Journal of Marine Science, 67: 1948–1956.

Wallace, I. S., Donald, K., Munro, L. A., Murray, W., Pert, C. C., Stagg, H., Hall, M. et al. 2015. A survey of wild marine fish identifies a potential origin of an outbreak of viral haemorrhagic septicaemia in wrasse, Labridae, used as cleaner fish on marine Atlantic salmon, *Salmo salar* L., farms. Journal of Fish Diseases, 38: 515–521.

Waples, R. S., and Gaggiotti, O. 2006. INVITED REVIEW: what is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology, 15: 1419–1439.

Woll, A., Skiftesvik, A. B., Solevåg, S. E., Hansen Aas, G., Bakke, S., and Bjelland, R. 2013. Fiskestørrelse og rømming fra laksemerd. Norsk Fiskeoppdrett August 2013 (in Norwegian).

Young, E. F., Belchier, M., Hauser, L., Horsburgh, G. J., Meredith, M. P., Murphy, E. J., Pascoal, S. et al. 2015. Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. Evolutionary Applications, 8: 486–509.

Handling editor: Lorenz Hauser