*ICES Journal of Marine Science*, 2023, **80**, 1103–1112
DOI: 10.1093/icesjms/fsad042
Advance access publication date: 22 March 2023
**Original Article**

ICES CIEM International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

# Genetic study reveals local differentiation persisting in the face of high connectivity and a genomic inversion likely linked with sexual antagonism in a common marine fish

**Eeva Jansson [1,*], Carl André [2], María Quintela [1], Kim T. Halvorsen [3], François Besnier [1], Fernando Ayllon [1], Ellika Faust [2], Halvor Knutsen [4], Åsa Strand [5], and Kevin A. Glover [1]**

[1]Institute of Marine Research, P. O. Box 1870 Nordnes, N-5817 Bergen, Norway
[2]Department of Marine Sciences-Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden
[3]Institute of Marine Research, Austevoll Research Station, Storebø N-5392, Norway
[4]Institute of Marine Research, Flødevigen, Nye Flødevigveien 20, N-4817 His, Norway
[5]IVL Swedish Environmental Research Institute, Kristineberg 566, 45178 Fiskebäckskil, Sweden

*Corresponding author:tel: +4755238500; e-mail: eeva.jansson@hi.no.

Sustainable harvest of wild populations requires knowledge of the underlying population structure. The focus of this study is on goldsinny wrasse (*Ctenolabrus rupestris*), a small marine fish inhabiting coastal waters of the north-eastern Atlantic. This species is caught in large numbers to serve as cleaner fish in salmonid aquaculture. We genotyped 2073 goldsinny wrasse from 43 sites along the Scandinavian coastline with 143 SNPs. Seven of the SNPs were linked and likely reside within a large genomic inversion dominated by one haplotype. The heterokaryotype of the putative inversion displayed sex-specific growth patterns, potentially resolving sexual antagonism for this trait. The unlinked 134 SNPs showed modest isolation-by-distance with samples from the northernmost locations showing highest divergence, whereas sites farther south were much more interconnected. Genetic divergence ($F_{ST}$) was highly variable among sites within regions, suggesting a varying degree of connectivity and local divergence. We conclude that despite a high degree of gene-flow mediated through pelagic dispersal in early life stages, regional and some local population structure remains due to limited adult movement in addition to other unidentified factors. Consequently, the species might be more vulnerable to local disturbances than previously anticipated.

**Keywords:** cleaner fish, *Ctenolabrus rupestris*, genetic patchiness, IBD, SNP, structural variant.

## Introduction

Efficient management of marine resources requires consideration of the underlying population structures but in reality, this knowledge is often lacking. The feasibility of using fishes from the wrasse family (*Labridae*) to control parasitic sea lice infestations on salmonid farms was first discovered in late 1980s (Darwall *et al.*, 1992; Sundt and Jørstad, 1998). Since then, both the wrasse fisheries and commercial aquaculture of these species have developed substantially in Norway (Halvorsen *et al.*, 2021a) and elsewhere (Bolton-Warberg, 2018). In Norway, three species of wrasses are predominantly used for delousing: goldsinny wrasse (*Ctenolabrus rupestris*), corkwing wrasse (*Symphodus melops*), and ballan wrasse (*Labrus bergylta*). About 18 million wild-caught wrasses are currently deployed in salmonid farms each year (Supplementary Figure S1). This equals the maximum national fishing quota set in 2018 to protect the wild fish stocks from overfishing. Seven to eight million goldsinny and corkwing wrasses are caught annually each, together with about two million ballan wrasses. Moreover, a few hundred thousand wrasses of Swedish origin are transported to salmonid farms in mid-Norway yearly and used before the local wrasse fishing season opens in July (Sandlund *et al.*, 2022).

The large-scale use of cleaner fish in the aquaculture industry was well-established before management guidelines for their sustainable use were considered (Halvorsen *et al.*, 2017). Fundamental knowledge of demography, abundance, and population genetics was largely lacking, likely due to their low prior commercial value (Darwall *et al.*, 1992). These small-bodied wrasses are often abundant in coastal marine ecosystems, and as intermediate predators are likely to serve an important role as both, prey and predator species (Olsen *et al.*, 2019). The distributions of wrasse overlap on rocky reefs but the species have largely varying life histories and biological traits (Darwall *et al.*, 1992) making generalizations difficult and management challenging. In past years, several studies have significantly increased our knowledge of cleaner wrasses' ecology (Halvorsen *et al.*, 2017; Halvorsen *et al.*, 2020) and genetics (Knutsen *et al.*, 2013; Jansson *et al.*, 2017; Seljestad *et al.*, 2020; Faust *et al.*, 2021), leading to better integrated-management advice (Halvorsen *et al.*, 2021a).

Goldsinny wrasse is the smallest of the cleaner fishes, grows slowly (Skiftesvik *et al.*, 2013), and rarely becomes larger than 16 cm (Halvorsen *et al.*, 2017) but can live up to ~20 years (Darwall *et al.*, 1992). It has a wide distribution in shallow waters of the north-east Atlantic stretching from North Africa to mid-Norway, into the Black, Mediterranean, and Baltic Seas (Pollard, 2010). Male and female goldsinny wrasse look similar, without clear sexual dimorphism but males tend to grow faster (Olsen *et al.*, 2019). Goldsinny wrasse are stationary during the breeding season (Halvorsen *et al.*, 2021b), when males defend small (<2m$^2$) territories (Hilldén, 1981; 1984)

**Received:** 26 October 2022; **Revised:** 28 February 2023; **Accepted:** 1 March 2023

into which they return every breeding season likely guided by their sense for magnetic fields (Cresci *et al.*, 2021). During winter months, goldsinny wrasse are inactive and hibernate in deeper waters (Hilldén, 1984). Unlike other cleaner wrasses, goldsinny wrasse is a broadcast-spawner with pelagic eggs (Hilldén, 1984), do not show parental care (like corkwing wrasse that build nests), or change sex (like protogynous ballan wrasse). These characteristics, together with the species' broad distribution (Halvorsen *et al.*, 2021a) and the wide depth coverage it inhabits, exceeding typical fishing depths (Halvorsen *et al.*, 2020) suggest that goldsinny wrasse is likely more resilient against human disturbances than other wrasse. However, Halvorsen *et al.* (2017) showed that intensive wrasse fisheries had a considerable impact on the target populations: the abundance of goldsinny wrasse was significantly less in harvested sites compared to control sites.

Previous studies of goldsinny wrasse have detected genetic differences between geographically distant populations (Jansson *et al.*, 2020), which is likely due to limited adult migration and passive drift leading to isolation-by-distance (IBD). However, adaptive differences between goldsinny wrasse populations were postulated in an allozyme study (Sundt and Jørstad, 1998) and with an outlier approach using genome-wide SNP data (Jansson *et al.*, 2020). As seen in corkwing wrasse (Faust *et al.*, 2021), large-scale translocation of goldsinny wrasse may cause human-mediated gene flow into aquaculture-intense areas and influence natural population genetic structure (Jansson *et al.*, 2017). In this study, we provide additional genetic information to support the development of goldsinny wrasse management strategies. We use a combination of extensive sampling and new markers to better resolve the genetic population structure of the goldsinny wrasse in Scandinavia. This was achieved by genotyping over 2000 individuals from this area with a set of putatively discriminatory SNPs developed by Jansson *et al.* (2020).

## Materials and methods

### Sampling

In all, 2147 goldsinny wrasse were collected from 2013 to 2017 at 43 locations along the Scandinavian coast (Figure 1). Of these, 1148 fish from 26 sites were collected from roughly evenly spaced locations, and formed the reference baseline in this study (called "reference dataset"; Supplementary Table S1). An additional 925 fish were included from 19 locations in mid-Norway in 2017 (Supplementary Table S1 and Supplementary Figure S2). This area was chosen for intensive sampling due to its high density of salmonid farms, high local fishing pressure, and import of wild wrasses from southern Scandinavia. Data including all 2073 fish are indicated as the "full dataset." Sampling was opportunistic and included bycatches from research cruises; thus, sample size per location and individual metadata varied (Supplementary Table S1, Supplementary Figure S3, and Supplementary Table S2).

### DNA isolation, genotyping, and data filtering

Genomic DNA was extracted from fin clips stored in absolute ethanol with the Qiagen DNeasy Blood & Tissue Kit. In all, 2147 goldsinny wrasse were genotyped for 173 SNPs using the Agena MassARRAY iPLEX Platform. Details of the marker development and selection, as well as genotyping procedure, appear in Jansson *et al.* (2020). Six SNP loci were genotyped twice to estimate the mean error rate for genotyping. Loci that did not produce clear clustering patterns were removed. Call rates were checked with *dartR* (Gruber *et al.*, 2018). Loci and individuals with ≥30% missing data were discarded. After filtering, 143 SNPs and 2073 individuals remained. This full dataset was then tested for locus independence with the *poppr* 2.9.1 (Kamvar *et al.*, 2014) by calculating the index of association over all loci in the dataset using 199 permutations, as well as per each locus pair. Locus pairs showing strong associations were scrutinized.

### Population genetic analysis

Both LD pruned and linked datasets were analysed, unless otherwise stated. Whether all 43 sampling sites or the reference dataset with 26 sites was used depended on the analysis.

#### *Haplotype reconstruction for linked loci*

The likely haplotype for each individual of the linked seven SNPs was reconstructed using the software *PHASE*, v2.1.1 (Stephens *et al.*, 2001). This method includes a Bayesian approach that can infer phase and reconstruct the most likely haplotypes. As each SNP is located on a different assembly contig, the order of the SNPs on the chromosome was unknown. Haplotype reconstruction was thus repeated once for each permutation of SNP. The permutation that minimized the estimated recombination rate in the whole window was accepted as the likely order of SNPs along the chromosome. The positioning of the recovered haplotypes was scrutinized against the observed population patterns for the linked dataset. Inferred haplotypes were used as a single locus dataset to estimate genetic differentiation between samples.

#### *Genetic variation and its division*

Expected ($H_E$) and observed ($H_O$) heterozygosities, allelic richness (*Ar*), and inbreeding ($F_{IS}$) were estimated with *diveRsity* (Keenan *et al.*, 2013). $F_{IS}$ 95% confidence intervals were calculated with 1000 bootstraps. Fit of loci and sampling sites to expected Hardy–Weinberg proportions was tested with exact tests with 1000 Markov Chain Monte Carlo (MCMC) permutations with *pegas* v.1.1 (Paradis, 2010). Pairwise $F_{ST}$-values were calculated with *StAMPP* (Pembleton *et al.*, 2013). *P*-values and 95% confidence intervals were estimated with 999 bootstraps and corrected for multiple comparisons with the false discovery rate Benjamini and Hochberg (1995) approach.

#### *IBD and population structure*

First, we tested for correlation between genetic and oceanographic distances using a Mantel approach in *vegan* 2.5-7 (Oksanen *et al.*, 2020) with 999 permutations. Minimum waterway distances between all samples were estimated with the least path method in *marmap* 1.0.5 (Pante and Simon-Bouhet, 2013) in R 4.1.2 (https://www.R-project.org/). Corresponding Euclidean genetic distances were calculated using *gl.dit.pop()* function in *dartT*. Besides estimating the statistical significance of the IBD model as a whole, we divided the data pairs into several roughly equally long-distance classes. Significance was tested with Mantel test in *ecodist*, 2.0.7 (Goslee and Urban, 2007) using 500 permutations and 500 iterations to calculate the corresponding 95% confidence intervals. This Mantel correlogram checks whether there is a correlation between

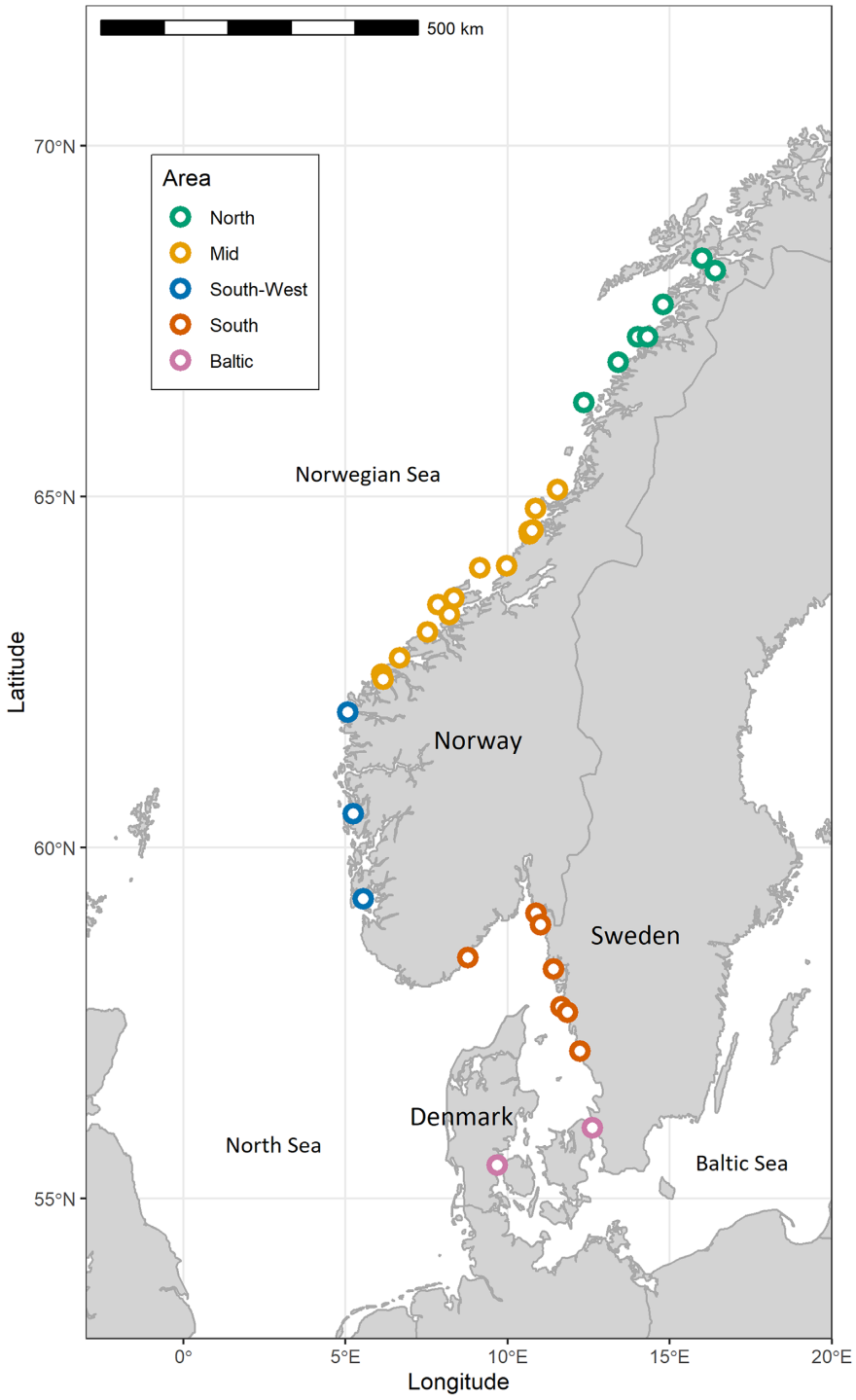


**Figure 1.** Sampling locations ($N = 43$) divided into five areas. Note that many circles in the "Mid" region of Norway overlap and are inseparable. The five geographic regions do not represent biological boundaries but were initially set for analytical exploration. See Supplementary Table S1 for more detailed information. See Supplementary Figure S2 for a more detailed map of sampling sites in mid-Norway.

two distance matrices by measuring the correlation between each class of distances.

We used two approaches to explore and visualize genetic differentiation among individuals. First, we performed a principal component analysis (PCA) with *ade4* (Chessel *et al.*, 2004). Without underlying assumptions on populations or their boundaries. We used *factoextra* 1.0.7 (https://CRAN.R-project.org/package=factoextra) to visualize individual contributions of SNPs on each principal component (PC) axis and to filter individual coordinates from the observed clusters based on their positioning. Next, we used Bayesian clustering in *STRUCTURE* 2.3.4 (Pritchard *et al.*, 2000) to search through predefined numbers of $K$ clusters to estimate the most likely number of genetic groups and to partition individuals into the $K$ groups. *STRUCTURE* runs were conducted with the default admixture model with correlated allele frequencies and with a priori location given. A total of 50000 MCMC repetitions were used after 20000 repeats were discarded as burn-in.

$K$ was explored for values from 1 to 10, with five iterations. To determine the optimal $K$, bar plots were inspected visually and runs analyzed with the StructureSelector (Li and Liu, 2018) that summarizes results as the optimal *Ln Pr(X|K)* from *STRUCTURE* and *ad hoc* summary statistic $\Delta K$ (Evanno *et al.*, 2005). MedMed, MedMean, MaxMed, and MaxMean were calculated as described by Puechmaille (2016). Values of $K$ from repeated runs were averaged with *CLUMPAK* (Kopelman *et al.*, 2015) with the LargeKGreedy algorithm and 2000 repeats. Besides reference and full datasets, the genetic structure was investigated for the linked dataset separately.

### Analysis of individual growth

The presence of genomic inversions has been linked with size-related adaptation (Zong *et al.*, 2021). Therefore, size-adjusted growth rates of individuals with different assumed inversion karyotypes were compared. This was done for 291 fish caught in mid-Norway from where most of the needed measurements were available (total length in mm and age; Supplementary Table S2). Sagittal otoliths were used for age determination and growth models were fitted to the length and age data using the typical parameterization of the von Bertalanffy growth equation with *FSA* 0.9.3 (Ogle *et al.*, 2022). This was done for both sexes separately after outlier observations were removed. We then extracted the residuals from the growth models, which were used as a standardized measure of individual growth. Linear models were then applied to test whether individual growth differed between karyotypes and applied the Tukey method for multiple comparison adjustments of $p$-values in pairwise contrasts using *emmeans* 1.8.0 (Russell, 2022).

## Results

### Splitting of data

The final dataset (Supplementary Table S2) consisted of 143 SNPs and 2073 fish (96.5%) that passed the quality controls. The index of association over all loci in the full dataset was highly significant ($p = 0.005$, $\bar{r}_d = 0.0067$) suggesting significant linkage between some of the markers (Supplementary Figure S4a). Of the 10153 pairwise $\bar{r}_d$ values, 113 (1.1%) were above the selected association threshold of 0.06 (Supplementary Figure S4b). Eleven SNPs (7.7%) showed strong and/or repeated associations: Two SNP pairs that showed exceptionally strong association (>0.75; pairs Gold-60: Gold-87 and Gold-165: Gold-224), and one SNP from each pair was discarded. The majority of the strongest associations were observed repeatedly among a group of loci: Gold-8, Gold-27, Gold-61, Gold-153, Gold-156, Gold-190, and Gold-206. The observed strong association between the seven SNP loci together with the observed diversity patterns, and population structures suggest that they are possibly within a chromosomal inversion. Therefore, in subsequent analyses, data were split into a non-linked 134 SNPs and linked seven SNPs datasets. Results for the entire combined dataset are included when applicable.

### Genetic variation and its division

Genetic diversity measured as expected heterozygosity ($H_E$) was on average ~0.38 (Supplementary Table S3). Observed heterozygosities ($H_O$) were significantly smaller across populations with sample sizes >20, and $H_O$ averaged ~0.34. This general heterozygote deficient was significant across sampling sites with $p \leq 0.001$. The pattern was not driven by just a few loci as a proportion of loci significantly deviating from HWE was averaged 9.2% (±4.8). The dataset of linked loci showed completely different patterns: Heterozygosity was clearly larger with $H_E$ averaging 0.446 (±0.015), $H_O$ averaging 0.451 (±0.042), and with no significant deviations from HWE in any sampling site. Also, allelic richness ($A_r$) was significantly larger for the LD dataset averaging 1.974 (±0.016) compared with the LD pruned dataset (1.877 ± 0.040).

Comparison of genetic differentiation among sampling sites revealed somewhat patchy patterns (Supplementary Tables S4a–c and Supplementary Figures S5a and b). Regardless of the dataset, the northernmost sampling site, Stefjorden (STE_2014) was generally the most divergent, and pairwise $F_{ST}$ were greatest at ~0.06. Other sampling sites in the area north (Supplementary Table S1) were significantly differentiated from the sites farther south (Supplementary Table S4a): Mean divergence between all sites including area north vs. rest of the sites was 0.018 (±0.011), whereas for all the rest of the pairwise comparisons it was 0.005 (±0.004).

### IBD and population structure

For all datasets, an association between the pairwise genetic divergence among sampling locations and their associated oceanographic distances was observed. However, the correlations were weak, not strictly linear, and the patterns varied between the datasets (Figure 2). Up to ~1000 km, a positive spatial autocorrelation was detected, whereas over longer distances, the relationship was significantly negative. The distinction of the northernmost sites was high regardless of the dataset analysed, dragging the slope upwards (Supplementary Figures S6b and c).

Decomposing SNP genotype likelihoods through a PCA showed that the first PC contained 3.5% of the total genetic variance, separating the data into three clusters (Figure 3). There were no obvious spatial patterns among these clusters nor were they diverged between sexes (Supplementary Figures S7a and b). Closer inspection of the SNP contributions along the first axes revealed that the pattern was driven by the linked seven loci, and without them, only a slight separation of samples from area north remained (Supplementary Figures S7c and d and S8a and b), which defined the second PC axes with 1.9% of the total genetic variation.

*STRUCTURE* analysis revealed the same basic pattern as the PCA. When all 143 loci or just the seven linked SNPs were used, a dichotomous pattern appeared (Figure 3, Supplementary Figures S9a–c) supporting $K = 2$. All fish were divided with high assignment probability either into one group or another or equally into both with no obvious spatial differences. This was the case regardless of the dataset used and indicates that this partitioning was uppermost level of genetic structuring. As such a pattern matches well with a presence of genomic inversion (Huang *et al.*, 2020; Merot, 2020), the detected groups were named as karyotype AA, AB, or BB. As there is no complete reference genome for the species available, phasing of linked SNPs into haplogroups was resolved probabilistically.

For the remaining 134 unlinked loci, a north to south divergence was observed (Figure 3). $K = 2$ and $K = 3$ were both supported, and showed the same differentiation of the northernmost sampling sites. Moreover, samples from northern parts of

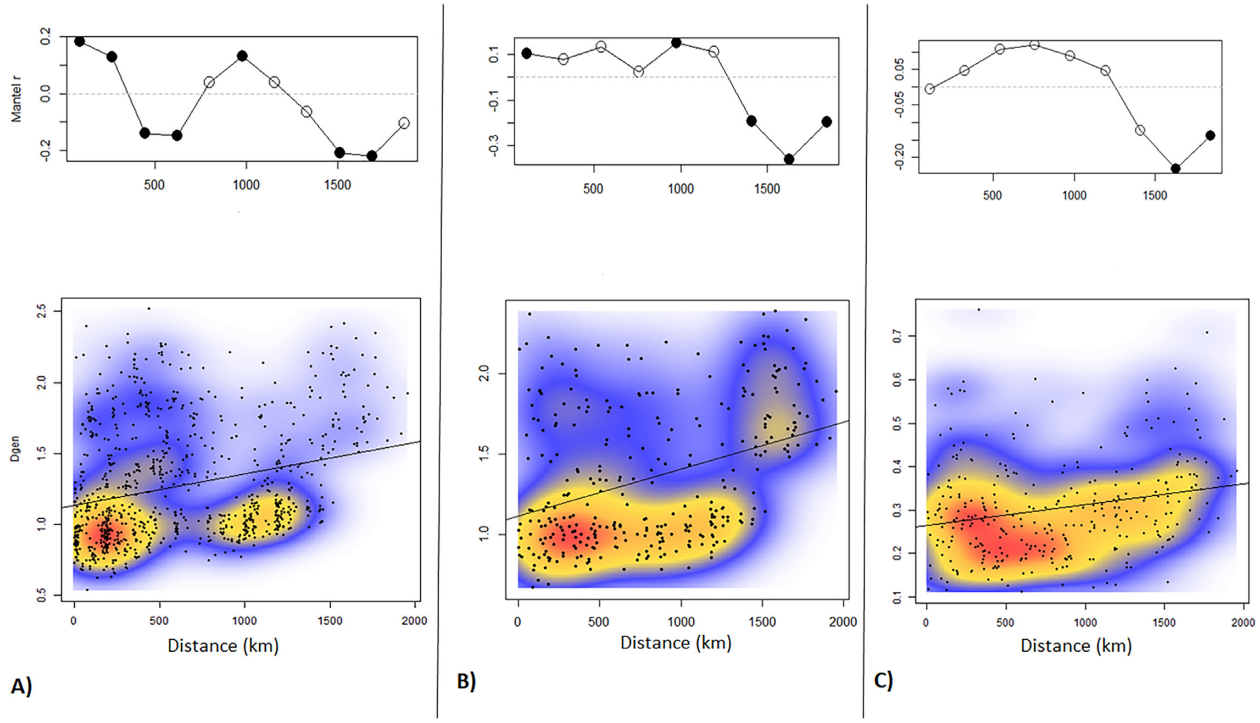


**Figure 2.** Relationship between oceanographic distance (km) and genetic differentiation between goldsinny wrasse sampling sites. Upper figures show Mantel correlograms where black dots indicate statistical significance for that distance class. Positive significant values indicate a positive autocorrelation, whereas significant negative values have the opposite interpretation. Figures below show corresponding IBD plots for the (A) pruned full dataset (134 SNPs and 43 sites), (B) pruned reference dataset (134 SNPs and 26 sites), and (C) for the linked dataset of the reference sites (seven SNPs and 26 sites). Dgen stands for Euclidean genetic distances. Colour contours indicate local kernel density estimates, where higher densities are shown by increasing degrees of red. Lines show the least-squares linear relationship between parameters. Please note the scale differences between plots.

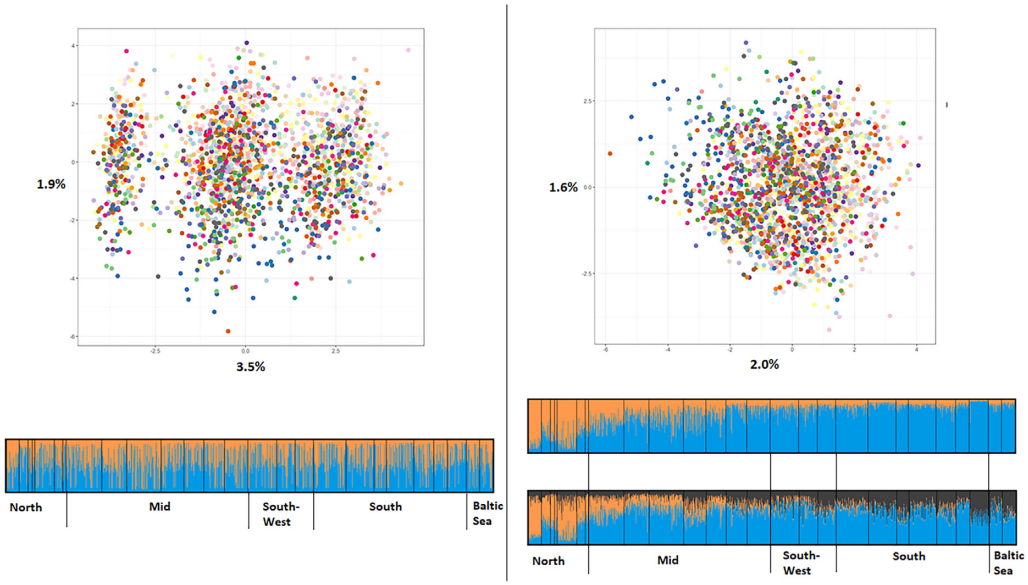


**Figure 3.** Structure of goldsinny wrasse populations based on PCA and Bayesian clustering method. Plots on left are for the dataset including linked SNPs, and those on right for the LD-pruned dataset. Figures above show PCA plots (two first axes) including all 2073 fish in 43 sampling sites shown in different colours. Figures below show *STRUCTURE* bar plots for the 26 reference sites arranged from north to south with individual assignments of fish into most likely genetic *K* clusters (2 for dataset including linked loci and 2/3 for LD pruned dataset). Results for other datasets are shown in Supplementary Figures 9a–c.

mid-Norway showed variable degree of individual assignment (between 10 and 70%) into this cluster typical for north. If an additional level of division was included, $K = 3$ (following the Puechmaille methods), some separation of southernmost sites was observed.

## Haplotype reconstruction and variability within karyotypes

Little change was observed between the reconstructed haplotypes after exploring SNP permutations, indicating a robust outcome from the haplotype reconstruction regardless

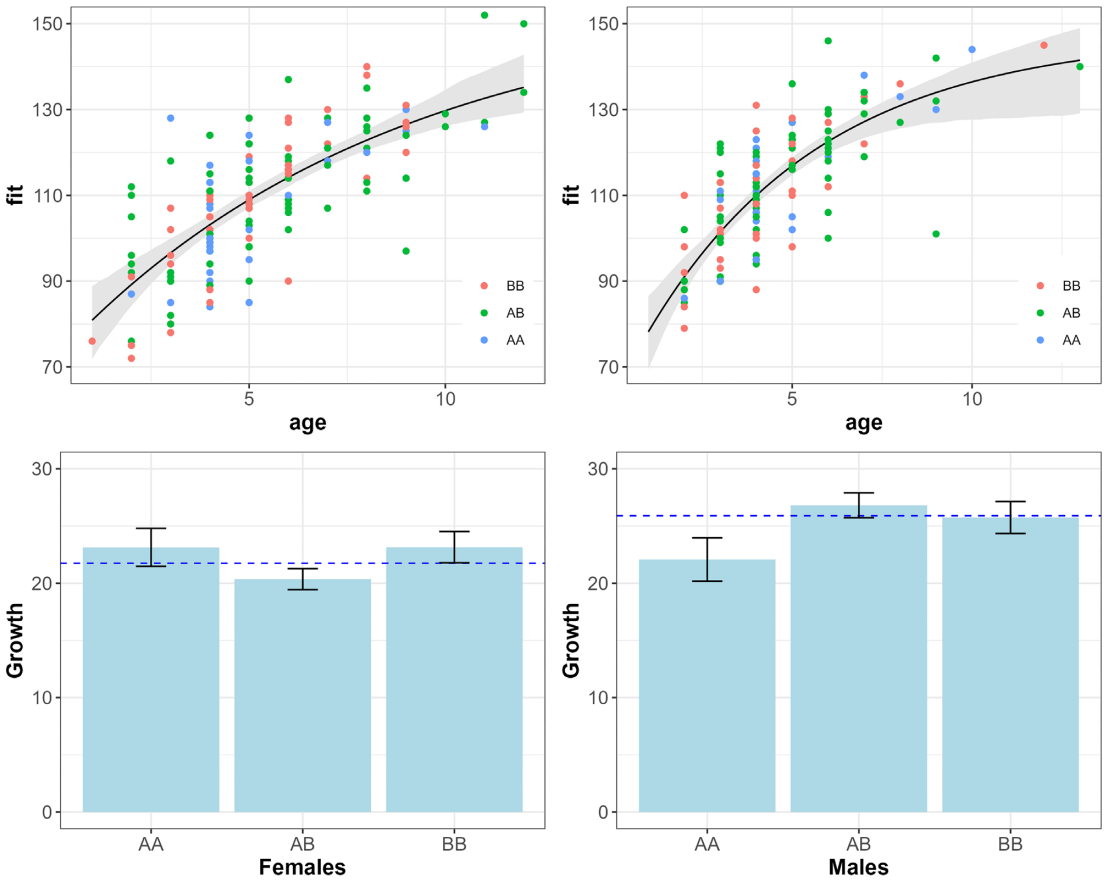


**Figure 4.** Growth of goldsinny wrasse by sex and karyotype. Figures for females are shown on left and for males on right. The figures on top show the fitted growth curves, and bar plots below them show measured growth rates (mm per year after first year) for each karyotype. Dotted blue line shows sex-specific means. On top of each bar, corresponding standard errors are shown. Analysis based on 139 and 152 males and females, respectively.

of SNP order. Therefore, the order minimizing the total recombination rate was kept as the most likely solution. Based on reconstruction, the seven linked SNPs generated 83 haplotypes (Supplementary Figure S10a) indicating recombination in this genomic region. This is consistent with the association between the loci, measured as pairwise $r\bar{r}_d$, was on average 0.233 s.d. (±0.154; Supplementary Figure S4b), far from complete linkage. An explanatory pattern for the observed linkage appeared with a comparison of individual haplotypes with the observed three genetic clusters (Figure 3 left, Supplementary Figure S10b): One genetic group—karyotype AA—consisted largely of one haplotype, hap10 (89.1%, Supplementary Table S5), and all homozygous individuals for this haplotype (271 fish) were in this group. A total of 15 reconstructed haplotypes could be assigned to karyotype AA, all of which closely resembled the reconstructed sequence of hap10 (Supplementary Figures S10a and b). Together with fewer haplotypes than for the other two karyotypes, AB and BB, the observed heterozygosity within AA was small, 0.083. The deficit of heterozygotes was significant ($p < 0.0001$). The estimated $F_{ST}$ between AA and BB for the seven loci was 0.542, between AA and AB 0.186, and 0.135 between AB and BB.

Fish with heterokaryotype AB had high diversity, $H_O = 0.661$, and many haplotypes (76). There was a significant homozygote deficiency ($p < 0.0001$) in this group. In total 37.3% of the AB karyotypes had hap10. The second homokaryotype, BB, showed statistically significant heterozygote deficit but with greater variability than AA with $H_O = 0.293$ and with 51 haplotypes. Hap10 was not detected for any BB fish.

### *Distribution of karyotypes and their association with growth rate*

No differences appeared in comparisons of the distribution of the three karyotypes between regions, sexes, and young and adult fish (Supplementary Table S5). Next, we investigated if growth rate (measured in mm per year) was associated with karyotype. Due to significant regional differences in size (Supplementary Figure S11a), comparisons were made only for fish from mid-Norway for which we had most data (139 males and 152 females). Females and males did not differ in size (Supplementary Figure S11b), but males grew faster in general (Supplementary Figure 12a; $F = 15.384$, $p < 0.001$). We detected a sexual difference between karyotypes (Figure 4): Among fish with karyotype AB, males grew faster, whereas in females the opposite was true (Supplementary Figure S12b). The difference between karyotypes was significant in females for AB and BB ($p = 0.014$), and indicative in comparison of AA and BB ($p = 0.057$), but not in other comparisons ($p > 0.1$).

## Discussion

### Large-scale genetic patterns reveal high general connectivity

We investigated the genetic population structure of goldsinny wrasse, an abundant mesopredatory fish along the rocky

shoreline of the northeast Atlantic. Our results show population structure on multiple levels and align with earlier genetic studies (Sundt and Jørstad, 1998; Jansson *et al.*, 2017; Jansson *et al.*, 2020). However, while the IBD identified here is in accordance with observations by Jansson *et al.* (2017), the pattern was weaker ($r \leq 0.361$), and not simply linear throughout the sampling range. We observed a greatest divergence for the northernmost sites, which impacted the overall pattern greatly (Supplementary Figures S6b and c) and masked weaker patterns on a local scale. The relationship between genetic and geographic distances was weaker but strongly supported ($R^2 = 0.192$, $p << 0.001$) when samples from the north were omitted. Therefore, we conclude that the populations in northern Norway are somewhat isolated, whereas the rest of the Scandinavian goldsinny wrasse in mid, southwestern, and southern Scandinavia show greater connectivity albeit still with IBD. Strong self-retention of drifting eggs and/or larvae in northern Norway is likely an important factor (Jansson *et al.*, 2017), but other contributing factors are discussed below.

Within Scandinavia, wrasses are currently expanding northward (Halvorsen *et al.*, 2021a) presumably due to changing climate. Populations in peripheral areas are often subject to founder effects and low effective population sizes (Dupoué *et al.*, 2020), with greater genetic drift. They may also be subjected to different evolutionary forces as they are expanding out of their natural range. Moreover, human-mediated gene flow via translocation of cleaner fish into fish farms between distant areas has been shown to genetically alter wild corkwing wrasse populations (Faust *et al.*, 2021). Millions of goldsinny wrasse are similarly caught and transported great distances to salmonid farms each year (Halvorsen *et al.*, 2021a; Sandlund *et al.*, 2022). This practice has been on-going since the 1990s (Sundt and Jørstad, 1998), with limited regulations (Sandlund *et al.*, 2022). Goldsinny wrasse are small, and can escape from sea cages. Consequently, this species has likely been exposed to a considerable human-mediated translocation over an extended period of time (>10 generations; Darwall *et al.*, 1992). An earlier study concluded that this possibly impacted the wild populations as illustrated by lower-than-expected level of genetic differentiation between southern and mid-Norway (Jansson *et al.*, 2017). Jointly these mid-Norwegian sampling sites fell below the expected linear IBD (Supplementary Figure S6c), suggesting greater connectivity to southern sites. Due to the low level of differentiation observed in this species, effects of translocation are difficult to disentangle from natural connectivity, however. This is exacerbated by the lack of baseline samples prior to any translocations. Unmonitored translocations represent a serious threat to biodiversity in general (Laikre *et al.*, 2010), and this practise is now declining in Norway (Halvorsen *et al.*, 2021a; Sandlund *et al.*, 2022) possibly caused by increased media attention in recent years.

## Metapopulation-like dynamics likely behind the patchy local patterns

Despite strong connectivity, we uncovered genetic divergence among most of the sampling sites (Supplementary Table S4a and b). Contrast to Jansson *et al.* (2017), the level of divergence was variable in all distance classes, also within areas (Supplementary Figure S6 and Supplementary Tables S4a–c). For example, in Flatanger divergence between samples separated by 3–4 years was observed, suggesting notable local genetic change over short time (~one generation), whereas other sites were strongly distinct in most comparisons. Significant heterozygosity deficit in all populations indicated mixing. These patterns are unexpected because large, well-connected marine population are expected to be stable (Le Corre *et al.*, 2015), and show only a weak spatial genetic differentiation (Lowe and Allendorf, 2010).

There are several possible explanations for these observations. First, the varying patterns of genetic divergence among populations indicate fluctuation in the degree of connectivity in space and time. Even relatively small $F_{ST}$ values may be associated with a notable amount of differentiation among the subpopulations (Lowe and Allendorf, 2010). The transition from demographic dependence to independence occurs in a region of strong connectivity where genetic methods have relatively little power (Waples and Gaggiotti, 2006). Adult goldsinny wrasse are territorial, long-lived, and stationary; hence, oceanographic processes such as larval retention influence local population structure (Le Corre *et al.*, 2015; Vendrami *et al.*, 2021). Substantial amounts of local retention have been suggested and that 90% of the eggs sink to the bottom (Hilldén, 1981). Therefore, demographic connectivity patterns may be local, and their genetic outcome—if the larval drift leads to gene flow—also likely affected by the population densities, fishing pressure, and translocations.

Knutsen *et al.* (2022) showed that habitat patchiness and larval dispersal likely determine the general connectivity patterns between populations of coastal fishes in this area. Unlike in their study, we did not observe discontinuity in the IBD pattern on the west coast of Norway but only in the far north. There was a great variation in connectivity between populations, with some being more isolated than others. This patchy connectivity, caused by differences in local demography resembles that of metapopulations, described for fish populations in habitats with source-sink dynamics and local extinctions, such as in coral reefs (Saenz-Agudelo *et al.*, 2012). Therefore, we hypothesize that goldsinny wrasses likely form local gene pools in areas where larval connectivity is weak. This indirectly supports the observation of localized growth patterns (Olsen *et al.*, 2019). A combination of allopatric, partially differentiated gene pools would also explain the observed significant heterozygosity deficits via the Wahlund's effect. The possibility of localized gene pools should be investigated further, as it would be an integral part of the species ecology and could have important management implications.

## Growth-related inversion is likely under sexually antagonistic selection

The significance of genomic rearrangements in enabling and maintaining local adaptation in the face of gene flow has gained recent support (Sodeland *et al.*, 2022). The most common type of rearrangement associated with adaptation is large inversions, where a segment of DNA is reversed end-to-end relative to a reference sequence (Merot *et al.*, 2020b). The inverted fragment is sheltered from recombination and inherited as a single haploblock. There are many indirect methods to search for and validate inversions using SNPs (Nowling *et al.*, 2020; Merot *et al.*, 2020b) that are based on a few features: physical linkage causing linkage disequilibrium, PCA showing three-cluster pattern corresponding with the three karyotypes,

as well as local heterozygosity excess due to suppressed recombination around the inversion.

We did not expect to find genomic rearrangements with the modest number of SNPs employed here. The used loci were originally selected to be in different contigs of >20 kb (Jansson *et al.*, 2020), and tested for independence. However, inversions can be large (Nowling *et al.*, 2020), and when enough individuals were sampled, the three-cluster PCA became visible (Figure 3) and overlayd other structures because samples clustered by their inversion karyotypes (Supplementary Figures S7a and b). In concurrence, a significant portion of the SNPs (7.7%) were strongly linked and that the observed genetic patterns matched the criteria for a chromosomal inversion (Supplementary Table 5; Figure 3; Supplementary Figure 4b). This suggests that there is at least one, likely large, chromosomal inversion in the goldsinny wrasse genome. We noticed that the observed patterns were almost exclusively connected to a single haplotype (hap10), and that the corresponding karyotype AA showed a likely suppression of recombination with a few haplotypes, all resembling the dominant haplotype.

Previous population genetic studies have shown diverse functions of inversions in regard to for example adaptation to environmental variation (Huang *et al.*, 2020), formation of distinct morphotypes (Sanchez-Donoso *et al.*, 2022), and variation in life-history characteristics (Merot *et al.*, 2020a). Here we show that the inversion karyotype is likely under strong balancing selection with similar frequencies detected across regions, sexes, and age classes (Supplementary Table S5, Figure 2c). This may be a case of antagonistic selection for divergent sex-specific growth patterns (Figure 4, Supplementary Figure S12b). Although possible causative genomic regions remain unknown, it is a common notion that sexual dimorphism evolves in response to sexual selection and/or natural selection arising from sex differences in reproductive roles (Cox and Calsbeek, 2009). Since goldsinny wrasse males are territorial and defend the same breeding patch throughout their lives (Hilldén, 1981), growing fast early in life could help them to gain high-quality territories, and thus maximize their life-time breeding success. This would be especially important with high population densities, thus suggesting possible density-dependent selection (Lorenzen and Enberg, 2002). Females, on the other hand, may gain higher fitness by investing surplus energy in egg mass rather than somatic growth.

### Conclusions with management implications

The central role of genetic diversity in sustainable management of marine populations is increasingly acknowledged as it can inform the managers of population structure, connectivity, local adaptation, and resilience. The use of cleaner fishes developed rapidly from almost non-existing to large-scale. The long-term impacts of the extensive fishing and translocations on these species populations, and local ecosystems are unknown. Studies are complicated because cleaner wrasses are ecologically diverse species and differ by their key life history attributes (Hilldén, 1984; Darwall *et al.*, 1992). Thus, there is an evident need of species-specific management in which genetic studies can provide invaluable insights.

In this study, we showed that goldsinny wrasse populations in the northern Scandinavia are distinct and somewhat isolated, and that there is a global IBD pattern. Populations also show a surprisingly high level of local divergence which supports the notion that goldsinny wrasse is an extremely stationary species with strong local retention. We suggest that the commonness of the species with high population densities mostly explains the observed large-scale population patterns, and that the significance of drifting eggs and larvae for connectivity may be less than previously thought (Jansson *et al.*, 2017). Local gene pools have important management implications, including greater susceptibility to fishing pressure, and stronger negative effect of translocations. The observed sex-specific, and possibly density-dependent growth patterns complicate this further (Eikeset *et al.*, 2016). Commercial harvest is currently regulated by minimum size, opening times, and maximum quota but not by other means such as local maximum take. Further genetic and physiological studies are needed to explore whether this assumed locality relates to adaptation and to better understand of the ecological functions of such adaptations.

## Acknowledgement

We thank Lars Måvik, Nils Marius Holm, Eva Farestveit, Lisbeth Sælemyr, Torkel Larsen, and Gunnar Didriksen for their contribution in the field. We acknowledge Fredrik Staven and Per Andersen for their assistance in sampling, and Stein Mortensen, Reidun Bjelland, and Anne Berit Skiftesvik for contributing samples. We are grateful to Editor Stewart Grant and two anonymous reviewers for their constructive comments and suggestions that notably improved this manuscript.

## Supplementary data

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

## Conflict of Interest statement

The authors declare that they have no conflict of interest.

## Funding

This study was funded by the Norwegian Ministry for Trade, Industry and Fisheries, the Swedish research council FORMAS and the European Regional Development Fund (Interreg Project "Margen II").

## Author contribution

K.A.G. and E.J. designed the study. E.J., M.Q., K.T.H., and Å.S. contributed in sampling. E.J. conducted the laboratory work and majority of the analyses but with significant input from F.B., K.T.H., F.A, M.Q., and E.F. E.J. wrote the article with notable contribution and help from all coauthors.

## Data availability

The dataset used in this study together with the collected individual metadata is included as Supplementary Table S2.

## References

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B, 57: 289–300.

Bolton-Warberg, M. 2018. An overview of cleaner fish use in Ireland. Journal of Fish Diseases, 41: 935–939.

Chessel, D., Dufour, A.-B., and Thioulouse, J. 2004. The ade4 package - I : one-table methods. R News, 7: 47–52.

Cox, R. M., and Calsbeek, R. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. The American Naturalist, 173: 176–187.

Cresci, A., Larsen, T., Halvorsen, K. T., Durif, C. M. F., Bjelland, R., Browman, H. I., and Skiftesvik, A. B. 2022. Goldsinny wrasse (*Ctenolabrus rupestris*) have a sex-dependent magnetic compass for maintaining site fidelity. Fisheries Oceanography, 31: 164–171.

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

Dupoué, A., Trochet, A., Richard, M., Sorlin, M., Guillon, M., Teulieres-Quillet, J., Vallé, C. *et al.* 2021. Genetic and demographic trends from rear to leading edge are explained by climate and forest cover in a cold-adapted ectotherm. Diversity and Distributions, 27: 267–281.

Eikeset, A. M., Dunlop, E. S., Heino, M., Storvik, G., Stenseth, N. C., and Dieckmann, U. 2016. Roles of density-dependent growth and life history evolution in accounting for fisheries-induced trait changes. Proceedings of the National Academy of Sciences, 113: 15030–15035.

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.

Faust, E., Jansson, E., André, C., Halvorsen, K. T., Dahle, G., Knutsen, H., Quintela, M. *et al.* 2021. Not that clean: aquaculture-mediated translocation of cleaner fish has led to hybridization on the northern edge of the species' range. Evolutionary Applications, 14: 1572–1587.

Goslee, S.C, and Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software, 22: 1–19.

Gruber, B., Unmack, P. J., Berry, O. F., and Georges, A. 2018. dartr: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources, 18: 691–699.

Halvorsen, K., Skiftesvik, A., Durif, C., Faust, E., Wennhage, H., André, C., Linnemann Rønfeldt, J. *et al.* 2021a. Towards a sustainable fishery and use of cleaner fish in salmonid aquaculture: challenges and opportunities. Report to Nordic Council of Ministers. 545. https://www.norden.org/en/publication/towards-sustainable-fishery-and-use-cleaner-fish-salmonid-aquaculture (last accessed 13 March 2023).

Halvorsen, K. T., Larsen, T., Browman, H. I., Durif, C., Aasen, N., Vollestad, L. A., Cresci, A. *et al.* 2021b. Movement patterns of temperate wrasses (*Labridae*) within a small Marine protected area. Journal of Fish Biology, 99: 1513–1518.

Halvorsen, K. T., Larsen, T., Sørdalen, T. K., Vøllestad, L. A., Knutsen, H., and Olsen, E. M. 2017. Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. Marine Biology Research, 13: 359–369.

Halvorsen, K. T., Sørdalen, T. K., Larsen, T., Browman, H. I., Rafoss, T., Albretsen, J., and Skiftesvik, A. B. 2020. Mind the depth: the vertical dimension of a small-scale coastal fishery shapes selection on species, size, and sex in wrasses. Marine and Coastal Fisheries, 12: 404–422.

Hilldén, N.-O. 1981. Territoriality and reproductive behaviour in the Goldsinny, Ctenolabrus rupestris L, Behavioural Processes, 6: 207–221.

Hilldén, N.-O. 1984. Behavioural Ecology of the Labrid Fishes (Teleostei: Labridae) at Tjärnö on the Swedish West Coast. University of Stockholm, Department of Zoology and Tjärnö Marine Biological Laboratory. 57 pp.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., and Rieseberg, L. H. 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. Molecular Ecology, 29: 2535–2549.

Jansson, E., Besnier, F., Malde, K., André, C., Dahle, G., and Glover, K. A. 2020. Genome wide analysis reveals genetic divergence between Goldsinny wrasse populations. BMC Genetics, 21: 118.

Jansson, E., Quintela, M., Dahle, G., Albretsen, J., Knutsen, H., André, C., Strand, Å. *et al.* 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.

Kamvar, Z. N., Tabima, J. F., and Grunwald, N. J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2: e281.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., Prodöhl, P. A., and O'Hara, R. B. 2013. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4: 782–788.

Knutsen, H., Catarino, D., Rogers, L., Sodeland, M., Mattingsdal, M., Jahnke, M., Hutchings, J. A. *et al.* 2022. Combining population genomics with demographic analyses highlights habitat patchiness and larval dispersal as determinants of connectivity in coastal fish species. Molecular Ecology, 31: 2562–2577.

Knutsen, H., Jorde, P. E., Gonzalez, 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: e67492.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources, 15: 1179–1191.

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

Le Corre, N., Johnsonn, L. E., Smith, G. K., and Guishard, F. 2015. Patterns and scales of connectivity: temporal stability and variation within a marine metapopulation. Ecology, 96: 2245–2256.

Li, Y. L., and Liu, J. X. 2018. StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources, 18: 176–177.

Lorenzen, K., and Enberg, K. 2002. Density-dependent growth as a key mechanism in the regulation of fish populations: evidence from among-population comparisons. Proceedings of the Royal Society of London B, 269: 49–54.

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

Merot, C. 2020. Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. Molecular Ecology, 29: 2513–2516.

Merot, C., Llaurens, V., Normandeau, E., Bernatchez, L., and Wellenreuther, M. 2020a. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. Nature Communications, 11: 670.

Merot, C., Oomen, R. A., Tigano, A., and Wellenreuther, M. 2020b. A roadmap for understanding the evolutionary significance of structural genomic variation. Trends in Ecology & Evolution, 35: 561–572.

Nowling, R. J., Manke, K. R., and Emrich, S. J. 2020. Detecting inversions with PCA in the presence of population structure. PLoS ONE, 15: e0240429.

Ogle, D. H., Doll, J. C., Wheeler, P., and Dinno, A. 2022. FSA: fisheries stock analysis. R package version 0.9.3, https://github.com/fishR-Core-Team/FSA (last accessed 13 March 2023).

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R. *et al.* 2020. vegan: community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan (last accessed 13 March 2023).

Olsen, E. M., Halvorsen, K. T., Larsen, T., Kuparinen, A., and Juanes, F. 2019. Potential for managing life history diversity in a commercially exploited intermediate predator, the goldsinny wrasse (*Ctenolabrus rupestris*). ICES Journal of Marine Science, 76: 410–417.

Pante, E., and Simon-Bouhet, B. 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS ONE,e73051, 8.

Paradis, E. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 26: 419–420.

Pembleton, L. W., Cogan, N. O., and Forster, J. W. 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13: 946–952.

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 10 May 2022).

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

Puechmaille, S. J. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources, 16: 608–627.

Russell, V. L. 2022. emmeans: estimated Marginal means, aka Least-Squares means. R package version 1.8.0. https://CRAN.R-project.org/package=emmeans (last accessed 13 March 2023).

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., and Planes, S. 2012. Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. Molecular Ecology, 21: 4695–4705.

Sanchez-Donoso, I., Ravagni, S., Rodriguez-Teijeiro, J. D., Christmas, M. J., Huang, Y., Maldonado-Linares, A., Puigcerver, M *et al.* 2022. Massive genome inversion drives coexistence of divergent morphs in common quails. Current Biology, 32: 462–469.e6.

Sandlund, N., Perez Garcia, E., Halvorsen, K. T., Jansson, E., Nedreaas, K., Skiftesvik, A. B. *et al.* (Eds). 2022. 10 - Risiko knyttet til miljøeffekter ved bruk av villfanget leppefisk i fiskeoppdrett. In Risikorapport norsk fiskeoppdrett 2022 – risikovurdering—Effekter på miljø og dyrevelferd i norsk fiskeoppdrett. Ed. by Grefsrud E.S Institute of Marine Research, (In Norwegian)

Seljestad, G. W., Quintela, M., Faust, E., Halvorsen, K. T., Besnier, F., Jansson, E., Dahle, G. *et al.* 2020. "A cleaner break": genetic divergence between geographic groups and sympatric phenotypes revealed in ballan wrasse (*Labrus bergylta*). Ecology and Evolution, 10: 6120–6135.

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.

Sodeland, M., Jentoft, S., Jorde, P. E., Mattingsdal, M., Albretsen, J., Kleiven, A. R., Synnes, A. W. *et al.* 2022. Stabilizing selection on Atlantic cod supergenes through a millennium of extensive exploitation. Proceedings of the National Academy of Sciences of the United States of America, 119: e2114904119.

Stephens, M., Smit, N. J., and Donnelly, P. 2001. A New statistical method for haplotype reconstruction from population data. The American Journal of Human Genetics, 68: 978–989.

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.

Vendrami, D. L. J., Peck, L. S., Clarck, M. S., Eldon, B., Meredith, M., and Hoffman, J. I. 2021. Sweepstake reproductive success and collective dispersal produce chaotic genetic patchiness in a broadcast spawner. Science Advances, 7: eabj4713.

Waples, R. S., and Gaggiotti, O. 2006. 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.

Zong, S. B., Li, Y. L., and Liu, J. X. 2021. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. Molecular Biology and Evolution, 38: 1317–1329.

*Handling Editor: W. Stewart Grant*