

# Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe

Morten Mattingsdal<sup>1</sup>  | Per Erik Jorde<sup>2</sup>  | Halvor Knutsen<sup>1,2</sup> | Sissel Jentoft<sup>3</sup>  |  
Nils Christian Stenseth<sup>1,3</sup> | Marte Sodeland<sup>1</sup> | Joana I. Robalo<sup>4</sup> |  
Michael M. Hansen<sup>5</sup>  | Carl André<sup>6</sup> | Enrique Blanco Gonzalez<sup>1,7</sup> 

<sup>1</sup>Department of Natural Sciences, Centre for Coastal Research, University of Agder, Kristiansand, Norway

<sup>2</sup>Institute of Marine Research, Flødevigen, Norway

<sup>3</sup>Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

<sup>4</sup>Marine and Environmental Sciences Centre, ISPA Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Lisboa, Portugal

<sup>5</sup>Department of Bioscience, Aarhus University, Aarhus C, Denmark

<sup>6</sup>Department of Marine Sciences-Tjärnö, Göteborg University, Strömstad, Sweden

<sup>7</sup>Norwegian College of Fishery Science, UiT The Arctic University of Norway, Tromsø, Norway

## Correspondence

Morten Mattingsdal, Centre for Coastal Research, Department of Natural Sciences, University of Agder, Kristiansand, Norway.  
Email: morten.mattingsdal@uia.no

## Funding information

Norges Forskningsråd, Grant/Award Number: 234328 and 280453; Swedish Research Council FORMAS; Centre for Marine Evolutionary Biology at the University of Gothenburg (CeMEB)

## Abstract

Understanding the biological processes involved in genetic differentiation and divergence between populations within species is a pivotal aim in evolutionary biology. One particular phenomenon that requires clarification is the maintenance of genetic barriers despite the high potential for gene flow in the marine environment. Such patterns have been attributed to limited dispersal or local adaptation, and to a lesser extent to the demographic history of the species. The corkwing wrasse (*Symphodus melops*) is an example of a marine fish species where regions of particular strong divergence are observed. One such genetic break occurred at a surprisingly small spatial scale ( $F_{ST} \sim 0.1$ ), over a short coastline (<60 km) in the North Sea-Skagerrak transition area in southwestern Norway. Here, we investigate the observed divergence and purported reproductive isolation using genome resequencing. Our results suggest that historical events during the post-glacial recolonization route can explain the present population structure of the corkwing wrasse in the northeast Atlantic. While the divergence across the break is strong, we detected ongoing gene flow between populations over the break suggesting recent contact or negative selection against hybrids. Moreover, we found few outlier loci and no clear genomic regions potentially being under selection. We concluded that neutral processes and random genetic drift e.g., due to founder events during colonization have shaped the population structure in this species in Northern Europe. Our findings underline the need to take into account the demographic process in studies of divergence processes.

## KEYWORDS

aquaculture, fish, landscape genetics, phylogeography, population genetics—empirical

## 1 | INTRODUCTION

Many marine species present a pelagic stage during their life cycle (Hauser & Carvalho, 2008), with high potential for dispersal

and gene flow. While such life cycles should generally result in panmixia and weak population divergence (Palsboll, Berube, & Allendorf, 2007), some species display genetic patterns of reproductive isolation indicative of barriers to random mating (Ravinet

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Molecular Ecology published by John Wiley & Sons Ltd

et al., 2017; Storfer, Murphy, Spear, Holderegger, & Waits, 2010). The observed patterns of divergence may be characterized by: (a) isolation-by-distance, where spatially separated individuals are less likely to encounter and hence mate; (b) isolation-by-adaptation, where locally adapted populations produce maladaptive or unviable hybrids when faced with gene flow, including Dobzhansky-Muller models of hybrid incompatibility; and (c) isolation-by-colonization, where the path and colonization history across the seascape and barriers may continue to restrict gene flow (Nadeau, Meirmans, Aitken, Ritland, & Isabel, 2016; Orsini, Vanoverbeke, Swillen, Mergeay, & De Meester, 2013; Spurgin, Illera, Jorgensen, Dawson, & Richardson, 2014).

After the last glacial maximum (~21 kya), serial colonization and founding events along the recolonization routes have shaped the biota on the northern hemisphere (Hewitt, 2000; Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998). Re-colonization has some common features among species, such as a general loss of genetic variation with increasing latitude, but the reconstructed histories tend to be quite complex and sometimes species-specific, involving glacial refugia, isolated pockets and secondary contact, as exemplified by terrestrial plants (Francois, Blum, Jakobsson, & Rosenberg, 2008; Kyrkjeeide, Stenøien, Flatberg, & Hassel, 2014; Petit et al., 2002). Similarly, many marine species also carry clear genetic signals of post-glacial range-expansions (Jenkins, Castilho, & Stevens, 2018). During the last glacial maximum, cold-adapted fish species are believed to have persisted in Northern Europe, while temperate fish species, such as the wrasses, found refuge in the Mediterranean and the surrounding coast of the Iberian Peninsula (Kettle, Morales-Muñiz, Roselló-Izquierdo, Heinrich, & Vøllestad, 2011).

The genetic makeup of several temperate wrasse fish species follow this classical pattern of loss of genetic variation with increasing latitude, as seen for ballan wrasse, *Labrus bergylta*, (Almada et al., 2017) and corkwing wrasse, *Symphodus melops* (Robalo et al., 2012). The corkwing wrasse has emerged as a particularly interesting case due to two substantial genetic breaks, across the North Sea ( $F_{ST} = 0.15$ ) and over a narrow coastal barrier with unsuitable sandy habitats (~60 km;  $F_{ST} = 0.11$ ) in southwestern Norway (Blanco Gonzalez, Knutsen, & Jorde, 2016). In addition, as the corkwing wrasse is currently exploited as “cleaner fish” in aquaculture (Blanco

Gonzalez & de Boer, 2017), this conspicuous genetic break demands clarification, in particular as individuals are translocated across the genetic break and members of the two populations can interbreed (Blanco Gonzalez et al., 2019; Faust, Halvorsen, Andersen, Knutsen, & Andre, 2018).

Genome-wide patterns of differentiation are particularly informative in elucidating if reproductive isolation is driven by directional selection (Feder & Nosil, 2010) or random genetic drift (Nielsen, 2005). Somewhat simplified, a classical strong selective sweep should display a local genomic signal, with “hitchhiking” neutral markers in proximity of the beneficial variant (Feder & Nosil, 2010). On the other hand, isolation-by-colonization should demonstrate a global and random pattern of genome-wide differentiation, a result of the stochastic fluctuations of variant frequencies imposed by for instance a founding event (Nielsen, 2005).

While population genetic methods are typically used to investigate patterns of population divergence, analyses using demographic inference to explicitly test different scenarios of divergence are rarely undertaken (Rougemont & Bernatchez, 2018). Here, we make use of whole genome resequencing methods to analyze the divergence between populations of corkwing wrasse in Northern Europe and to investigate demographic histories and putative patterns of reproductive isolation of this rocky shore marine fish.

## 2 | MATERIALS AND METHODS

### 2.1 | Samples and genotyping

Sixty-five corkwing wrasses were sampled from eight coastal locations from three regions: the British Isles, western and southern Scandinavia (Table 1). Samples from southern Norway were collected by beach seine, while those from the west coast of Norway, Sweden and the British Isles were collected by fish pots, as described in (Blanco Gonzalez et al., 2016). Muscle tissues were taken from fresh or frozen specimens and stored in 96% ethanol prior to DNA extraction. Total genomic DNA was extracted with the DNeasy kit (Qiagen) or the E.Z.N.A. Tissue DNA kit (Omega Bio-Tek) and resuspending the DNA in TE buffer. The extractions

**TABLE 1** Regional groupings of corkwing sampling locations by region, location, code, sampling year, latitude, longitude, sample size ( $n$ ) and number of variable single nucleotide polymorphisms (SNPs) per site (minor allele count > 1)

Region	Location	Code	Year	Latitude	Longitude	$n$	SNPs
British Isles	Ardtoe	ARD	2010	N 56.40	W 5.50	7	704,073
Western Scandinavia	Smøla	SM	2015	N 63.32	E 8.11	8	592,767
	Norheimsund	NH	2014	N 60.39	E 6.48	8	584,422
	Stavanger	ST	2015	N 59.01	E 5.56	8	597,438
Southern Scandinavia	Egersund	EG	2008	N 58.45	E 5.53	8	480,354
	Arendal	AR	2014	N 58.41	E 8.74	8	431,556
	Tvedestrand	TV	2010	N 58.62	E 9.06	8	440,905
	Gullmarsfjorden	GF	2008	N 58.18	E 11.32	10	463,654

were analyzed with Qubit (Thermo Fisher Scientific) for assessment of the DNA quality and concentration. After normalization to 1,200 ng with Qiagen EB buffer (10 mM Tris-cl; pH = 8.0) the samples were fragmented to ~350 bp using a Covaris S220 (Life Technologies). Library construction was performed using the Illumina TruSeq DNA PCR Free protocol and checked on Bioanalyser High sensitivity chip and TapeStation (both Agilent) followed by Kapa Biosystems qPCR assay for Illumina libraries quantification.

Whole-genome resequencing was conducted on the Illumina HiSeq platform, generating  $2 \times 125$  bp paired-end reads to an average depth of  $\sim 9.16\times$  per sample ( $595\times$  in total across the 65 sample libraries). The mean read insert size across samples was 347 (range: 246–404). Reads were mapped to the corkwing wrasse reference genome assembly (Mattingsdal et al., 2018) using BWA-MEM (v0.7.5a; Li & Durbin, 2009) followed by duplicate removal by Picard (<http://broadinstitute.github.io/picard/>). Single nucleotide polymorphisms (SNPs) were called across all samples with FREEBAYES (v1.0.2-33; Garrison & Marth, 2012), using the following quality control criteria: (a) quality  $>40$ ; (b) minimum and maximum read depth of  $\times 4$  and  $\times 30$ ; (c) maximum 5% missing genotypes; (d) minimum minor allele count of 3 (MAF  $>2\%$ ). Two data sets were made: (a) all SNPs with ancestral states and (b) a thinned data set keeping random SNPs equally spaced by 10,000 bp and excluding rare variants (MAF  $>2\%$ , thinned with “-bp-space 10,000”).

The ancestral allele states were inferred using whole-contig alignments between the corkwing and ballan wrasse (*L. bergylta*) genome assemblies (Lie et al., 2018; Mattingsdal et al., 2018) constructed by LAST (v923; Frith, Hamada, & Horton, 2010); both species are members of the *Labridae* family. First, the genomes were indexed specifying the “YASS” and “R11” options, optimizing for long and weak similarities and masking low-complexity regions. Then, a pairwise genome-wide alignment between corkwing- and ballan wrasses was made, setting minimum *E*-value to 0.05 and maximum matches per query position = 100. The “last-split” function was run twice to ensure 1-1 alignments. The multiple alignments were converted to bam format and SNP positions in the corkwing wrasse genome used to extract “genotypes” in the corkwing and ballan wrasse alignment using SAMTOOLS and BCFTOOLS (Li et al., 2009). The inferred ancestral states were manually controlled and PLINK v1.90b3.40 (Purcell et al., 2007) was used to annotate the ancestral state as the reference allele. Missing data were imputed and phased using BEAGLE default settings (Browning & Browning, 2013). To elucidate demographic relationships between the populations, we searched for identical-by-descent (IBD) haplotypes inferred by BEAGLE (Browning & Browning, 2013), which accounts for haplotype phase uncertainty.

## 2.2 | Population structure and admixture

Single nucleotide polymorphism-wise  $F_{ST}$  values between populations were calculated using the  $F_{ST}$  of Weir and Cockerham (1984)

implemented in VCFTOOLS (v0.1.13; Danecek et al., 2011). Patterns of population structure were investigated by Multidimensional scaling (MDS) analysis and inbreeding coefficients using PLINK v1.90b3.40 (Purcell et al., 2007). Proportion of ancestry for each individual,  $Q$ , for each putative ancestral population,  $K$ , was estimated using ADMIXTURE (v1.3.0; Alexander, Novembre, & Lange, 2009), making use of the integrated five-fold cross-validation scheme for 10 iterations each for  $K = 2-6$ , each using different random seed.

In an idealized diploid population, the identity-by-descent (IBD) haplotype lengths are exponentially distributed in an organism with a mean of  $1/(2 \times \text{generations})$  Morgans (Thompson, 2013). Therefore, IBD lengths and their distribution are of interest in inferring the ancestry of populations. Pairwise IBD segments between individuals were estimated by BEAGLE (v. 08 Jun17; Browning & Browning, 2013) using a minimum segment length of 0.01 cM, LOD score  $>3$ , overlap = 100 and ibdtime = 40. To assess the extent and length of IBD segment sharing between populations, a subset of seven random individuals from the most distant sampling locations (ARD, SM and GF; cf. Table 1 for sample information) were selected.

Gene flow and diversity between locations relative to geographical distance were estimated using EEMS (Petkova, Novembre, & Stephens, 2016), which models effective rates of gene flow using the pairwise dissimilarity matrix calculated by the embedded bed2diffs tool. The number of demes was set to 300 and several iterations were performed using default settings (100,000 burnin iterations, 200,000 MCMC iterations, 9,999 thinning interval) to ensure consistent and converging results. To detect admixture, the  $f_3$ -statistic (Reich, Thangaraj, Patterson, Price, & Singh, 2009), implemented in TREEMIX/THREEPOP (v 0.1; Pickrell & Pritchard, 2012), was used as a formal test for admixture between all population triplets using a block size of 200.

## 2.3 | Gene flow

Gene flow across the genetic break was estimated by calling discrete local ancestry using PCADMIX (Brisbin et al., 2012). Individuals from the ST and EG sites defined the admixed sample ( $N = 16$ ) and the remaining samples from Scandinavia used as “South” ( $N = 24$ ) and “West” ( $N = 16$ ) ancestors. Simplified, PCADMIX uses a PCA based algorithm of phased SNPs in a sliding window projecting the admixed samples using PCA loadings from the ancestral populations and calls local ancestry. Here, we used our previously phased dense SNPs data set and specified a fixed window size of 50 Kbp (-wMb 0.05). To reduce bias introduced by artificial breakpoints by our genome assembly, we used SNPs located on contigs  $>N50$  (461 Kbp). Fifty Kbp bins of southern origin were denoted by “0” and bins with a western origin “1”. We inserted a flag “9” to signify breaks introduced by a new contig. The R function “rle” was used to count the length of consecutive “0” and “1” for each haplotype for each contig (R Core Team, 2017).

## 2.4 | Demographic history

Demographic histories were estimated using two Markovian coalescent methods,  $\text{PSMC}$  (v 0.6.5-r67; Li & Durbin, 2011) and  $\text{SMC}++$  (v1.12.1; Terhorst, Kamm, & Song, 2017). In the  $\text{PSMC}$  analysis, the minimum read depth was increased to  $\times 6$  (Alex Buerkle & Gompert, 2013) and a maximum missing rate increased to 20% (Nadachowska-Brzyska, Burri, Smeds, & Ellegren, 2016). Population substructure can induce spurious signals of population bottlenecks and expansions (Mazet, Rodriguez, Grusea, Boitard, & Chikhi, 2016), so the analyses were performed separately for each of the three regions excluding possible admixed samples. In the  $\text{PSMC}$  analysis, one random individual from each of the most geographically distant locations were selected (ARD15, SM111 and GF01) using a similar approach to the one described in (Barth, Damerau, Matschiner, Jentoft, & Hanel, 2017), setting minimum and maximum read depth at six and 30 and base quality  $>30$ . Then the resulting fastq files were converted to  $\text{PSMC}$  input format specifying quality threshold  $>20$ .  $\text{PSMC}$  was run using default parameters, followed by 100 bootstraps.

For the  $\text{SMC}++$  analysis four random individuals from each region were combined into a composite likelihood for each population (South: GF01, GF49, TV69, TV70. West: NH61, NH60, SM111, SM114 and the British Isles: ARD20, ARD21, ARD18 and ARD15). Only SNPs situated on contigs  $>N50$  (461,652 bp) were included.  $\text{SMC}++$  was run using the options “thinning 50, unfold, knots 30”, specifying an unfolded frequency spectrum, reducing LD (approx. SNP density after thinning 1 SNP pr. 25,350 bp) and fixating the number of spline knots used in smoothing. For both  $\text{PSMC}$  and  $\text{SMC}++$  the site mutation rate was set to  $1 \times 10^{-8}$  and generation time to 3 years (Halvorsen et al., 2016, 2017; Uglen, Rosenqvist, & Wasslavik, 2000).

A diffusion approximation method was implemented,  $\text{DADI}$  (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), to determine the most likely population history scenario and its coalescent parameters. Four classical models were tested: (a) Strict Isolation (SI); (b) Isolation with Migration (IM); (c) Ancient Migration (AM); and (d) Secondary Contact (SC). The scenario obtaining the best Akaike information criterion (AIC) was deemed the most probable model. To reduce the effect of linkage disequilibrium (LD), we used the thinned data set. An optimization function (Optimize\_Functions.Optimize\_Routine) which sequentially refines the perturbation of parameters, was used (Portik et al., 2017). The optimization function included four rounds each with 10, 20, 30 and 40 replications, increasing maximum iterations (three, five, 10 and 15) and decreasing fold in parameter generation (3, 2, 2 and 1), resulting in 100 replications. We looped the aforementioned algorithm 10 times, yielding 1,000 local minima from the four models. The best model and its parameters were subjected to a goodness-of-fit test (Optimize\_Functions\_GOF) generating simulated parameters and these were used to assess the significance of the empirical parameters (Portik et al., 2017). Coalescent parameters were converted as follows: ancestral effective population size ( $N_e$ ) was calculated by  $N_e = \theta/4\mu$ , where  $\theta$  was the scaled population parameter,  $\mu$  was the mutation rate per

site and per generations, and  $l$  the length of analyzed sequence. Thinning the data set to one SNP per 10 kbp effectively reduced the length of the analyzed sequence by a factor of  $\sim 18$ , resulting in 35 Mbps. Migration was calculated as  $m = M/2N_e$  and time in years as  $t = 2TN_e \times g$ , using  $g = 3$  as generation time (Halvorsen et al., 2016, 2017; Uglen et al., 2000).

## 2.5 | Local patterns of differentiation and adaptation

Selective sweeps should display localized, elevated and linked  $F_{ST}$  values between populations (Sabeti et al., 2006). SNP-wise Weir and Cockerham's  $F_{ST}$  values were calculated by  $\text{vcftools}$  (v0.1.13; Danecek et al., 2011). In addition, the  $F_{ST}$  outlier test implemented in  $\text{BAYESCAN}$  was conducted (Foll & Gaggiotti, 2008) using default settings. Finally, a haplotype based test,  $\text{HAPFLK}$  (Fariello, Boitard, Naya, SanCristobal, & Servin, 2013), was also used. First, we calculated the Reynolds distance matrix using the thinned data set. No outgroups were defined, 20 local haplotype clusters ( $K = 20$ ) were specified and the  $\text{HAPFLK}$  statistic computed using 20 EM iterations ( $\text{nfit} = 20$ ). Statistical significance was determined through the script “scaling\_chi2\_hapflk.py”. To adjust for multiple testing, we set the false discovery rate (FDR) level to 5% using  $Q_{\text{VALUE/R}}$  (Storey, Bass, Dabney, & Robinson, 2019). Samples from the western and southern locations were grouped into their respective groups (South,  $N = 34$  and West,  $N = 24$ ) in all three tests.

## 3 | RESULTS

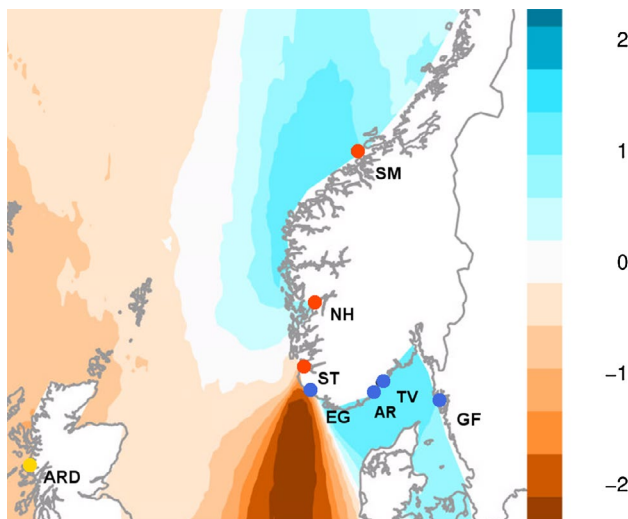
### 3.1 | Genotyping

The whole genome resequencing analysis generated a total of 3,048 million reads. Approximately 0.8% of these reads were duplicated and thus discarded. Of the remaining reads in the merged data set (3,024,360,818 reads), 97.19% mapped to the genome, and 93.27% were correctly paired. The mean depth of coverage per individual was  $\times 9.16$ . In total, 13.2 million sequence variants were detected, of which, 5.55 million had a quality metric  $>40$ . After applying min/max depth and maximum missing filters, 2.69 million variants were kept, of which 2.25 million SNPs were biallelic. We successfully inferred the ancestral state of 1,210,723 SNPs. Excluding rare SNPs, minor allele count (MAC)  $>3$ , resulted in 836,510 SNPs. We denominate this as the “all SNPs” data set. This highly dense data set was further reduced to keeping one SNP per 10 Kbp, using  $\text{vcftools}$  (“bp-thin 10,000”), yielding a reduced data set of 50,130 SNPs, denominated as the “thinned data set”. Due to a relatively low minimum read depth filter ( $\times 4$ ) it is likely that the proportion of heterozygous SNPs is underestimated, which can introduce a systematic error especially in windowed analyses which rely on breakpoints like IBD haplotypes (Meynert, Bicknell, Hurler, Jackson, & Taylor, 2013).

### 3.2 | Population structure and sequential loss of genetic variation

The number of SNPs within each sampling location suggests a pattern of sequential loss of diversity among regions, initially from the British Isles to western Scandinavia and followed by a further reduction to southern Scandinavia (Table 1). Of the 894 k SNPs (MAC >3 across all samples), ~704 k were found to be polymorphic (MAC >1) in the British Isles, ~590 k polymorphic in western Scandinavia (MAC >1) and ~450 k polymorphic in southern Scandinavia (MAC >1). We chose ARD ( $n = 7$ ), SM ( $n = 8$ ) and TV ( $n = 8$ ) as representative samples to count the overlap and unique SNPs between populations. Of the 704 k SNPs detected in the British Isles, 69% (485 k) were found in the West (SM) and 51% (360 k) in the South (TV). The proportion of unique SNPs in the British Isles, western and southern regions were 18%, 6% and 3%, respectively. A total of 327 k SNPs (39%) were found to be polymorphic in all three populations. The dramatic loss of genetic variation in Scandinavia as compared to the British Isles, especially in southern Scandinavia, was also revealed by the pairwise  $F_{ST}$  estimates (Table S1).

The simulation of effective migration surfaces (Figure 1) and MDS plot (Figure 2) identified three distinct groups corresponding to the British Isles, southern and western Scandinavia, as previously reported (Blanco Gonzalez et al., 2016; Knutsen et al., 2013), with some evidence of contact between the western and southern populations at the ST-EG site of south-western Norway. The ADMIXTURE analysis suggested  $K = 3$ , as the most likely number of ancestral populations with lowest mean cross validation of 0.368. The mean cross validation error for each  $K$ -value were,  $K2 = 0.378$ ,

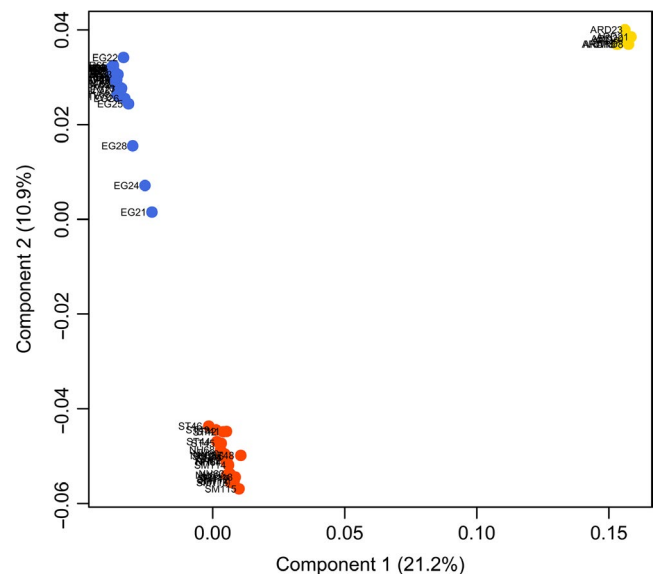


**FIGURE 1** Map showing sampling locations (see also Table 1) together with estimation of effective migration surfaces inferred by EEMS, where brown colour indicates a reduction and cyan colour indicates an increase in gene flow on the log<sub>10</sub> scale. Note that gene flow between differentiated populations would appear as a barrier (for example between ST and EG) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$K3 = 0.368$ ,  $K4 = 0.424$ ,  $K5 = 0.461$  and  $K6 = 0.471$  (for  $K2$  and  $K3$ , see Figure 3). The results from ADMIXTURE added further evidence for some gene flow across the contact zone between southern and western Scandinavian sample localities. The  $f_3$ -statistic test for admixture revealed that EG had the most negative  $f_3$ -statistic and  $Z$ -score in any combination with western (SM, NH, ST) and southern samples (AR, TV, GF), suggesting the EG population as a candidate admixed population in Scandinavia (mean:  $-0.0024$ ). The inbreeding coefficient ("plink -het") also revealed that the EG site was somewhat less homozygous compared to the other southern Scandinavian sites (Figure S1).

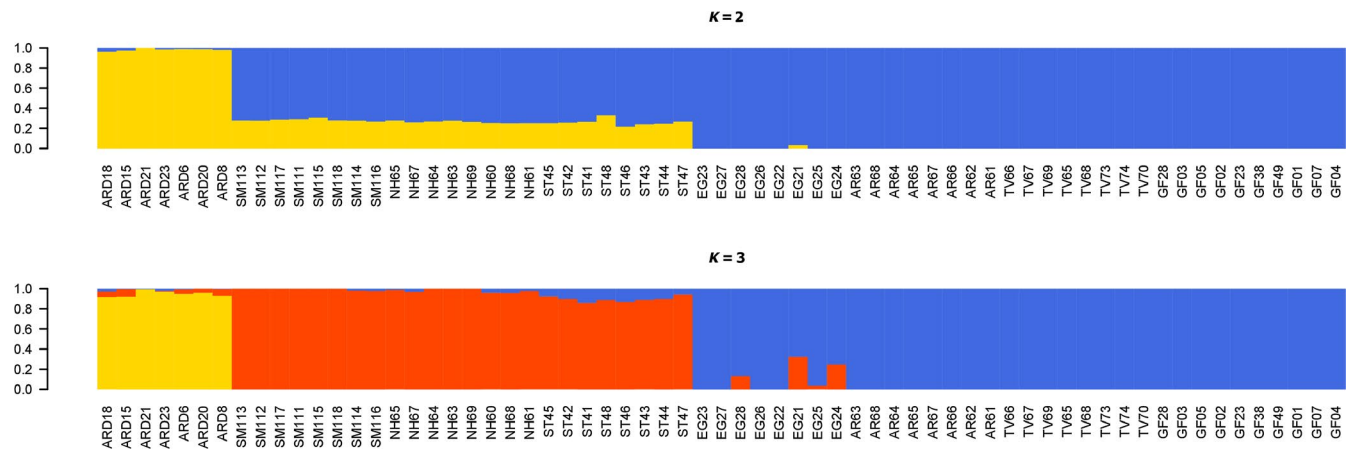
### 3.3 | Stochastic genome-wide differentiation

Searching for localized signals of differentiation and candidate regions of selection we explored the genome-wide pattern of variation between the two Scandinavian populations. The analysis revealed a strong and global genome-wide pattern of differentiation (Figure S2). Across the genome five regions showed  $F_{ST}$  values >0.9 and 32 regions  $F_{ST}$  >0.8. The haplotype based test, HAPFLK, returned a similar pattern but with more distinct candidate regions, albeit none passed the threshold for statistical significance ( $q$ -value <0.05). Testing for outliers between the western and southern populations, BAYESCAN results yielded two significant loci possibly under diversifying selection, SYMME\_00001686\_632632 and SYMME\_00023564\_399441 (Figure S3). The frequency of these two loci is 0.72 and 0.89 in the western population, and both loci were monomorphic in the southern population. The most differentiated SNPs can be informative in population discrimination and are listed in Table S2.



**FIGURE 2** Multidimensional scaling (MDS) plot using PLINK and the thinned data set (SNPs = 50,130). Individuals sampled in the British Isles (yellow), western Scandinavia (red) and southern Scandinavia (blue). SNP, single nucleotide polymorphism [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 3** ADMIXTURE results using the thinned single nucleotide polymorphism (SNP) data set (SNPs = 50,130).  $K = 3$  represents the most likely number of putative ancestral populations. Here, we show results at  $K = 2$  and  $K = 3$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 3.4 | Gene flow across the genetic break

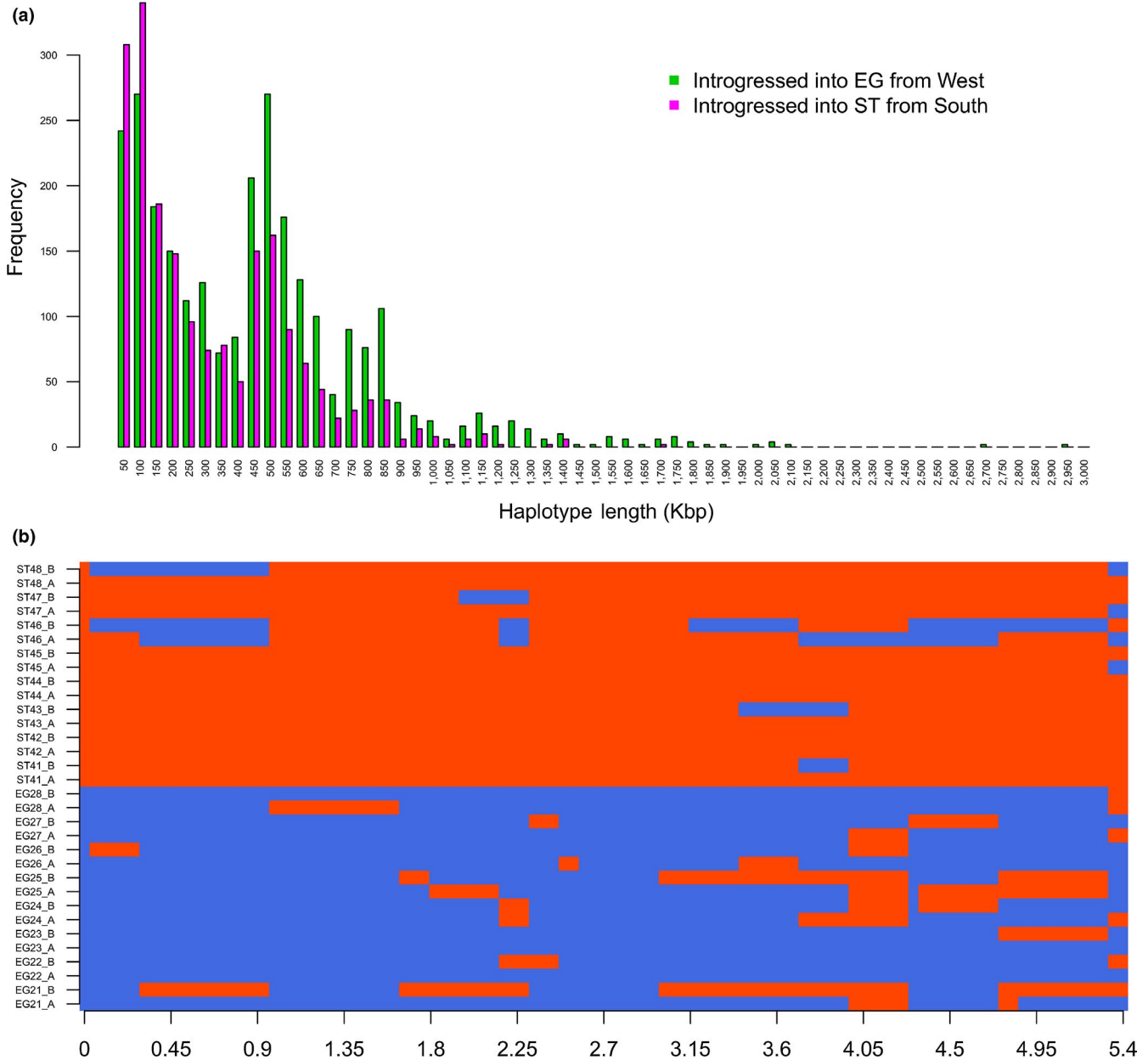
We used the default parameters in PCADMIX thereby removing SNPs in high LD ( $r^2 > .8$ ) and monomorphic SNPs in the ancestral samples. Of the 501,177 SNPs located on large contigs, 123,831 SNPs passed PCADMIX filters and were used for inference of local ancestry. They were located on 343 contigs, representing half the genome of the species (307 Mbp). Approximately 21.7 SNPs remained per bin of 50 Kbp ( $N$  bins = 5,695), a SNP quantity per bin recommended in the PCADMIX manual. A total of 27% of the genetic composition in the EG population was classified as “western” and 13% of the genetic composition in the ST sample was classified as “southern”. The overall mean length of consecutive western haplotypes in EG was 9.28 bins or 464 Kbp ( $SD = 7.2$ , median = 9, mean bins = 334) and southern haplotypes in ST was 6.34 bins or 317 Kbp ( $SD = 5.2$ , median = 5, mean bins = 246). The EG population has thus both longer and more regions of western origin than the ST population has of southern origin, clearly demonstrating introgression from the West into the South (Figure 4a). Some EG individuals appeared highly admixed (EG21 and EG24) with a 50.1% and 46.7% western ancestry, also suggested in the MDS plot (Figure 2) and admixture graph (Figure 3). Inspecting these individuals as potential F1 hybrids, revealed numerous heterozygous bins (~40%), but approximately ~60% of bins were homozygous from either southern or western ancestry, suggesting that these individuals were not F1 hybrids but instead admixed individuals. The ancestral calls for the ST and EG individuals can be obtained through <http://doi.org/10.6084/m9.figshare.9741641.v1>.

### 3.5 | Demographic history and founding events

The analysis of PSMC and SMC++ is based on the “all SNP” data set, while DADI analysis was conducted using the “thinned SNPs” data set. Demographic history estimated by PSMC suggests that all populations overall reduction of effective population size ( $N_e$ ) in all populations during the last ice age approximately 50 kya

(Figure 5a). The population in the British Isles experienced a more recent recovery (~5 kya), while the decline of  $N_e$  continued in the Scandinavian wrasses. There is also a distinct phase shift between the Scandinavian and British Isles population. The PSMC has limitations in inferring recent histories, as addressed by SMC++ (Terhorst et al., 2017). The inferred histories of SMC++ are remarkably similar (Figure 5b), suggesting that all populations have experienced a decline at different points in time, possibly reflecting sequential founding events. The most pronounced reduction of  $N_e$  was in southern Scandinavian, approximately 10 kya (blue line, Figure 5b). SMC++ offers flexibility, thus we experienced variation in the results (data not shown) depending on the options used. However, some patterns remained constant regardless of settings and included: (a) decline in all populations started approximately 30 kya, first in the British Isles, then in western and southern Scandinavia; (b) the magnitude of the decline was smallest in the British Isles, followed by West and finally largest in southern Scandinavia. SMC++ seems to present limitations to detect the two independent declines presumably experienced by the southern corkwing population, due to the algorithmic smoothing of inferred history. Even though SMC++ allows a folded frequency spectrum, we experienced a one order of magnitude improvement of the log likelihood by inferring the ancestral states of SNPs and specifying an unfolded spectrum during the simulations.

The isolation with migration model was the most likely scenario for the three comparisons analyzed in DADI. Among all 2D (two populations) models, the secondary contact projection yielded the best log likelihood and AIC statistic (Table S3). Nevertheless, we observed increased residuals on the rare frequency range, suggesting difficulties in modelling the loss of variation (Figure S5). We converted the coalescent values for the best model, the West and South Scandinavia secondary contact, resulting in an ancestral population ( $N_{ref}$ ) of 384. The size of the populations after the split were 3,980 and 1,275 for the West and South Scandinavia, respectively. The total time of divergence was  $T = 2 * N_{ref} * \text{generation time} * (T1 + T2) = 68,659$  years of which the first 63,102 were spent



**FIGURE 4** Introgression and local ancestry inferred by *PCADMIX* (Brisbin et al., 2012). (a) Displays the distribution of introgressed haplotype lengths in individuals from the sites across the genetic break (ST and EG). The distribution clearly shows larger haplotypes introgressed from the West into the EG site, including several large haplotypes >1,000 Kbp. (b) Displays the inferred local ancestry per haplotype using the largest contig as example region (SYMME\_00023145, 5.4 Mbp) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

in isolation while during the most recent 5,527 years the populations have experienced gene flow. The estimated migration rate was quite low (i.e.,  $m_{12}/2^*N_{ref}$ ) as the proportion of new migrants are  $4.1 \times 10^{-4}$  in the southern population and  $6.7 \times 10^{-4}$  in the western population (Table S3).

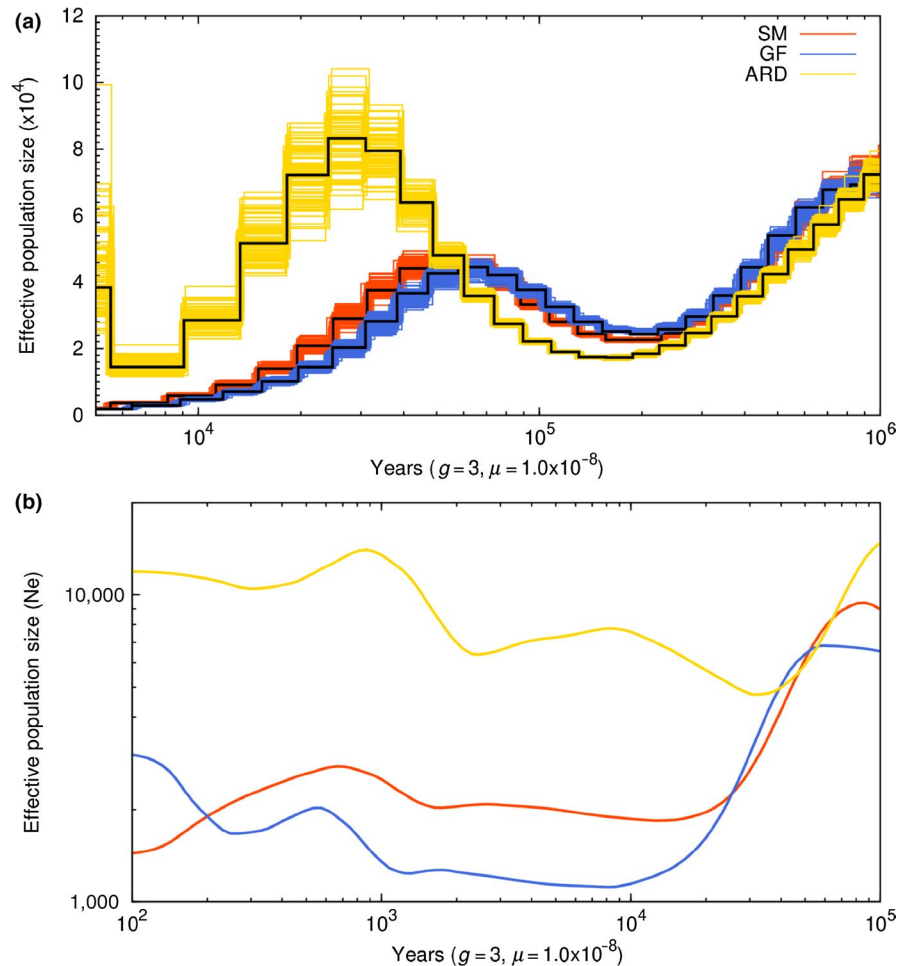
## 4 | DISCUSSION

Several marine species display cryptic population structure in parts of their range, and uncovering the underlying mechanisms behind

such genetic breaks are often nontrivial. Using whole genome sequencing and analyses of demographic history, we clarify the genetic underpinnings of reproductive isolation and differentiation of a marine fish, the corkwing wrasse. As a result of the cumulative evidence from our analyses, a clear picture of genetic drift has emerged as the dominant evolutionary force shaping contemporary patterns of population differentiation in corkwing wrasse.

The first line of evidence is the clear geographical pattern of global loss of genetic variation (number of polymorphic SNPs per sampling location, Table 1) and the increase in homozygosity from the British Isles, to western Scandinavia and finally to southern

**FIGURE 5** Demographic history inferred using the sequentially Markovian coalescent implemented in *psmc* and *smc++* using all single nucleotide polymorphisms (SNPs), a generation time of 3 years and a mutation rate of  $1 \times 10^{-8}$ . (a) Results from *psmc* with bootstraps using three individuals from the most distant sites. (b) Estimated histories in *smc++* using the composite likelihood of four individuals from each population. The yellow line represents the site from the British Isles, the red line corresponds to western Scandinavian samples, and the blue line comprises southern Scandinavian samples. Beware the differences on the axes between the top and bottom figures, as these two methods capture variation in effective population sized through different time scales [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



Scandinavia (Figure S1). The loss of SNPs is dramatic, as ~700 k SNPs detected in the British Isles are reduced to ~590 k (~16% less) SNPs in western Scandinavia with a further reduction to ~450 k (~35% less) SNPs in southern Scandinavia (Table 1), suggesting the direction and sequence of possible founding events to follow the British Isles-western Scandinavia-southern Scandinavia route, as previously suggested (Robalo et al., 2012). The pattern of genome-wide divergence ( $F_{ST}$  and  $H_{APFLK}$ ; Figure S2) did not show any fixed variation or clearly localized genomic regions that may suggest hard selective sweeps. Instead it showed a stochastic pattern of differentiation, probably imposed by strong drift (Figure 5), indicative of historical events shaping contemporary populations. Distinguishing between the genomic effects of bottlenecks with that of selective sweeps remains unresolved and are even discouraged (Pavlidis & Alachiotis, 2017; Poh, Domingues, Hoekstra, & Jensen, 2014). That being said, a polygenic model of adaptation remains a possibility although notoriously hard to detect and intrinsically difficult to distinguish from drift and population structure (Hollinger, Pennings, & Hermisson, 2019). The patterns of sequential loss of variation and lack of any missing fixed SNPs are also demonstrated in the site-frequency-spectra (Figure S4).

A second line of evidence is associated with the reduction of the effective population sizes in line with founding events detected

using the sequentially Markovian coalescent methods in *psmc* and *smc++* (Figure 5b). Our results suggest that corkwing wrasse colonized western Scandinavia about 11 kya, possibly from the British Islands. Based on the number of biallelic and heterozygous SNPs and inbreeding coefficients, Stavanger site (ST; Table 1, Figure 1) may be close to the point of entry into Scandinavia. Then, over a stretch of newly formed coastline, southern Scandinavia was subsequently (~10 Kya) colonized from the western Scandinavian population. The post-glacial colonization pattern in Scandinavia is similar to the colonization routes suggested for other marine species which depend on rocky habitats, such as seaweed, invertebrates and other fishes (Almada et al., 2017; Evankow et al., 2019; Hoarau, Coyer, Veldsink, Stam, & Olsen, 2007; Kettle et al., 2011; Maggs et al., 2008; Quintela et al., 2016). The colonization of Scandinavia ~10 kya, coincides with the deglaciation in western Norway (Stroeven et al., 2016). The demographic history date estimates inferred by the two Markovian approaches should be considered approximations, as the simulations rely on accurate generation time, mutation rates and sex ratio (Spence, Steinrücken, Terhorst, & Song, 2018). These values are intrinsically challenging for a species like corkwing wrasse, considering the variance in reproductive behaviour and generation time displayed by the species along the latitudinal gradient covered in this study (Halvorsen et al., 2017).



The fact that  $PSMC$  and  $SMC++$  do not adjust for periods of gene flow between populations and assumes clean population splits demands some care when interpreting changes in effective population sizes, and validation of findings using other methods are encouraged (Beichman, Phung, & Lohmueller, 2017). The method using diffusion approximations of the joint frequency spectrum implemented in  $DADI$  is frequently used to model complex scenarios of gene flow between populations (Rougemont et al., 2017; Tine et al., 2014). Here, we only tested simple scenarios, as more complicated models (Rougeux, Bernatchez, & Gagnaire, 2017) failed to converge and tended to produce artificial fits and parameters (data not shown). Using the site-frequency spectrum (SFS), the model which best fitted the empirical spectrum of all three comparisons was the secondary contact model (British Isles vs. western Scandinavia, British Isles vs. southern Scandinavia and western Scandinavia vs. southern Scandinavia, Table S3).

The third line of evidence is in the distribution of shared haplotypes (identical-by-decent) between the populations which corroborate the findings from the demographic history (Figure S6; Harris & Nielsen, 2013). Mean length of shared haplotypes was longer between the two Scandinavian populations, compared to the mean length between the British Isles and either Scandinavian population, suggesting a more recent split between the Scandinavian populations. The frequency of shared haplotypes also indicated the sequential loss of shared haplotypes and the direction of founding events.

Finally, we detected ongoing gene flow in both directions across the genetic break (Figure 4b). The contact was asymmetrical with increased gene flow from the West into the South. By using half the genome ( $>N50$ ) and bin size of 50 Kbp, we detected 1,568 bins of a total of 5,694 bins (27.5%) were of western origin in the EG population, and 785 bins in the ST population were of southern origin (13.8%).

Gene flow across genetic breaks can be an indicator of secondary contact after divergence (Sexton, McIntyre, Angert, & Rice, 2009). This strongly suggests that the genetic break is a hybrid zone with ongoing secondary contact after divergence. Findings from the IBD analysis suggests that the southern population descends from the western population. Our limited geographical sampling scheme does not, however, exclude the scenario of a ring-like colonization pattern, possibly surrounding the Norwegian Trench, where the southern population could descend from an unstudied population from the coastline in the southern parts of the North Sea. The persistence of the break remains intriguing and suggests that the contact is recent or actively selected against (Abbott et al., 2013).

In conclusion, our findings shed new light on the dynamics underlying the presence of two genetic breaks of this species in Northern Europe (Blanco Gonzalez et al., 2016; Knutsen et al., 2013; Robalo et al., 2012). It also serves to remind us that more simple scenarios involving sequential recolonization and associated founder events combined with secondary contact could underlie instances of strong genetic breaks, without having to

invoke more elaborate scenarios of selection and environmental adaptation (Hewitt, 1999; Ravinet et al., 2017; Schluter & Conte, 2009). Yet, while we could associate contemporary patterns of genetic differentiation to historical demographic events rather than adaptation, isolating mechanisms between western and southern Scandinavian populations still need further clarification, including a possible polygenic model of adaptation. In conclusion, corkwing wrasse could become an interesting future model for complementing and exploring the full span of possible dynamics that can lead to distinct contact zones, ranging from selectively neutral population history and structure to strong selection.

## ACKNOWLEDGEMENTS

The sequencing service was provided by the Norwegian Sequencing Centre ([www.sequencing.uio.no](http://www.sequencing.uio.no)), a national technology platform hosted by the University of Oslo supported by the Research Council of Norway and the Southeastern Regional Health Authorities. Computational work was performed on the Abel Supercomputing Cluster operated by the University of Oslo IT-department. This work was supported by funds from the Norwegian Research Council (grants 234328 and 280453), the Swedish Research Council FORMAS and by the Centre for Marine Evolutionary Biology at the University of Gothenburg (CeMEB).

## AUTHOR CONTRIBUTIONS

The study was conceived and designed by P.E.J, H.K, S.J and E.B.G. Sample collection were performed by P.E.J, H.K, C.A and E.B.G and genomic analyses performed by M.M. The manuscript was written by M.M with contributions from all authors. All authors read, revised and approved the manuscript.

## ORCID

Morten Mattingsdal  <https://orcid.org/0000-0003-4440-0324>  
 Per Erik Jorde  <https://orcid.org/0000-0001-5515-7257>  
 Sissel Jentoft  <https://orcid.org/0000-0001-8707-531X>  
 Michael M. Hansen  <https://orcid.org/0000-0001-5372-4828>  
 Enrique Blanco Gonzalez  <https://orcid.org/0000-0002-2631-2331>

## DATA AVAILABILITY STATEMENT

Sequence reads are available through NCBI sequence read archive by accession number PRJNA354496. SNPs (Mattingsdal, 2019) can be obtained through: [doi.org/10.6084/m9.figshare.7570907.v1](https://doi.org/10.6084/m9.figshare.7570907.v1).

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Alex Buerkle, C., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, 22(11), 3028–3035. <https://doi.org/10.1111/mec.12105>

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Almada, F., Francisco, S. M., Lima, C. S., FitzGerald, R., Mirimin, L., Villegas-Rios, D., ... Robalo, J. I. (2017). Historical gene flow constraints in a northeastern Atlantic fish: Phylogeography of the ballan wrasse *Labrus bergylta* across its distribution range. *Royal Society of Open Science*, 4(2), 160773. <https://doi.org/10.1098/rsos.160773>
- Barth, J. M. I., Damerou, M., Matschiner, M., Jentoft, S., & Hanel, R. (2017). Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations. *Genome Biology and Evolution*, 9(4), 1084–1098. <https://doi.org/10.1093/gbe/evx067>
- Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3*, 7(11), 3605–3620. <https://doi.org/10.1534/g3.117.300259>
- Blanco Gonzalez, E., & de Boer, F. (2017). The development of the Norwegian wrasse fishery and the use of wrasses as cleaner fish in the salmon aquaculture industry. *Fisheries Science*, 83(5), 661–670. <https://doi.org/10.1007/s12562-017-1110-4>
- Blanco Gonzalez, E., Espeland, S. H., Jentoft, S., Hansen, M. M., Robalo, J. I., Stenseth, N. C., & Jorde, P. E. (2019). Interbreeding between local and translocated populations of a cleaner fish in an experimental mesocosm predicts risk of disrupted local adaptation. *Ecology Evolution*, 9(11), 6665–6677. <https://doi.org/10.1002/ece3.5246>
- Blanco Gonzalez, E., Knutsen, H., & Jorde, P. E. (2016). Habitat discontinuities separate genetically divergent populations of a rocky shore marine fish. *PLoS ONE*, 11(10), e0163052. <https://doi.org/10.1371/journal.pone.0163052>
- Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., ... Bustamante, C. D. (2012). PCADMIX: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Human Biology*, 84(4), 343–364. <https://doi.org/10.3378/027.084.0401>
- Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, 194(2), 459–471. <https://doi.org/10.1534/genetics.113.150029>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Evankow, A., Christie, H., Hancke, K., Brysting, A. K., Junge, C., Fredriksen, S., & Thaulow, J. (2019). Genetic heterogeneity of two bioeconomically important kelp species along the Norwegian coast. *Conservation Genetics*, 20(3), 615–628. <https://doi.org/10.1007/s10592-019-01162-8>
- Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3), 929–941. <https://doi.org/10.1534/genetics.112.147231>
- Faust, E., Halvorsen, K. T., Andersen, P., Knutsen, H., & Andre, C. (2018). Cleaner fish escape salmon farms and hybridize with local wrasse populations. *Royal Society of Open Science*, 5(3), 171752. <https://doi.org/10.1098/rsos.171752>
- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, 64(6), 1729–1747. <https://doi.org/10.1111/j.1558-5646.2010.00943.x>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Francois, O., Blum, M. G., Jakobsson, M., & Rosenberg, N. A. (2008). Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics*, 4(5), e1000075. <https://doi.org/10.1371/journal.pgen.1000075>
- Frith, M. C., Hamada, M., & Horton, P. (2010). Parameters for accurate genome alignment. *BMC Bioinformatics*, 11, 80. <https://doi.org/10.1186/1471-2105-11-80>
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint, arXiv:1207.3907* ([q-bio.GN]).
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Halvorsen, K. T., Larsen, T., Sørtdalen, T. K., Vøllestad, L. A., Knutsen, H., & Olsen, E. M. (2017). Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. *Marine Biology Research*, 13(4), 359–369. <https://doi.org/10.1080/17451000.2016.1262042>
- Halvorsen, K. T., Sørtdalen, T. K., Durif, C., Knutsen, H., Olsen, E. M., Skiftesvik, A. B., ... Vøllestad, L. A. (2016). Male-biased sexual size dimorphism in the nest building corks wing wrasse (*Symphodus melops*): Implications for a size regulated fishery. *ICES Journal of Marine Science*, 73(10), 2586–2594. <https://doi.org/10.1093/icesjms/fsw135>
- Harris, K., & Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics*, 9(6), e1003521. <https://doi.org/10.1371/journal.pgen.1003521>
- Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4), 333–362. <https://doi.org/10.1111/j.1467-2979.2008.00299.x>
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68(1), 87–112. <https://doi.org/10.1006/bjil.1999.0332>
- Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, 405(6789), 907–913. <https://doi.org/10.1038/35016000>
- Hoarau, G., Coyer, J. A., Veldsink, J. H., Stam, W. T., & Olsen, J. L. (2007). Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology*, 16(17), 3606–3616. <https://doi.org/10.1111/j.1365-294X.2007.03408.x>
- Hollinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics*, 15(3), e1008035. <https://doi.org/10.1371/journal.pgen.1008035>
- Jenkins, T. L., Castilho, R., & Stevens, J. R. (2018). Meta-analysis of north-east Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ*, 6, e5684. <https://doi.org/10.7717/peerj.5684>
- Kettle, A. J., Morales-Muñiz, A., Roselló-Izquierdo, E., Heinrich, D., & Vøllestad, L. A. (2011). Refugia of marine fish in the northeast Atlantic during the last glacial maximum: Concordant assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the Past*, 7(1), 181–201. <https://doi.org/10.5194/cp-7-181-2011>
- Knutsen, H., Jorde, P. E., Gonzalez, E. B., Robalo, J., Albretsen, J., & Almada, V. (2013). Climate change and genetic structure of leading edge and rear end populations in a northwards shifting marine fish species, the corks wing wrasse (*Symphodus melops*). *PLoS ONE*, 8(6), e67492. <https://doi.org/10.1371/journal.pone.0067492>
- Kyrkjeeide, M. O., Stenøien, H. K., Flatberg, K. I., & Hassel, K. (2014). Glacial refugia and post-glacial colonization patterns in European bryophytes. *Lindbergia*, 2, 47–59. <https://doi.org/10.25227/linbg.01046>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMTOOLS. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lie, K. K., Tørresen, O. K., Solbakken, M. H., Rønnestad, I., Tooming-Klunderud, A., Nederbragt, A. J., ... Sæle, Ø. (2018). Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics*, 19(1), 186. <https://doi.org/10.1186/s12864-018-4570-8>
- Maggs, C. A., Castilho, R., Foltz, D., Henzler, C., Jolly, M. T., Kelly, J., ... Wares, J. (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, 89(11 Suppl), S108–S122. <https://doi.org/10.1890/08-0257.1>
- Mattingsdal, M. (2019). *SNPs from corkwing in Northern Europe*. Retrieved from [https://figshare.com/articles/SNPs\\_from\\_Corkwing\\_in\\_Northern\\_Europe/7570907/1](https://figshare.com/articles/SNPs_from_Corkwing_in_Northern_Europe/7570907/1)
- Mattingsdal, M., Jentoft, S., Tørresen, O. K., Knutsen, H., Hansen, M. M., Robalo, J. I., ... Gonzalez, E. B. (2018). A continuous genome assembly of the corkwing wrasse (*Symphodus melops*). *Genomics*, 110(6), 399–403. <https://doi.org/10.1016/j.ygeno.2018.04.009>
- Mazet, O., Rodriguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, 116(4), 362–371. <https://doi.org/10.1038/hdy.2015.104>
- Meynert, A. M., Bicknell, L. S., Hurler, M. E., Jackson, A. P., & Taylor, M. S. (2013). Quantifying single nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics*, 14, 195. <https://doi.org/10.1186/1471-2105-14-195>
- Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, 25(5), 1058–1072. <https://doi.org/10.1111/mec.13540>
- Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., & Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecology and Evolution*, 6(24), 8649–8664. <https://doi.org/10.1002/ece3.2550>
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999. <https://doi.org/10.1111/mec.12561>
- Palsboll, P. J., Berube, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology and Evolution*, 22(1), 11–16. <https://doi.org/10.1016/j.tree.2006.09.003>
- Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki*, 24, 7. <https://doi.org/10.1186/s40709-017-0064-0>
- Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., ... Kremer, A. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, 156(1), 49–74. [https://doi.org/10.1016/S0378-1127\(01\)00634-X](https://doi.org/10.1016/S0378-1127(01)00634-X)
- Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48(1), 94–100. <https://doi.org/10.1038/ng.3464>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Poh, Y. P., Domingues, V. S., Hoekstra, H. E., & Jensen, J. D. (2014). On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS ONE*, 9(11), e110579. <https://doi.org/10.1371/journal.pone.0110579>
- Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., ... Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26(19), 5245–5263. <https://doi.org/10.1111/mec.14266>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Quintela, M., Danielsen, E. A., Lopez, L., Barreiro, R., Svasand, T., Knutsen, H., ... Glover, K. A. (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(2), 162–172. <https://doi.org/10.1111/1749-4877.12186>
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489–494. <https://doi.org/10.1038/nature08365>
- Robalo, J. I., Castilho, R., Francisco, S. M., Almada, F., Knutsen, H., Jorde, P. E., ... Almada, V. C. (2012). Northern refugia and recent expansion in the North Sea: The case of the wrasse *Symphodus melops* (Linnaeus, 1758). *Ecology and Evolution*, 2(1), 153–164. <https://doi.org/10.1002/ece3.77>
- Rougemont, Q., & Bernatchez, L. (2018). The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. *Evolution*, 72(6), 1261–1277. <https://doi.org/10.1111/evo.13486>
- Rougemont, Q., Gagnaire, P. A., Perrier, C., Genthon, C., Besnard, A. L., Launey, S., & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26(1), 142–162. <https://doi.org/10.1111/mec.13664>
- Rougeux, C., Bernatchez, L., & Gagnaire, P. A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057–2074. <https://doi.org/10.1093/gbe/evx150>
- Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., ... Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, 312(5780), 1614–1620. <https://doi.org/10.1126/science.1124309>
- Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Suppl 1), 9955–9962. <https://doi.org/10.1073/pnas.0901264106>
- Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 415–436. <https://doi.org/10.1146/annurev.ecolsys.110308.120317>
- Spence, J. P., Steinrücken, M., Terhorst, J., & Song, Y. S. (2018). Inference of population history using coalescent HMMs: Review and outlook. *Current Opinion in Genetics & Development*, 53, 70–76. <https://doi.org/10.1016/j.gde.2018.07.002>

- Spurgin, L. G., Illera, J. C., Jorgensen, T. H., Dawson, D. A., & Richardson, D. S. (2014). Genetic and phenotypic divergence in an island bird: Isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23(5), 1028–1039. <https://doi.org/10.1111/mec.12672>
- Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2019). *QVALUE: Q-value estimation for false discovery rate control*. R package version 2.16.0.
- Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: Where are we now? *Molecular Ecology*, 19(17), 3496–3514. <https://doi.org/10.1111/j.1365-294X.2010.04691.x>
- Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., ... Caffee, M. W. (2016). Deglaciation of fennoscandia. *Quaternary Science Reviews*, 147, 91–121.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7(4), 453–464. <https://doi.org/10.1046/j.1365-294x.1998.00289.x>
- Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49(2), 303–309. <https://doi.org/10.1038/ng.3748>
- Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2), 301–326. <https://doi.org/10.1534/genetics.112.148825>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., ... Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5, 5770. <https://doi.org/10.1038/ncomms6770>
- Uglen, I., Rosenqvist, G., & Wasslavik, H. S. (2000). Phenotypic variation between dimorphic males in corkwing wrasse. *Journal of Fish Biology*, 57(1), 1–14. <https://doi.org/10.1111/j.1095-8649.2000.tb00771.x>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Mattingsdal M, Jorde PE, Knutsen H, et al. Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe. *Mol Ecol*. 2020;29:160–171. <https://doi.org/10.1111/mec.15310>