Received: 12 September 2018 | Revised: 2 November 2018 | Accepted: 5 November 2018

DOI: 10.1111/eva.12736

ORIGINAL ARTICLE

WILEY Evolutionary Applications Open Access

Detailed insights into pan-European population structure and inbreeding in wild and hatchery Pacific oysters (*Crassostrea gigas*) revealed by genome-wide SNP data

David L. J. Vendrami[1] | Ross D. Houston[2] | Karim Gharbi[3] | Luca Telesca[4,5] | Alejandro P. Gutierrez[2] | Helen Gurney-Smith[6] | Natsuki Hasegawa[7] | Pierre Boudry[8] | Joseph I. Hoffman[1,5]

[1]Department of Animal Behavior, Bielefeld University, Bielefeld, Germany

[2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

[3]Edinburgh Genomics, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK

[4]Department of Earth Sciences, University of Cambridge, Cambridge, UK

[5]British Antarctic Survey, High Cross, Cambridge, UK

[6]Department of Fisheries and Aquaculture, Vancouver Island University, Nanaimo, British Columbia, Canada

[7]National Research Institute of Aquaculture, Japan Fisheries Research Agency, Minami-Ise, Japan

[8]Ifremer, Laboratoire des Sciences de l'Environnement Marin (UBO/CNRS/IRD/Ifremer), Plouzané, France

Correspondence
David Vendrami, Department of Animal Behavior, Bielefeld University, Bielefeld, Germany.
Email: david.vendrami@student.unife.it

Funding information
Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/M026140/1; Natural Environment Research Council, Grant/Award Number: NE/P010695/1; European Union Marie Curie Seventh Framework Programme, Grant/Award Number: 605051; BBSRC Institute Strategic Program Grants, Grant/Award Number: BBS/E/D/20002172 and BBS/E/D/30002275

Abstract

Cultivated bivalves are important not only because of their economic value, but also due to their impacts on natural ecosystems. The Pacific oyster (*Crassostrea gigas*) is the world's most heavily cultivated shellfish species and has been introduced to all continents except Antarctica for aquaculture. We therefore used a medium-density single nucleotide polymorphism (SNP) array to investigate the genetic structure of this species in Europe, where it was introduced during the 1960s and has since become a prolific invader of coastal ecosystems across the continent. We analyzed 21,499 polymorphic SNPs in 232 individuals from 23 localities spanning a latitudinal cline from Portugal to Norway and including the source populations of Japan and Canada. We confirmed the results of previous studies by finding clear support for a southern and a northern group, with the former being indistinguishable from the source populations indicating the absence of a pronounced founder effect. We furthermore conducted a large-scale comparison of oysters sampled from the wild and from hatcheries to reveal substantial genetic differences including significantly higher levels of inbreeding in some but not all of the sampled hatchery cohorts. These findings were confirmed by a smaller but representative SNP dataset generated using restriction site-associated DNA sequencing. We therefore conclude that genomic approaches can generate increasingly detailed insights into the genetics of wild and hatchery produced Pacific oysters.

KEYWORDS

aquaculture, *Crassostrea gigas*, genetic structure, high-density genotyping array, inbreeding, Pacific oyster, restriction site-associated DNA (RAD) sequencing, single nucleotide polymorphism (SNP)

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.
© 2018 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Oysters are among the most economically important aquaculture species, with worldwide annual production exceeding 600,000 tonnes (FAO, http://www.fao.org). In particular, the Pacific cupped oyster (*Crassostrea gigas*), which is native to the Pacific coast of eastern Asia, was introduced into many countries worldwide for commercial cultivation. Starting in the 1960s, *C. gigas* was introduced into Europe to support oyster farming after severe declines of the two previously cultivated oyster species–the Portuguese oyster (*C. angulata*) and the flat oyster (*Ostrea edulis*, Grizel & Héral, 1991, Nehring, 1999, Wolff & Reise, 2002). Large quantities of seed as well as adult oysters were brought to France and the Netherlands from the Miyagi prefecture in Japan and from British Columbia in Canada, where *C. gigas* was also introduced from Japan in the 1920s (Quayle, 1988) and became quickly established in the wild. Concurrently, several small importations of less than a hundred individuals at a time also took place into the United Kingdom for hatchery propagation (Walne & Helm, 1979).

Subsequently, Pacific oysters produced in UK hatcheries were farmed in the German Wadden Sea (Reise, 1998) as well as in Denmark (Nehring, 2006), while oysters produced in French farms were transferred to various locations in the Mediterranean Sea (Grizel & Héral, 1991; Šegvić-Bubić et al., 2016) including southern Portugal, where hybridization with *C. angulata* is known to occur (Batista, Fonseca, Ruano, & Boudry, 2017; Huvet, Fabioux, McCombie, Lapegue, & Boudry, 2004). More recently, *C. gigas* also reached the southern coasts of Sweden and Norway (Troost, 2010), where it arrived as a consequence of both natural dispersal from Denmark and human-mediated translocation from British hatcheries (d'Auriac et al., 2017). Consequently, Pacific oysters have become widespread across the Atlantic and Mediterranean coasts of Europe, where they are responsible for major changes to coastal ecosystems (Troost, 2010) and are considered an invasive species (Goulletquer, Bachelet, Sauriau, & Noel, 2002).

Several studies have used genetic markers such as mitochondrial DNA and microsatellites to investigate the population structure of the Pacific oyster in Europe, mainly with a view toward understanding the history of invasion (d'Auriac et al., 2017; Faust et al., 2017; Lallias et al., 2015; Meistertzheim, Arnaud-Haond, Boudry, & Thébault, 2013; Rohfritsch et al., 2013) as well as interrelationships between wild populations and hatcheries (Kochmann, Carlsson, Crowe, & Mariani, 2012; Lallias et al., 2015; Moehler, Wegner, Reise, & Jacobsen, 2011). Many of these studies uncovered evidence for two distinct genetic clusters: one in southern Europe (subsequently referred to as the "southern group") that includes populations from the Mediterranean, Spain, France, the Netherlands, and the southwestern coast of England, and one in northern Europe (subsequently referred to as the "northern group") that consists of the remaining British, German, and Scandinavian populations (Lallias et al., 2015; Meistertzheim et al., 2013; Moehler et al., 2011; Rohfritsch et al., 2013). Furthermore, no genetic differences were found between the southern group and the source populations of Japan and British Columbia, suggesting that the original mass introduction may not have resulted in a founder effect (Rohfritsch et al., 2013). Additionally, the northern group was found to have lower genetic diversity, suggesting that it probably arose locally in Europe and more specifically in the UK as a consequence of repeated small introduction events that may have acted as bottlenecks due to hatchery propagation followed by genetic drift (Faust et al., 2017; Lallias et al., 2015).

Although previous studies have provided important insights into the population structure of Pacific oysters in Europe, many focused on local scales and, even though Rohfritsch et al. (2013), Lallias et al. (2015), and Faust et al. (2017) sampled extensively along the western Atlantic seaboard, there is still a need for more comprehensive studies encompassing the full latitudinal range of the species in Europe and including hatcheries from both Britain and France. Furthermore, classical approaches such as mitochondrial sequencing and microsatellite genotyping have limited power to detect population structure, especially over fine geographic scales where genetic differences may be too subtle to be captured with a handful of markers (Vendrami et al., 2017). By contrast, new genomic approaches capable of genotyping tens of thousands of single nucleotide polymorphisms (SNPs) have been proven to have far greater power to resolve genetic differences among populations (Morin, Luikart, & Wayne, 2004; Rašić, Filipović, Weeks, & Hoffmann, 2014) and therefore allow more in-depth studies of population genetic structure.

One of the most commonly used approaches for generating large SNP datasets for nonmodel organisms is to use genotyping by sequencing methods such as restriction site-associated DNA (RAD) sequencing (Baird et al., 2008), which allows concurrent SNP identification and genotyping via high-throughput sequencing of flanking regions of restriction enzyme digestion sites dispersed throughout the genome. These methods have democratized the study of population genomics but are not without their disadvantages (da Fonseca et al., 2016) such as the need for extensive bioinformatic processing, high rates of missing data, and the issue of uncertainty in genotype calling, which can affect downstream analyses (Shafer et al., 2017). A convenient alternative where available is therefore to use a medium- or high-density SNP array, in which the probe sequences of many tens or hundreds of thousands of SNPs are "printed" onto a slide against which the genomic DNA is hybridized. SNP arrays typically generate very high-quality data with relatively few missing genotypes, but they also have some downsides. Arguably, the most important of these is ascertainment bias, which occurs when not all of the genetic diversity present in a population can be captured by the array due to the use of a limited pool of individuals in the original SNP discovery phase (Lachance & Tishkoff, 2013).

Another drawback of small panels of nuclear markers like microsatellites is that their sampling variance is usually too large to accurately quantify variation in inbreeding (Balloux, Amos, & Coulson, 2004). This may be relevant to aquaculture because moderate to high levels of inbreeding have been shown to have detrimental effects on a variety of commercially important traits, such as harvest body size and larval growth, in several species, including turbot (Lyu,

Wu, Hu, & Wang, 2018), Pacific white shrimp (Moss, Arce, Otoshi, Doyle, & Moss, 2007), coho salmon (Gallardo, Garcıa, Lhorente, & Neira, 2004), and Pacific abalone (Deng, Liu, Zhang, & Guo, 2005) as well as in flat oysters (Lallias, Boudry, Lapegue, King, & Beaumont, 2010) and Pacific oysters (Evans, Matson, Brake, & Langdon, 2004; Launey & Hedgecock, 2001; Plough & Hedgecock, 2011). Hence, inbreeding depression could conceivably be a problem if aquaculture practices, such as the use of restricted numbers of parents as broodstock and/or the crossing of related individuals, lead to increased inbreeding in cultured populations (Norris, Bradley, & Cunningham, 1999; Taris, Batista, & Boudry, 2007).

Given the limited power of microsatellites to quantify inbreeding, the method of choice until recently has been to derive individual inbreeding coefficients (*f*) from deep pedigrees (Pemberton, 2008). However, pedigrees can be costly and time-consuming to construct and may also be unworkable for many aquaculture species due to their high fecundity and broadcast spawning life-histories. Fortunately, recent simulation (Kardos, Luikart, & Allendorf, 2015; Wang, 2016) and empirical (Hoffman et al., 2014; Huisman, Kruuk, Ellis, Clutton-Brock, & Pemberton, 2016) studies suggest that inbreeding can now be directly and accurately quantified from genomic data, with around ten thousand or more SNPs being preferable under most circumstances even to a high-quality pedigree. Consequently, the increasing availability of SNP arrays for non-model species provides an exciting new opportunity to elucidate how different aquaculture practices influence inbreeding, as well as to identify suitable sources of individuals for use as broodstock to establish effective management and breeding protocols.

Recently, Gutierrez et al. (2017) developed a medium-density combined species SNP array for Pacific and flat oysters (*Ostrea edulis*). Whole genome sequencing of pooled genomic samples from eight European *C. gigas* populations led to the discovery of 1.2 million putative SNPs, of which 40,625 were printed on the array and 27,697 were validated as being polymorphic and of high quality. This array represents an important resource for selective breeding programs as well as more generally for population genetic studies of oysters. We therefore used it to investigate population genetic structure and inbreeding in *C. gigas* sampled from wild European populations and hatcheries. Specifically, we genotyped 192 individuals from 13 populations spanning a European latitudinal cline from Portugal in the south to Norway in the north. We then incorporated existing data from Gutierrez et al. (2017) to generate a combined dataset of 273 individuals sampled from 23 populations, of which just over half were wild. Our results may be useful for designing exchanges

TABLE 1 Table of sampling locations including coordinates, origin classified as either wild or farmed, and the number of samples that were retained for analysis after quality control.

Population ID	Location	Latitude	Longitude	Origin	Sample size passing QC
POR	Faro, Portugal	37.002	−7.583	Wild	9
ITA	Ravenna, Italy	44.402	12.122	Wild	10
SPA	Santoña, Spain	43.426	−3.543	Wild	8
FRA	Brest, France	48.215	−4.462	Wild	10
ENG	Plymouth, UK	50.373	−3.441	Wild	9
NE1	Oosterschelde, Netherlands	51.608	3.91	Wild	12
NE2	Texel, Netherlands	53.001	4.474	Wild	10
IFR	Ifremer, France	48.351	−4.551	Hatchery	12
FH1	Hatchery 1, France	na	na	Hatchery	10
FH2	Hatchery 2, France	na	na	Hatchery	10
FH3	Hatchery 3, France	na	na	Hatchery	10
FH4	Hatchery 4, France	na	na	Hatchery	10
GUE	Guernsey, UK	49.497	−2.502	Hatchery	10
SCO	Oban, UK	55.534	−5.244	Hatchery	8
SES	Sea Salter, UK	51.378	1.212	Hatchery	9
MAL	Maldon, UK	51.724	0.71	Hatchery	9
WAL	Bangor, UK	53.098	−4.15	Hatchery	6
GER	Sylt, Germany	55.152	8.253	Wild	12
DEN	Limfjorden, Denmark	56.833	8.906	Wild	11
SWE	Kristineberg, Sweden	58.134	58.134	Wild	12
NOR	Arendal, Norway	58.428	8.793	Wild	11
JAP	Matsushima, Japan	38.367	141.066	Wild	12
CAN	Vancouver, Canada	50.164	−124.432	Wild	12
All	-	-	-	-	232

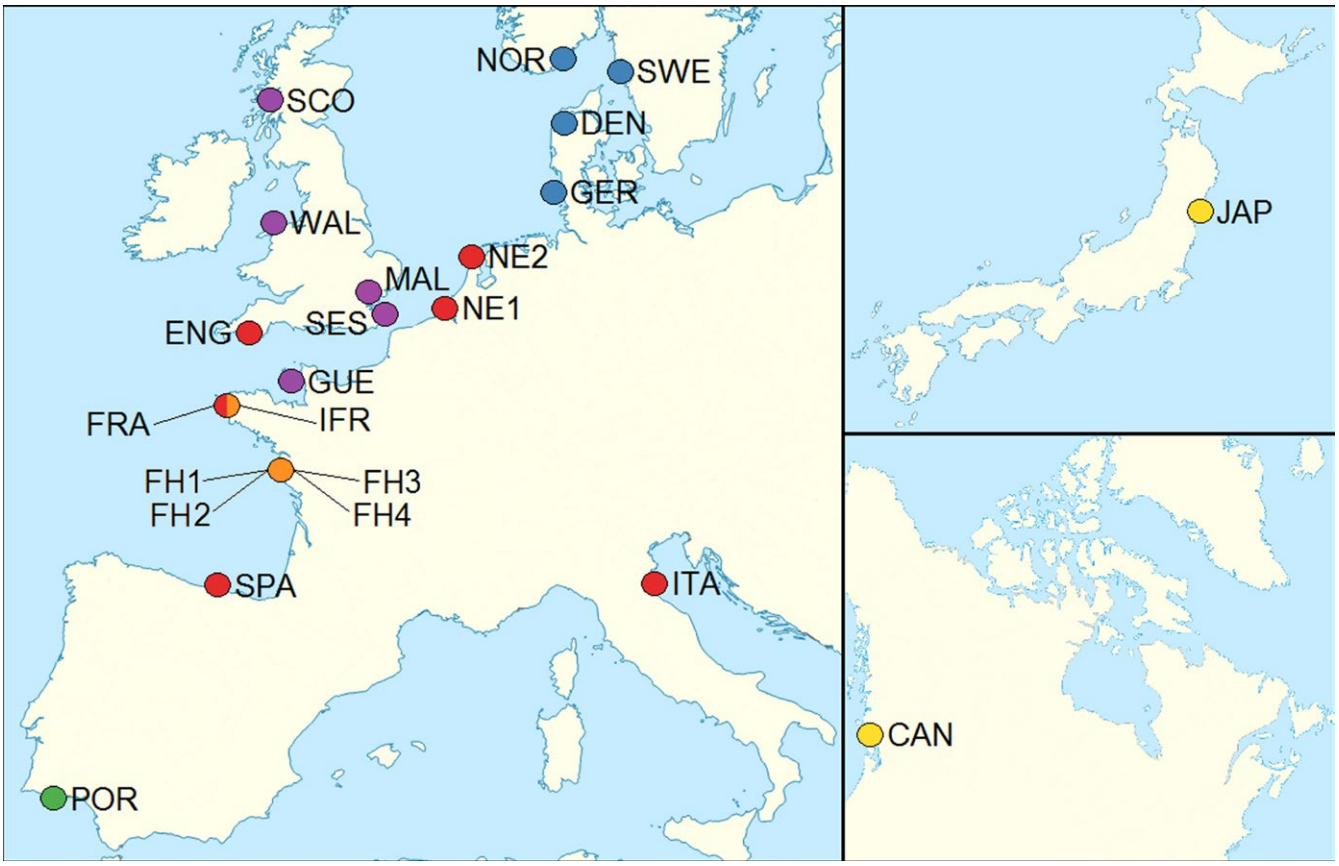

FIGURE 1 Map showing Pacific oyster sampling locations. The two source populations from the Miyagi Prefecture in Japan and British Columbia in Canada are indicated with yellow circles. Wild and hatchery populations within the southern group are indicated in red and orange, respectively, while wild populations and hatcheries within the northern group are represented by blue and purple circles, respectively. Finally, the population from Portugal, where hybridization between *C. gigas* and *C. angulata* is known to occur, is indicated by a green circle

among hatcheries, identifying potential sources of broodstock, and for the elaboration of other effective management strategies aimed at minimizing inbreeding within hatchery propagated Pacific oysters. Consequently, we believe this study provides important information for breeding programs as well as a baseline for future studies.

2 | MATERIALS AND METHODS

2.1 | Sample collection and DNA extraction

Pacific oyster samples were collected between November 2014 and March 2016 from 12 different sites along the Atlantic seaboard of mainland Europe as well as from one location in the Mediterranean (Table 1 and Figure 1). Samples from Scotland (SCO) and Wales (WAL) were from hatcheries, while the remaining 11 populations were wild. Specimens from Portugal (POR) originated from an area where hybridization between *C. angulata* and *C. gigas* is known to take place (Batista et al., 2017; Huvet et al., 2004) and could therefore represent *C. gigas* samples introgressed with *C. angulata*. For comparison, we also included samples from the Miyagi Prefecture in Japan (JAP) and from British Columbia in Canada (CAN).

2.2 | DNA extraction and SNP genotyping

Adductor muscle tissue was taken from each adult oyster and stored in 95% ethanol at −20°C. Whole genomic DNA was then extracted following an adapted phenol–chloroform protocol (Sambrook, Fritsch, & Maniatis, 1989) and sent to Edinburgh Genomics for genotyping at 40,625 SNPs on the custom Affymetrix SNP Array (Gutierrez et al., 2017). Out of a total of 204 DNA extracts, 192 (94%) passed quality checks and were therefore selected for genotyping using the protocol described by Gutierrez et al. (2017).

2.3 | Incorporation of existing data

We also incorporated data into our study from 81 oysters that were previously genotyped on the same array (Gutierrez et al., 2017). These samples were initially included in the discovery panel used to develop and validate the SNP array and originated from eight hatcheries, three from the UK (MAL, SES, and GUE) and five from France (FH1–4 and IFR, Table 1 and Figure 1). In general, most companies use breeding practices based on rotational crosses among yearly cohorts (P. Harray, personal communication, November 21, 2018).

Consequently, the genetic variability of a given cohort may not be representative of the population as a whole.

After the inclusion of these additional samples, our dataset consisted of (a) six wild populations from the southern group (red circles in Figure 1); (b) five hatcheries from the southern group (orange circles in Figure 1); (c) four wild populations from the northern group (blue circles in Figure 1); (d) five hatcheries from the northern group (purple circles in Figure 1); (e) the source populations of Japan and Canada (yellow circles in Figure 1); and (f) a single population from Portugal, where hybridization between *C. angulata* and *C. gigas* is known to occur (green circle in Figure 1).

2.4 | SNP calling

We imported raw output data for all 273 samples into the Axiom Analysis Suite (version 3.1, Affymetrix) for quality control and genotype calling. All thresholds for quality assessment were set to the values recommended in the Affymetrix best practice workflow (Supporting information Table S1) and allowed for the categorization of each SNP into one of six possible classes: (a) "polymorphic high resolution" where the SNP passed all quality controls; (b) "no minor homozygote" where the SNP passed quality checks but no homozygotes for the minor allele were found; (c) "off-target variant" where, in addition to the heterozygote and the two alternative homozygotes, a fourth genotype cluster was also observed; (d) "monomorphic high resolution" where the SNP passed quality checks but was uninformative; (e) "call rate below threshold" where the genotype call rate was below the specified threshold of 97%; and (f) "other" where the SNP failed to pass any other quality threshold. Following Affymetrix recommendations, SNPs from the first two categories were retained for further analysis, in addition to a subset of SNPs from the third category that were selected after applying the "off-target caller" tool that allows for off-target variant recalibration. The resulting dataset was then filtered to retain only SNPs genotyped in at least 90% of individuals and only samples with less than 5% missing data. Finally, the software PLINK (version 1.9, Purcell et al., 2007) was used to prune out linked loci using an r^2 threshold of 0.5. The final dataset therefore comprised 232 individuals genotyped at 21,499 polymorphic, unlinked SNPs.

2.5 | 2.5 RAD sequencing

To provide a comparison with the SNP array data, we also RAD sequenced a representative subset of 40 individuals from eight populations (Supporting information Table S2). Specifically, we included the source population of Japan (JAP), the potential hybrid population from Portugal (POR), two geographically distant wild populations from the southern group (SPA and NE2), two wild populations from the northern group (DEN and NOR), the Mediterranean population from Italy (ITA), and a hatchery from Scotland (SCO). Only one hatchery could be included because DNA from the other hatcheries was either not of high enough quality to pass thresholds for library construction, or it was not available due to the sample having been genotyped as part of a previous study (Gutierrez et al., 2017). Whole genomic DNA was shipped to the Beijing Genomics Institute (BGI) for library preparation and sequencing. The libraries were constructed using the restriction enzyme PstI and sequenced on an Illumina X Ten platform to generate a total of 869,113,776 100 bp paired-end sequence reads. Already demultiplexed sequence data were received from BGI and further sequence quality assessment was performed using the software FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We then conducted a de novo assembly of the data and called genotypes using the Stacks 2.1 pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Values for the three main parameters *–m*, *–M*, and *–n* were chosen following the optimization procedure described by Rochette and Catchen (2017). Briefly, *–m* was set to three, and increasing values for *–M* and *–n* were tested. The combination of these parameters for which the number of polymorphic loci present in at least 80% of the individuals reached a plateau was defined as optimal. Two different strategies were employed: *–n* was either set as equal to *–M* or one unit greater, to account for the potential presence of fixed *C. angulata* polymorphisms (Paris, Stevens, & Catchen, 2017). The combination yielding the highest plateau (*m* = 3, *M* = 5, and *n* = 6; Supporting information Figure S1) was selected for analyzing the entire dataset, from which PCR duplicates were then removed. The raw genotypes were subsequently quality filtered to retain only biallelic SNPs with both genotype quality and depth of coverage greater than 10 using VCFTools (Danecek et al., 2011). Subsequently, all SNPs and individuals with genotyping rates below 10% were removed and only variants with minor allele frequency (MAF) greater than 0.05 were retained. Finally, the software PLINK was employed to prune out linked loci using an r^2 threshold of 0.2.

2.6 | Analysis of population genetic structure

Three complimentary approaches were used to characterize the strength and pattern of population genetic structure. First, we calculated pairwise F_{st} values among populations and determined their statistical significance based on 1,000 permutations of the dataset using Arlequin version 3.5.2.2 (Excoffier & Lischer, 2010). We then performed an analysis of molecular variance (AMOVA) employing the R package "poppr" version 2.8.0 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014) to evaluate the proportion of genomic variation explained by different hierarchical levels of population structure. Specifically, we started by dividing our data into four regions corresponding to the northern group, southern group, Portugal, and source populations. Each was then subdivided into wild and hatchery samples, which were further split based on the population of origin. Variation among samples within each population and within individuals was also evaluated. A randomization test with 1,000 repetitions was used to determine statistical significance. Subsequently, we used Mantel tests to evaluate the significance of isolation-by-distance patterns for the full dataset as well as for the wild and hatchery oysters separately for the northern and southern groups. For this analysis, geographic distances between

populations were calculated as shortest coastline distances using "FreeMapTools (https://www.freemaptools.com/measure-distance.htm)." In the case of locations not connected by land, the shortest sailing distance between coasts was used. Second, we used the R package Adegenet version 2.1.1 (Jombart, 2008; Jombart & Ahmed, 2011) to conduct a principal component analysis (PCA) of the SNP dataset. Third, we utilized the software package fineRADstructure (Malinsky, Trucchi, Lawson, & Falush, 2018) to infer population structure using a model-based Bayesian clustering approach that groups together individuals with high levels of shared coancestry. A "coancestry matrix," defined as a summary of nearest neighbor haplotype relationships, is required as input and was generated using the "RADpainter" module of fineRADstructure. We used the default parameters of 100,000 Markov chain Monte Carlo (MCMC) iterations with a burn-in of 100,000 iterations and sampling occurring every 1,000 iterations. A tree was then constructed with 10,000 hill-climbing iterations, and the results were visualized using the scripts FINERADSTRUCTUREPLOT.R and FINESTRUCTURELIBRARY.R, which are available via http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html.

2.7 | Genomic inbreeding coefficients

We calculated $\hat{F}_l$, a genomic inbreeding estimator based on the variance of additive genotype values (Yang, Lee, Goddard, & Visscher, 2011), for each individual in our dataset based on the SNP data. To test for an association between levels of relatedness and inbreeding, we calculated mean pairwise relatedness among individuals within populations from the SNP data using GCTA (Yang et al., 2011) and correlated this with mean $\hat{F}_l$ values. Genomic inbreeding coefficients were compared between populations and groups using a Kruskal–Wallis test followed by *post hoc* pairwise Mann–Whitney tests, whose *p*-values were adjusted according to Benjamini and Hochberg (1995), to formally test for significant pairwise comparisons. As variation in inbreeding causes heterozygosity to be correlated across loci, we also estimated the extent of identity disequilibrium (ID, Weir & Cockerham, 1973) by calculating the two-locus heterozygosity disequilibrium, g_2 (David, Pujol, Viard, Castella, & Goudet, 2007) within the R package inbreedR (Stoffel et al, 2016). The same package was also used to calculate the 95% confidence interval of g_2 by bootstrapping the data 1,000 times over individuals, as described by Stoffel et al. (2016).

3 | RESULTS

To provide detailed insights into the pan-European population structure of *C. gigas* and facilitate comparisons between wild and hatchery oysters, we analyzed medium-density SNP array data for a total of 273 individuals sampled from 23 locations. Data from 192 individuals were newly generated, while the remaining data were incorporated from Gutierrez et al. (2017). Sampling sites were putatively assigned to either the northern or the southern group on the basis of previous genetic studies (Lallias et al, 2015; Rohfritsch et al, 2013). Application of the filtering criteria described in the Materials and methods resulted in the exclusion of an 17,411 SNPs that did not meet Affymetrix recommendations and of an additional 1,715 SNPs due to low genotyping rates or linkage disequilibrium. On average, 10 individuals were genotyped for each location, and the final dataset consisted of 232 samples (see Table 1 for a breakdown by population) genotyped at 21,499 SNPs.

3.1 | Population genetic structure

To investigate broad-scale patterns of genetic differentiation, we used AMOVA to quantify the proportion of genomic variation attributable to each of five hierarchical levels of population substructure. As expected, over 95% of the total variation was partitioned within individuals. The remaining variance was mainly partitioned among the northern and southern groups, the source populations, and

TABLE 2 Results of the hierarchical analysis of molecular variance (AMOVA)

Source of variation	*df*	Sum of squares	% variation	Φ	*p*-value
Among regions	3	48,727.66	1.7	0.017	0.024
Between origins within regions	2	17,734.93	0.39	0.004	0.005
Among sampling locations within origins	17	1,16,998.80	2	0.02	0.001
Among samples within sampling locations	209	10,14,805.34	0.22	0.002	0.406
Within samples	232	11,21,225.26	95.69	0.043	0.001
Total	463	23,19,491.99	100	-	-

Notes. Five different hierarchical levels were evaluated. First, the dataset was divided into four "regions" corresponding to the southern group, the northern group, the source populations, and Portugal. Next, each region was divided into "origins" depending whether the samples were from wild populations or hatcheries. Finally, the remaining variance was partitioned among sampling locations, individuals within sampling locations, and within individuals.

Portugal (Φ = 0.017, p = 0.024, Table 2), between wild populations and hatcheries (Φ = 0.004, p = 0.005, Table 2) and among sampling locations (Φ = 0.02, p = 0.001, Table 2). Furthermore, the majority of pairwise F_{st} values between populations were highly significant, even after correction for multiple tests (Supporting information Table S3), although a significant isolation-by-distance pattern was only detected among wild populations belonging to the southern group (Mantel's r = 0.971; p = 0.022).

To evaluate population structure at the individual level, we performed a principal component analysis (PCA). Consistent with the AMOVA, a number of clear differences were apparent. First, the northern and southern groups clearly separated apart from one another, as did oysters from Portugal, although no genetic differences were apparent between the southern group and the two source populations (Figure 2). Furthermore, within both the northern and southern groups, hatcheries showed consistently greater scatter than wild populations, indicating that they may have experienced stronger genetic drift.

We next used a model-based Bayesian clustering approach implemented in fineRADstructure to infer population structure via shared ancestry. The resulting cladogram and coancestry matrix shown in Figure 3 confirmed the results of the AMOVA and PCA while also uncovering the presence of more subtle structuring. Specifically, two major clades were identified. The first of these, shown on the left of the cladogram and represented by a cluster of individuals in the bottom left of the coancestry matrix, comprised individuals from the northern group. This was further subdivided into two distinct clusters, the first comprising mainly individuals from the Seasalter, Maldon, and Bangor hatcheries in the UK and the second comprising mostly wild individuals from Germany and Scandinavia. The remaining individuals were grouped together into a second major clade shown on the right of the cladogram, which in turn was subdivided into three main clusters comprising the southern group and source populations, oysters from the Guernsey, Seasalter, and two of the French hatcheries, and Portugal. The fact that samples from Seasalter were distributed across two different clusters is consistent with the fact that oysters have been exchanged between Seasalter and Guernsey (M. Dravers, personal communication, November 21, 2018). Most of the individuals from the remaining French hatcheries could also be clearly distinguished within the southern group. Samples from Scotland were in a different part of the cladogram, but always clustered together with other samples from hatcheries. Furthermore, levels of shared coancestry varied appreciably across the dataset, with oysters from the Guernsey hatchery having the highest levels, the remaining hatchery samples as well as oysters from Portugal having intermediate levels, and wild individuals having the lowest levels.

3.2 | Variation in inbreeding

To explore patterns of inbreeding in wild populations and hatcheries, we calculated genomic inbreeding coefficients for all of the individuals in our dataset. Mean genomic inbreeding coefficients were strongly positively correlated with average pairwise relatedness values within populations (linear regression, b = 1.39, r^2 = 0.79, p < 0.001, Figure 4), which in turn were tightly correlated with mean shared coancestry values extracted from the fineRADstructure analysis (linear regression, b = 48.99, r^2 = 0.99, p < 0.001). This indicates that, as expected, genomic inbreeding coefficients tend to be higher in populations with elevated levels of relatedness and shared coancestry.

Pooling individuals according to the six main groups described in the first paragraph of the results revealed highly significant differences (Figure 5a, Kruskal–Wallis test: χ^2 = 105.17, df = 4, p < 0.001). First of all, regardless of whether wild populations or hatcheries were considered, a significant tendency was found for genomic inbreeding coefficients to be higher in the northern than in the southern group (*post hoc* pairwise Mann–Whitney tests: wild: adjusted p < 0.001, hatchery: adjusted p = 0.005). This is consistent with previous studies showing that genetic diversity is lower in the northern group (see Discussion). Second, genomic inbreeding coefficients were significantly higher in hatcheries versus wild populations (*post hoc* pairwise Mann–Whitney tests: northern group: adjusted p < 0.001, southern group: adjusted p < 0.001). In line with this, identity disequilibrium was higher in individuals sampled from hatchery (g_2 = 0.0065, bootstrap 95% confidence interval = 0.0044–0.0089) than from wild populations (g_2 = 0.0022, bootstrap 95% confidence interval = 0.0014–0.0029), indicating that hatchery-reared oysters

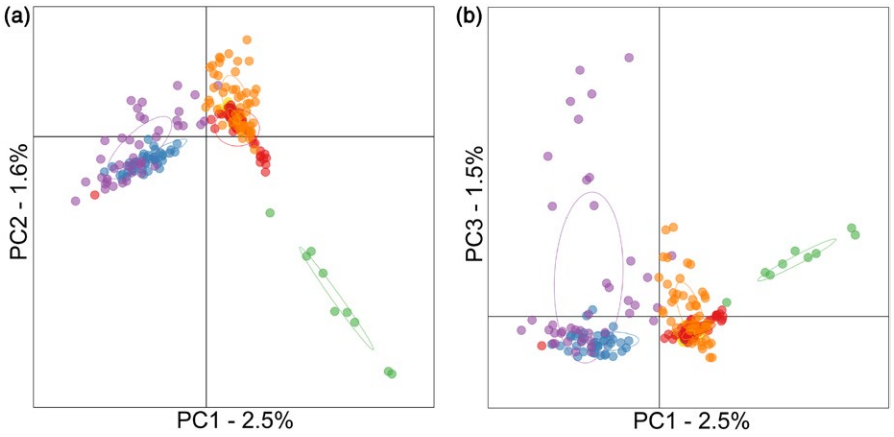

FIGURE 2 Scatterplot showing individual variation in principal component (PC) scores derived from a principal component analysis (PCA) of the genomic data. Panel (a) shows PC1 plotted against PC2, and panel (b) shows PC1 plotted against PC3. The amounts of variation explained by each PC are given as percentages. Samples are color coded as described in the legend of Figure 1

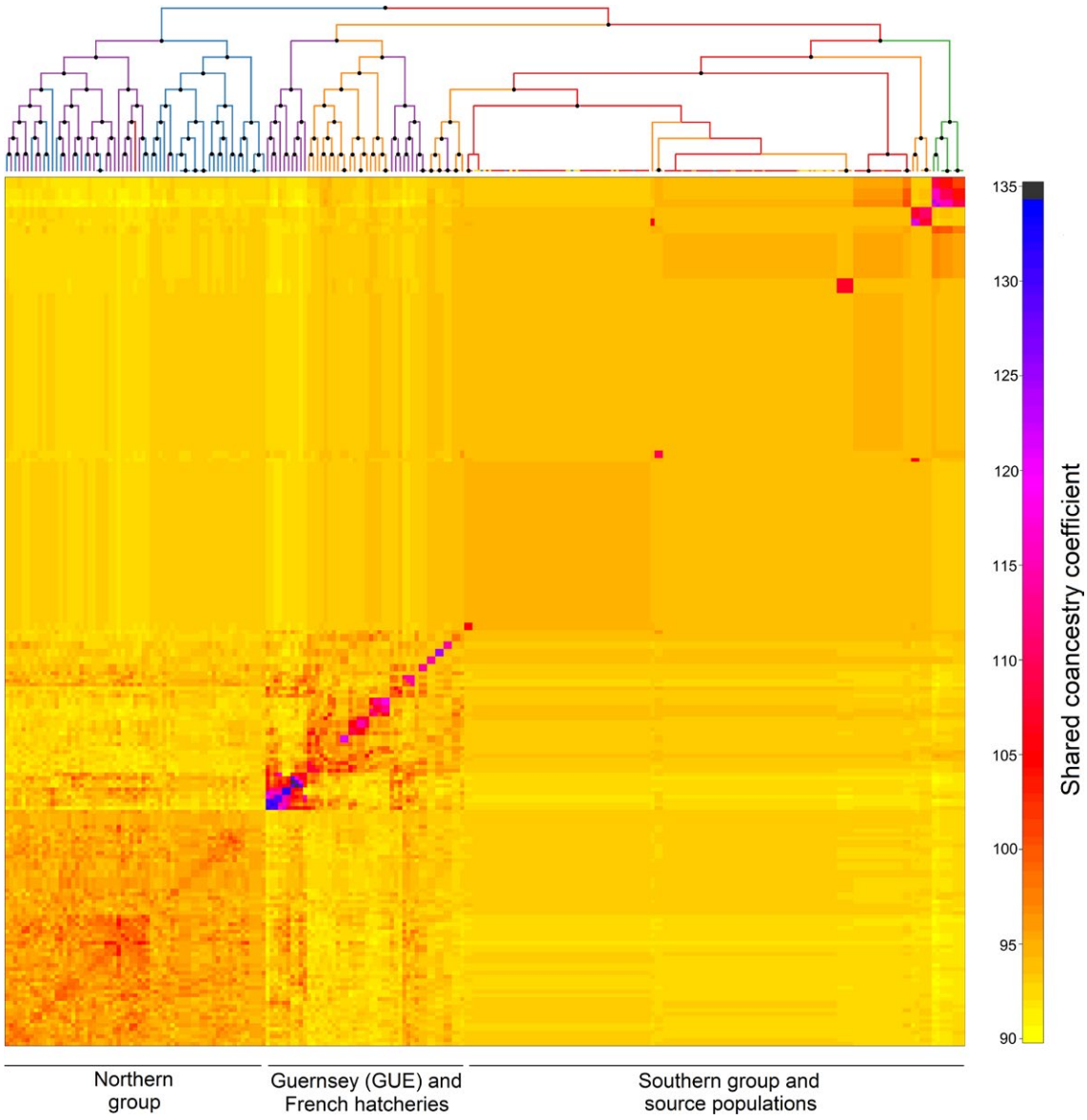

FIGURE 3 Output of the fineRADstructure analysis of the genomic data. In the cladogram, branches are color coded according to sampling origin as shown in Figure 1 and nodes with greater than 95% bootstrap support are marked by black points. The heatmap indicates pairwise coancestry between individuals, with blue and purple representing the highest levels, red and orange indicating intermediate levels, and yellow representing the lowest levels of shared coancestry

have greater overall variance in inbreeding (Figure 5b). This reflects a general tendency for the variance in inbreeding to be higher both within and among hatcheries relative to wild populations (Figure 5c). The highest genomic inbreeding coefficients were found in oysters from Guernsey, while in France and England, hatcheries with both intermediate (e.g., Fh1, FH3, SES, and MAL) and relatively low (e.g., IFR, FH2, FH4, WAL, and SCO) levels of inbreeding were present.

3.3 | RAD sequencing

Although we did not expect our results to be strongly affected by ascertainment bias because oysters from both the northern and southern groups were used in the discovery panel for the SNP array (Gutierrez et al., 2017), we nevertheless generated for comparison a parallel genetic dataset comprising RAD sequencing data for 40 individuals from eight populations (see Materials and methods for details). This resulted in a total of 869,113,776 high-quality paired-end reads that were assembled into 697,354 RAD loci. From these, we called a total of 7,322,935 SNPs, of which 115,087 were retained for further analyses after filtering. We found a virtually identical pattern of population structure, with the PCA clearly discriminating the northern group from the southern group, the wild populations from the hatcheries, and oysters from Portugal (Figure 6a and b). Similarly, fineRADstructure identified four main groups comprising the southern group and source population, the northern group, Portugal, and the Scottish hatchery (Figure 6c). Genomic inbreeding coefficients based on the RAD data showed a similar pattern to those calculated from

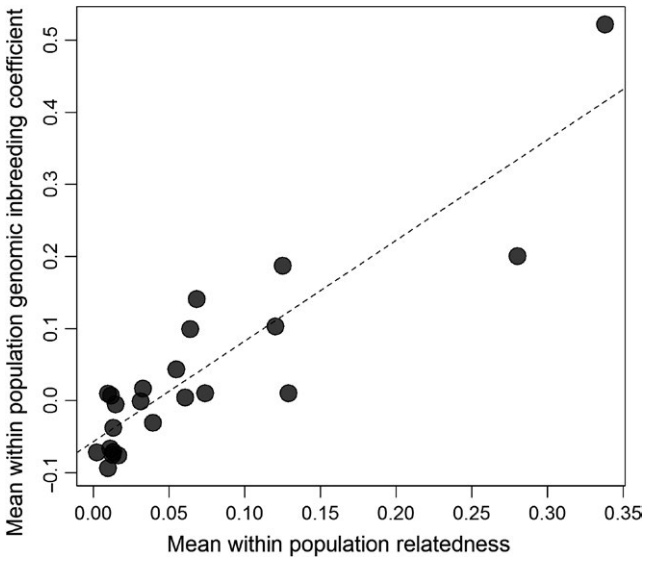

FIGURE 4 Relationship between average pairwise relatedness and average genomic inbreeding coefficient within populations (linear regression, $b = 1.39$, $r^2 = 0.79$, $p < 0.001$)

the SNP array (Figure 6d), although inbreeding levels appeared to be somewhat higher for the Danish and Norwegian populations. Notably, oysters from Portugal had the highest levels of inbreeding in our RAD analysis, which would not be expected if our sample contained hybrids. Nevertheless, due to the high pairwise F_{ST} values obtained in all comparisons involving oysters from Portugal (Supporting information Table S3), we cannot exclude the possibility that these samples may actually be pure *C. angulata* rather than pure *C. gigas*. *(see Discussion)*

4 | DISCUSSION

We used a medium-density SNP array to characterize the genetic structure of *C. gigas* populations across Europe as well as to evaluate levels of inbreeding in wild and hatchery oysters. Our comprehensive sampling design coupled with high-resolution genomic data allowed us to resolve patterns of genetic differentiation over both broad and fine geographic scales. Specifically, we found clear support for a northern and southern European group, with the latter being virtually identical to the Japanese and Canadian source populations, consistently with previous studies (Huvet, Lapegue, Magoulas, & Boudry, 2000; Moehler et al., 2011; Rohfritsch et al., 2013). We furthermore resolved substantial genetic differences between wild populations and hatcheries and compared genomic inbreeding coefficients to show that some of the sampled hatchery cohorts have higher levels of inbreeding than wild populations. Given that *C. gigas* carries a high genetic load that has been proposed to be responsible for substantial early mortality (Launey & Hedgecock, 2001; Plough & Hedgecock, 2011; Taris et al., 2007) as well as variation in commercially important adult traits (Evans et al., 2004), we believe that our findings could have important implications for aquaculture.

4.1 | Population genetic structure

Several studies have investigated the population genetic structure of Pacific oysters in Europe and interpreted their findings in the light of the known and rather complex history of multiple introductions and invasions. Our research compliments and builds upon these studies in a number of ways. First, we were able to confirm previous findings based on mitochondrial DNA as well as small panels of nuclear markers (Huvet et al., 2000; Lallias et al., 2015; Meistertzheim et al., 2013; Moehler et al., 2011; Rohfritsch et al., 2013) that European Pacific oyster populations are broadly structured into northern and southern groups. Although this is not necessarily surprising, studies based on one or a few markers can suffer from biases related to stochastic processes (Rokas & Carroll, 2005). Consequently, our study lends further weight to the conclusion that the north–south divide is a genome-wide phenomenon that is robust to different methodologies and repeatable across studies.

We were also able to confirm previous studies (Huvet et al., 2000; Moehler et al., 2011; Rohfritsch et al., 2013) reporting negligible genetic differentiation between the source population of Japan and the southern European group. Despite having analyzed samples from both Japan and British Columbia, which was a secondary site of introduction into Europe (Wolff & Reise, 2002), and having several orders of magnitude higher genetic resolution than previous studies, both PCA and fineRADstructure failed to detect any clear differences between the southern group and source populations. Furthermore, although a number of comparisons involving Japan and British Columbia yielded significant F_{st} values, the magnitude of these estimates was low. Our results therefore lend additional weight to the notion that Pacific oysters did not experience a pronounced founder effect when they were introduced into southern Europe. This is consistent with the observation that many thousands tonnes of spat were transferred into northern France from Japan as well as many hundreds of tonnes of adults from British Columbia (Grizel & Héral, 1991; Nehring, 1999; Wolff & Reise, 2002).

In addition to confirming previous findings, our genomic data also allowed us to resolve fine-scale patterns that could not be detected in previous studies. In particular, Huvet et al. (2004), Meistertzheim et al. (2013), and Rohfritsch et al. (2013) did not find any significant genetic differences among populations of the southern group, regardless of whether mitochondrial or nuclear markers were used. This apparent homogeneity was attributed to the prodigious reproductive potential of this species coupled with the possession of long-lived pelagic larvae and frequent transfers of farmed stocks (Meistertzheim et al., 2013; Rohfritsch et al., 2013). By contrast, we not only found that a number of pairwise population comparisons within the southern group yielded significant F_{st} values (Supporting information Table S3) but also uncovered a significant isolation-by-distance pattern among the wild populations from southern Europe. The highest F_{st} values in southern Europe were obtained for comparisons involving Italy, presumably due to the fact that *C. gigas* was introduced into this part of the Mediterranean during the late 1960 s (Šegvić-Bubić et al., 2016). We also found some evidence for the presence of fine-scale population structure within the

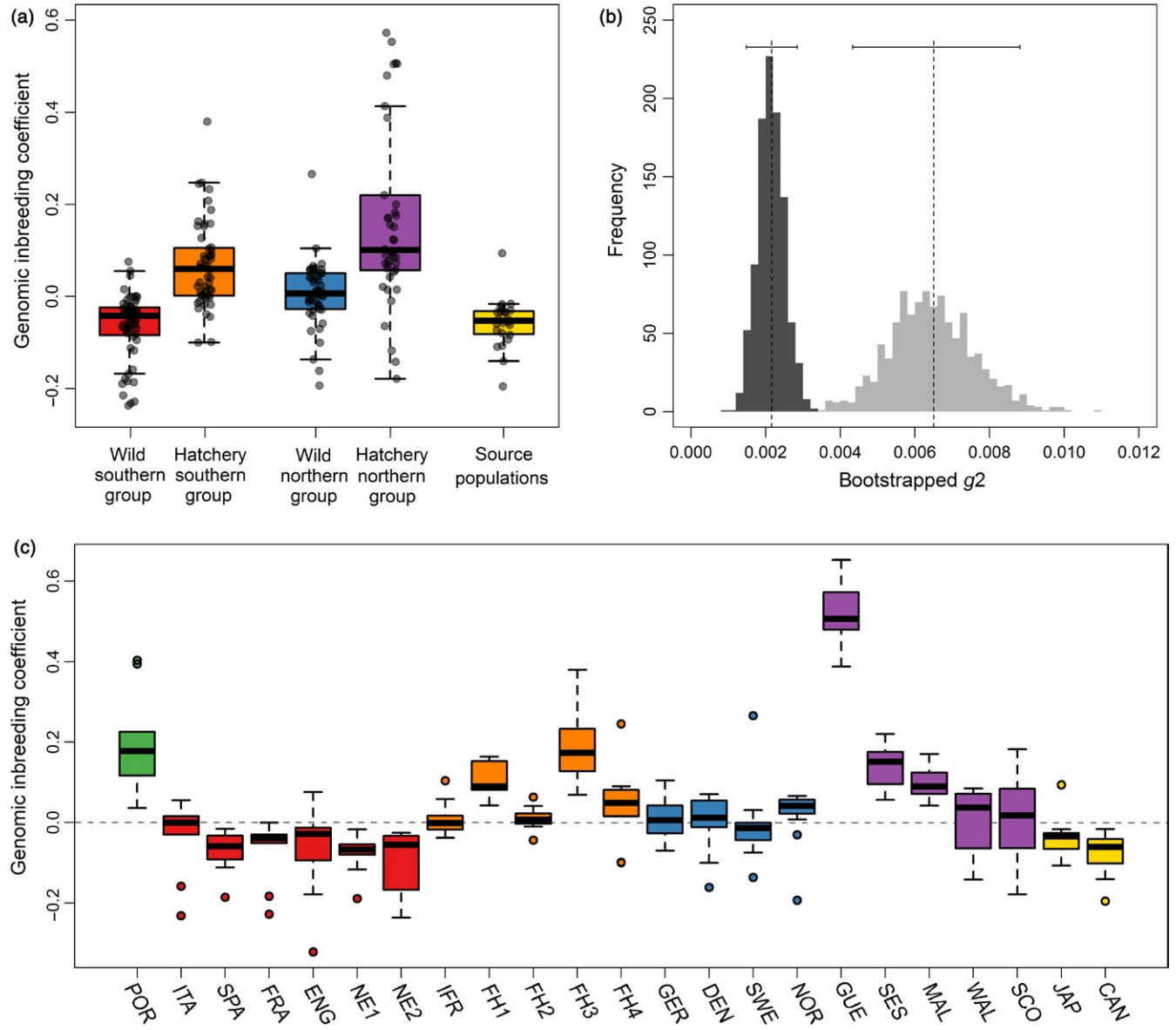

FIGURE 5 Levels of inbreeding in wild populations and hatcheries inferred from genome-wide SNP data. Panel (a) shows differences between wild and hatchery samples from the northern and southern groups separately as well as for the source populations. Raw data points are shown together with standard Tukey box plots. Panel (b) shows bootstrapped g_2 values for individuals sampled from wild populations (dark gray) versus hatcheries (light gray). The empirical g_2 values and their corresponding 95% confidence intervals are depicted by dashed vertical lines and horizontal bars, respectively. Panel (c) shows population-specific variation in inbreeding. In panels (a) and (c), the populations are color coded as described in the legend of Figure 1

northern cluster, although this was more equivocal. Specifically, most but not all of the individuals from Scandinavia and Germany clustered apart from the British hatcheries in the fineRADstructure analysis, although this distinction was not readily apparent in the PCA. Such a pattern is consistent with Pacific oysters having been imported repeatedly from UK hatcheries to Germany and Scandinavia (d'Auriac et al., 2017).

4.2 | Comparison of wild populations and hatcheries

Two innovations of our study were first to sample wild and hatchery oysters extensively enough to facilitate a meaningful and broad-scale comparison, and second to quantify inbreeding directly from genomic data. Repeated introductions of genetic material from different aquaculture broodstocks are commonplace and should in principle contribute toward the genetic homogenization of wild populations and hatcheries (Moehler et al., 2011). Moreover, a certain degree of genetic exchange between wild populations and hatcheries can be expected, at least in France where oyster production in some hatcheries is partially based on wild-caught spat and natural reproduction of farmed oysters occurs (Pouvreau et al., 2016). Set against this, however, temporal sweepstake effects (Hedgecock & Pudovkin, 2011) and far smaller numbers of breeding individuals in aquaculture populations (Kochmann et al., 2012) could potentially increase genetic drift and drive genetic differentiation from wild

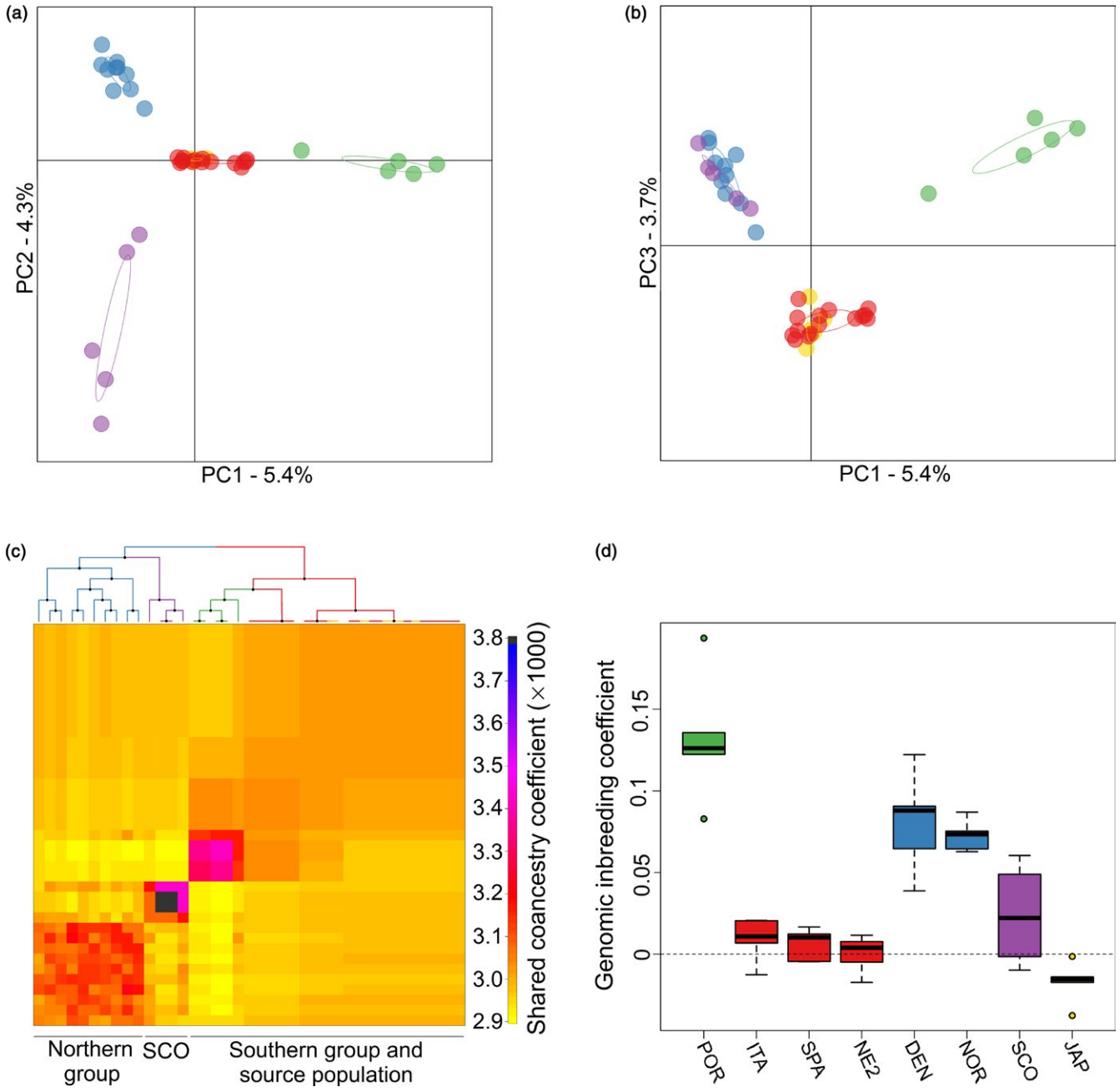

FIGURE 6 Results of repeated analyses of population genetic structure and inbreeding based on a subset of RAD sequenced individuals from eight populations (see Methods for details). Panels (a) and (b) show individual variation in principal component (PC) scores derived from a principal component analysis (PCA). Panel (a) shows PC1 plotted against PC2, while panel (b) shows PC1 plotted against PC3. The amounts of variation explained by each PC are given as percentages. Panel (c) presents the output of the fineRADstructure analysis, including both the cladogram and the heatmap representing pairwise coancestry values between individuals. Panel (d) shows population-specific variation in inbreeding. Samples are color coded as described in the legend of Figure 1

populations. Our data lend support to the latter scenario as we found that hatcheries showed a clear tendency to cluster apart from wild populations and were also characterized by elevated levels of shared coancestry and inbreeding.

Although small panels of genetic markers like microsatellites are capable of resolving population structure, under most circumstances they provide poor estimates of inbreeding (Balloux et al, 2004). This has hindered the study of inbreeding in wild populations lacking pedigrees (Pemberton, 2008). Consequently, we used our SNP data to calculate genomic inbreeding coefficients for the first time to our knowledge for a marine invertebrate in order to investigate how aquaculture practices may have influenced levels of inbreeding in oyster hatcheries. We uncovered a clear tendency for both the magnitude of inbreeding and its variance to be higher within the sampled hatchery cohorts. This might be considered surprising given that *C. gigas* is produced in vast numbers and is capable of long-distance

dispersal mediated by free-swimming planktotrophic larvae (FAO, http://www.fao.org/fishery/culturedspecies/Crassostrea_gigas). However, Pacific oysters also have one of the smallest documented effective to census population size ratios (10^{-6}, Frankham, 1995) reflecting a general tendency in marine invertebrates for highly variable sweepstakes reproductive success resulting from a combination of high fecundity and low larval survivorship (Hedgecock & Pudovkin, 2011). Concretely, a single oyster can produce several tens of millions eggs in a single season (FAO, http://www.fao.org/fishery/culturedspecies/Crassostrea_gigas), but mortality rates within commercial oyster hatchery cultures can be as high as 98% between fertilization and the spat stage (Plough & Hedgecock, 2011), which may lead to high variance in the reproductive success of hatchery broodstock (Boudry, Collet, Cornette, Hervouet, & Bonhomme, 2002).

Our findings are in line with a previous study documenting lower microsatellite allelic diversity in hatchery-reared relative to wild individuals within Loch Foyle in Northern Ireland (Kochmann et al., 2012), although a similar study did not find any differences between wild populations and hatcheries in the Wadden Sea (Moehler et al, 2011). Moreover, heterozygote deficiency has been observed in several previous studies of oysters (Lallias et al., 2015, 2010; Meistertzheim et al., 2013; Rohfritsch et al., 2013), which has been interpreted as being suggestive of inbreeding (Faust et al., 2017). Finally, experimental studies have observed massive distortions in marker segregation ratios in F_2 oyster families, consistent with a high genetic load comprising multiple recessive mutations under strong viability selection (Launey & Hedgecock, 2001; Plough & Hedgecock, 2011). Hence, our study contributes toward a growing body of evidence in support of Launey and Hedgecock's (2001) argument that inbreeding may be a biologically and economically important phenomenon in oysters as well as possibly in other marine invertebrates.

It is important to recognize that not all of the hatchery-reared oysters in our study showed higher levels of inbreeding than wild populations. By implication, inbreeding is not associated with hatchery propagation *per se* but may rather arise due to differences in management practices among facilities, which in many cases will reflect differing priorities. For example, many hatcheries minimize the risk of inbreeding by enhancing their broodstock with oysters collected from the wild (E. Vernier, *personal communication*, November 21, 2018) or by maximizing numbers of effective breeders, while others actively avoid these practices (M. Montergous, *personal communication*, June 6, 2018), presumably to minimize the risk of disease transmission. In other cases, it may be desirable to maintain particular families or lineages that have been selected based on specific characteristics, even if this results in somewhat higher levels of consanguinity (M. Dravers, *personal communication*, September 6, 2018).

4.3 | Practical implications for oyster aquaculture

Moderate to high levels of inbreeding are known to negatively impact a multitude of commercially relevant fitness traits, from individual growth rate through harvest body size to survival, in many aquaculture organisms (Deng et al., 2005; Gallardo et al., 2004; Lyu et al., 2018; Moss et al., 2007). More specifically, previous studies of oysters have found strong inbreeding depression for early viability (Plough & Hedgecock, 2011) as well as for yield, growth rate, and survival to harvest in adults (Evans et al., 2004). Consequently, elevated inbreeding levels in certain hatcheries are worthy of further exploration and it may be worth considering intervention strategies aimed at increasing genetic diversity.

With respect to the need for further exploration, it is worth bearing in mind that although our total sample size of oysters was reasonably large given the number of markers deployed, only around ten samples were analyzed on average from each population. While this is unlikely to have appreciably affected our inference of population structure (Willing, Dreyer, & Oosterhout, 2012), the inference of inbreeding levels within populations could be sensitive to the inadvertent sampling of highly related individuals when sample sizes are small, and this may be particularly true for hatcheries. Furthermore, high variance in reproductive success within and across generations could potentially lead to different cohorts from the same hatcheries varying substantially in their levels of inbreeding. Consequently, it would be worthwhile enlarging sample sizes within hatcheries, as well as collecting and analyzing samples from the same locations over multiple years in order to provide more robust inferences of average inbreeding levels and allow these to be interpreted in the light of temporal variation. This would further benefit from the development of a larger SNP array and a more contiguous *C. gigas* reference genome, which would allow inbreeding to be evaluated with even greater precision through the use of mapped genetic markers to quantify runs of homozygosity.

Having done so, a useful next step would be to evaluate in greater detail the effects of different levels of inbreeding on commercially important traits within hatcheries. So far, only a single study has evaluated the effects of inbreeding on adult traits in a commercial growing environment and this was based on a crossing design that maximized variation in inbreeding among families (Evans et al., 2004). By contrast, our approach of quantifying inbreeding directly from genomic data could in principle circumvent the need for experimental crosses, thereby allowing inbreeding depression to be directly quantified in real hatchery populations when phenotypic data are available. This could help to inform hatchery managers about the potential costs of inbreeding and the possible benefits of intervention strategies.

Finally, a number of potential intervention strategies aimed at reducing inbreeding and increasing genetic diversity could be envisaged. The first obvious approach would be to incorporate individuals from wild populations into hatchery broodstocks, as also discussed by Lallias et al. (2010) in the context of flat oysters. However, caution is warranted as selective breeding in captivity may lead to adaptive changes that are absent from wild populations (Lachambre et al., 2017) so the fitness consequences of such crosses remain unclear. A second possibility would be exchange individuals more extensively among hatcheries. Within Europe, the practice of exchanging oyster stocks between different countries is becoming more common, but we are not aware of any such exchanges between the United

Kingdom and the European mainland, probably due to the perceived risk of disease transmission. A third possibility would be to mitigate the risk of inbreeding by implementing oyster rearing based on molecular pedigree assignments (Boudry, 2009; Lapegue et al., 2014) as is common practice in fish farming (Vandeputte & Haffray, 2014). Clearly, hatchery managers need to balance the pros and cons of selective breeding and maximizing genetic diversity, but either way genomic tools such as SNP arrays provide a means of evaluating the genetic consequences of chosen management practices.

4.4 | Caveats

SNP arrays provide a cost-effective and convenient route to genome-wide investigations but can be prone to ascertainment bias when the samples used in the initial SNP discovery phase differ from those being interrogated on the array (Lachance & Tishkoff, 2013). However, we believe this is unlikely to substantially affect our main conclusions for two reasons. First, the discovery panel of individuals used to construct the array was unusually large, comprising over 200 individuals from eight different localities. By implication, much of the genetic diversity of the species in Europe should have been captured, including rare alleles that may easily be missed with smaller discovery panels but which could potentially be present at higher frequencies in unsampled populations. Second, although the discovery panel comprised primarily individuals from hatcheries, the northern and southern groups were roughly equally represented and we therefore have no reason to expect any broad-scale biases to be present. Two further points should also be recognized. First, our analyses of population structure will if anything be conservative, as ascertainment bias should lead to the underestimation of genetic differentiation when peripheral populations carry previously undetected alleles. Second, ascertainment bias cannot explain higher levels of inbreeding nor variation in inbreeding among hatcheries in the UK and France. This is because all of these populations were used to generate the array, so ascertainment bias if present would be expected to generate the opposite pattern of increased homozygosity in wild populations.

Nevertheless, we conservatively took into account the possibility that ascertainment bias could be responsible for the ostensibly high level of inbreeding in the putatively hybrid Portuguese population. To test this possibility as well as to confirm our broader findings, we RAD sequenced a subset of individuals and repeated all of our analyses. Our previous results were largely confirmed, with very similar patterns of population genetic structure and inbreeding being obtained, lending further weight to our main conclusions. Furthermore, oysters from Portugal were again found to have relatively high genomic inbreeding coefficients based on the RAD data. As we would expect hybrids to be relatively outbred, this finding points toward hybridization between *C. gigas* and *C. angulata* being negligible in our sample. Consequently, it appears that the Portuguese oysters could represent and inbred and isolated *C. gigas* population. However, we cannot discount the further possibility that we inadvertently sampled *C. angulata* from this location, as the two species are morphologically indistinguishable, the SNP array may include loci that cross amplify in *C. angulata* (Gagnaire et al., 2018), and F_{st} comparisons involving our Portuguese sample were consistently high.

5 | CONCLUSION

We harnessed some of the latest developments in genomics to shed new light on the population structure of Pacific oysters along a European latitudinal cline as well as to compare levels of inbreeding between wild and hatchery samples. The several orders of magnitude higher genetic resolution provided by the medium-density SNP array allowed us not only to confirm previous findings (Faust et al., 2017; Lallias et al., 2015; Meistertzheim et al., 2013; Rohfritsch et al., 2013) but also to detect fine-scale patterns including genetic differences between wild populations and hatcheries. We furthermore uncovered evidence for higher levels of inbreeding in sampled hatchery cohorts, which merits further investigation. Finally, our study contributes to a growing consensus that inbreeding could be more prevalent in animal populations than previously envisaged (Keller & Waller, 2002), even in highly fecund species with high dispersal.

ACKNOWLEDGEMENTS

We thank Carsten Fomsgaard, Jonathan King, Robert Mescam, Stein Mortensen, Kirti Ramesh, Yoshiko Shimahara, and Karin Troost for providing Pacific oyster samples or DNA extracts as well as SYSAAF for coordinating the provision of the French hatchery samples from Marinove, Satmar, France Naissain and Novostrea. We are also grateful to Melody Clark, Dennis Hedgecock and Pierrick Haffray as well as the participants of the 7th International Oyster Symposium (Bangor, Wales 2017) and the Wild Genomics Meeting (Bielefeld, Germany 2018) for useful discussions. Martin Stoffel provided helpful advice on data analysis, while Stéphane Duval, Mathieu Montergous, Emmanuel Vernier, Mark Dravers, and Gordon Goldsworthy shared information on Pacific oyster breeding practices. The research leading to these results has received funding from the European Union Marie Curie Seventh Framework Programme under grant agreement no. 605051. We also gratefully acknowledge funding from BBSRC and NERC (BB/M026140/1, NE/P010695/1) in addition to BBSRC Institute Strategic Program Grants (BBS/E/D/20002172 and BBS/E/D/30002275). Finally, we acknowledge the financial support of the German Research Foundation (DFG) and the Open Access Publication Fund of Bielefeld University for the article processing charge.

CONFLICT OF INTEREST

None Declared.

DATA ACCESSIBILITY

Individual genotype files obtained from the SNP array and from the analyses of the RAD sequencing data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6d778b6

ORCID

David L. J. Vendrami https://orcid.org/0000-0001-9409-4084

Pierre Boudry https://orcid.org/0000-0002-5150-2276

Joseph I. Hoffman https://orcid.org/0000-0001-5895-8949

REFERENCES

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Balloux, F., Amos, W., & Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, *13*(10), 3021–3031. https://doi.org/10.1111/j.1365-294X.2004.02318.x

Batista, F. M., Fonseca, V. G., Ruano, F., & Boudry, P. (2017). Asynchrony in settlement time between the closely related oysters *Crassostrea angulata* and *C. gigas* in Ria Formosa lagoon (Portugal). *Marine Biology*, *164*(5), 110.

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

Boudry, P. (2009). Genetic variation and selective breeding in hatchery-propagated molluscan shellfish. In M. Burnell, & A. Geoff (Eds.), *New technologies in aquaculture: Improving production efficiency, quality and environmental management* (pp. 87–108). Cambridge, UK: Woodhead Publishing Limited.

Boudry, P., Collet, B., Cornette, F., Hervouet, V., & Bonhomme, F. (2002). High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture*, *204*(3–4), 283–296. https://doi.org/10.1016/S0044-8486(01)00841-9

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

d'Auriac, M. B. A., Rinde, E., Norling, P., Lapègue, S., Staalstrøm, A., Hjermann, D. Ø., & Thaulow, J. (2017). Rapid expansion of the invasive oyster *Crassostrea gigas* at its northern distribution limit in Europe: Naturally dispersed or introduced? *PLoS One*, *12*(5), e0177481.

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., … Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3–13. https://doi.org/10.1016/j.margen.2016.04.012

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, *16*(12), 2474–2487. https://doi.org/10.1111/j.1365-294X.2007.03330.x

Deng, Y., Liu, X., Zhang, G., & Guo, X. (2005). Inbreeding depression and maternal effects on early performance of Pacific abalone. *North American Journal of Aquaculture*, *67*(3), 231–236. https://doi.org/10.1577/A04-021.1

Evans, F., Matson, S., Brake, J., & Langdon, C. (2004). The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture*, *230*(1–4), 89–98. https://doi.org/10.1016/j.aquaculture.2003.09.023

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Faust, E., André, C., Meurling, S., Kochmann, J., Christiansen, H., Jensen, L. F., … Strand, Å. (2017). Origin and route of establishment of the invasive Pacific oyster *Crassostrea gigas* in Scandinavia. *Marine Ecology Progress Series*, *575*, 95–105. https://doi.org/10.3354/meps12219

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. *Genetics Research*, *66*(2), 95–107. https://doi.org/10.1017/S0016672300034455

Gagnaire, P. A., Lamy, J. B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., … Lapègue, S. (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *GenomeBiology and Evolution*, *10*(9), 2518–2534. https://doi.org/10.1093/gbe/evy194

Gallardo, J. A., Garcıa, X., Lhorente, J. P., & Neira, R. (2004). Inbreeding and inbreeding depression of female reproductive traits in two populations of Coho salmon selected using BLUP predictors of breeding values. *Aquaculture*, *234*(1–4), 111–122. https://doi.org/10.1016/j.aquaculture.2004.01.009

Goulletquer, P., Bachelet, G., Sauriau, P. G., & Noel, P. (2002). Open Atlantic coast of Europe—a century of introduced species into French waters. In E. Leppäkoski S. Gollasch, & S. Olenin (Eds.), *Invasive aquatic species of Europe. Distribution, impacts and management* (pp. 276–290). Dordrecht: Springer.

Grizel, H., & Héral, M. (1991). Introduction into France of the Japanese oyster (*Crassostrea gigas*). *ICES Journal of Marine Science*, *47*(3), 399–403. https://doi.org/10.1093/icesjms/47.3.399

Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., Peñaloza, C., … Houston, R. D. (2017). Development of a medium density combined-species SNP array for pacific and european oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3: Genes, Genomes, Genetics*, *7*, 2209–2218.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: A review and commentary. *Bulletin of Marine Science*, *87*(4), 971–1002. https://doi.org/10.5343/bms.2010.1051

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A., … Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, *111*(10), 3775–3780. https://doi.org/10.1073/pnas.1318945111.

Huisman, J., Kruuk, L. E., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, *113*(13), 3585–3590. https://doi.org/10.1073/pnas.1518046113

Huvet, A., Fabioux, C., McCombie, H., Lapegue, S., & Boudry, P. (2004). Natural hybridization between genetically differentiated populations of *Crassostrea gigas* and *C. angulata* highlighted by sequence variation in flanking regions of a microsatellite locus. *Marine Ecology Progress Series*, *272*, 141–152. https://doi.org/10.3354/meps272141

Huvet, A., Lapegue, S., Magoulas, A., & Boudry, P. (2000). Mitochondrial and nuclear DNA phylogeography of *Crassostrea angulata*, the Portuguese oyster endangered in Europe. *Conservation Genetics*, *1*(3), 251–262.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis

on clonality. *Frontiers in Genetics*, *6*, 208. https://doi.org/10.3389/fgene.2015.00208

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity*, *115*(1), 63. https://doi.org/10.1038/hdy.2015.17

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kochmann, J., Carlsson, J., Crowe, T. P., & Mariani, S. (2012). Genetic evidence for the uncoupling of local aquaculture activities and a population of an invasive species—a case study of Pacific oysters (*Crassostrea gigas*). *Journal of Heredity*, *103*(5), 661–671. https://doi.org/10.1093/jhered/ess042

Lachambre, S., Day, R., Boudry, P., Huchette, S., Rio-Cabello, A., Fustec, T., & Roussel, S. (2017). Stress response of farmed European abalone reveals rapid domestication process in absence of intentional selection. *Applied Animal Behaviour Science*, *196*, 13–21. https://doi.org/10.1016/j.applanim.2017.07.004

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays*, *35*(9), 780–786. https://doi.org/10.1002/bies.201300014

Lallias, D., Boudry, P., Batista, F. M., Beaumont, A., King, J. W., Turner, J. R., & Lapègue, S. (2015). Invasion genetics of the Pacific oyster *Crassostrea gigas* in the British Isles inferred from microsatellite and mitochondrial markers. *Biological Invasions*, *17*(9), 2581–2595. https://doi.org/10.1007/s10530-015-0896-1

Lallias, D., Boudry, P., Lapegue, S., King, J. W., & Beaumont, A. R. (2010). Strategies for the retention of high genetic variability in European flat oyster (*Ostrea edulis*) restoration programmes. *Conservation Genetics*, *11*(5), 1899–1910. https://doi.org/10.1007/s10592-010-0081-0

Lapegue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., … Klopp, C. (2014). Development of SNP-genotyping arrays in two shellfish species. *Molecular Ecology Resources*, *14*(4), 820–830.

Launey, S., & Hedgecock, D. (2001). High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics*, *159*(1), 255–265.

Lyu, D., Wu, H., Hu, Y., & Wang, W. (2018). Inbreeding evaluation using microsatellite and its effect on growth traits in turbot *Scophthalmus Maximus*. *Aquaculture Research*, *49*(3), 1176–1181.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Meistertzheim, A. L., Arnaud-Haond, S., Boudry, P., & Thébault, M. T. (2013). Genetic structure of wild European populations of the invasive Pacific oyster *Crassostrea gigas* due to aquaculture practices. *Marine Biology*, *160*(2), 453–463. https://doi.org/10.1007/s00227-012-2102-7

Moehler, J., Wegner, K. M., Reise, K., & Jacobsen, S. (2011). Invasion genetics of Pacific oyster *Crassostrea gigas* shaped by aquaculture stocking practices. *Journal of Sea Research*, *66*(3), 256–262. https://doi.org/10.1016/j.seares.2011.08.004

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, *19*(4), 208–216. https://doi.org/10.1016/j.tree.2004.01.009

Moss, D. R., Arce, S. M., Otoshi, C. A., Doyle, R. W., & Moss, S. M. (2007). Effects of inbreeding on survival and growth of Pacific white shrimp *Penaeus (Litopenaeus) vannamei*. *Aquaculture*, *272*, S30–S37. https://doi.org/10.1016/j.aquaculture.2007.08.014

Nehring, S. (1999). Oyster beds and Sabellaria reefs. Wadden Sea Quality Status Report. *Wadden Sea Ecosystem*, *9*, 146–147.

Nehring, S. (2006). NOBANIS–Invasive alien species fact sheet–*Crassostrea gigas*. From: Online Database of the North European and Baltic Network on Invasive Alien Species-NOBANIS www. nobanis. org. *Fecha De Acceso*, *23*(08), 2010.

Norris, A. T., Bradley, D. G., & Cunningham, E. P. (1999). Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture*, *180*(3–4), 247–264. https://doi.org/10.1016/S0044-8486(99)00212-4

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373.

Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1635), 613–621.

Plough, L. V., & Hedgecock, D. (2011). Genome-wide analysis of stage-specific inbreeding depression in the Pacific Oyster *Crassostrea gigas*. *Genetics*, *189*, 1473–1486.

Pouvreau, S., Daniele, M., Auby, I., Lagarde, F., Le Gall, P., & Cochet, H. (2016). VELYGER database: The oyster larvae monitoring french project. SEANOE, 41888.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., 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

Quayle, D. B. (1988). Pacific oyster culture in British Columbia. In *Fisheries Research Board of Canada*, *Bulletin 169* (p. 192). Ottawa, Canada: The Queen's printer.

Rašić, G., Filipović, I., Weeks, A. R., & Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. *BMC Genomics*, *15*(1), 275. https://doi.org/10.1186/1471-2164-15-275

Reise, K. (1998). Pacific oysters invade mussel beds in the European Wadden SeaPazifische Austern bevölkern Miesmuschelbänke im europäischen Wattenmeer. *Senckenbergiana Maritima*, *28*(4–6), 167–175. https://doi.org/10.1007/BF03043147

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640. https://doi.org/10.1038/nprot.2017.123

Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F., & Lapegue, S. (2013). Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster *Crassostrea Gigas*. *Evolutionary Applications*, *6*(7), 1064–1078.

Rokas, A., & Carroll, S. B. (2005). More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*, *22*(5), 1337–1344. https://doi.org/10.1093/molbev/msi121

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (2nd ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Šegvić-Bubić, T., Grubišić, L., Zrnčić, S., Jozić, S., Žužul, I., Talijančić, I., … Katavić, I. (2016). Range expansion of the non-native oyster *Crassostrea gigas* in the Adriatic Sea. *Acta Adriatica-International Journal of Marine Sciences*, *57*(2), 321–330.

Shafer, A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907–917. https://doi.org/10.1111/2041-210X.12700

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: An R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, *7*(11), 1331–1339.

Taris, N., Batista, F. M., & Boudry, P. (2007). Evidence of response to unintentional selection for faster development and inbreeding

depression in *Crassostrea gigas* larvae. *Aquaculture*, *272*, S69–S79. https://doi.org/10.1016/j.aquaculture.2007.08.010

Troost, K. (2010). Causes and effects of a highly successful marine invasion: Case-study of the introduced Pacific oyster *Crassostrea gigas* in continental NW European estuaries. *Journal of Sea Research*, *64*(3), 145–165. https://doi.org/10.1016/j.seares.2010.02.004

Vandeputte, M., & Haffray, P. (2014). Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics*, *5*, 432.

Vendrami, D. L., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., … Hoffman, J. I. (2017). RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity. *Royal Society Open Science*, *4*(2), 160548. https://doi.org/10.1098/rsos.160548

Walne, P. R., & Helm, M. M. (1979). *Introduction of* Crassostrea gigas *into the United Kingdom. Exotic species in mariculture* (pp. 83–105). Cambridge, MA: MIT Press.

Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, *107*, 4–13. https://doi.org/10.1016/j.tpb.2015.08.006

Weir, B. S., & Cockerham, C. C. (1973). Mixed self and random mating at two loci. *Genetics Research*, *21*(3), 247–262. https://doi.org/10.1017/S0016672300013446

Willing, E. M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Wolff, W. J., & Reise, K. (2002). Oyster imports as a vector for the introduction of alien species into northern and western European coastal waters. In E. Leppakoski, S. Gollasch, & S. Olenin (Eds.), *Invasive aquatic species of Europe. Distribution, impacts and management* (pp. 193–205). Dordrecht: Springer.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Vendrami DLJ, Houston RD, Gharbi K, et al. Detailed insights into pan-European population structure and inbreeding in wild and hatchery Pacific oysters (*Crassostrea gigas*) revealed by genome-wide SNP data. *Evol Appl*. 2019;12:519–534. https://doi.org/10.1111/eva.12736