Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (Crassostrea gigas and Ostrea edulis)

Citation for published version:
Gutierrez, AP, Turner, F, Gharbi, K, Talbot, R, Lowe, NR, Peñaloza, C, McCullough, M, Prodöhl, PA, Bean, TP & Houston, RD 2017, 'Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (Crassostrea gigas and Ostrea edulis)' G3, vol. 7, no. 7, pp. 2209-2218. DOI: 10.1534/g3.117.041780

Digital Object Identifier (DOI):
10.1534/g3.117.041780

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
G3

G3: Genes|Genomes|Genetics Early Online, published on May 22, 2017 as doi:10.1534/g3.117.041780

Development of a medium density combined-species SNP array for Pacific and European oysters (*Crassostrea gigas & Ostrea edulis*)

Alejandro P. Gutierrez[*], Frances Turner [†], Karim Gharbi[†], Richard Talbot[†], Natalie R. Lowe[*], Carolina Peñaloza[*], Mark McCullough[‡], Paulo A. Prodöhl[‡], Tim P. Bean[§], & Ross D. Houston[*]

Affiliations

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

[†]. Edinburgh Genomics, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK

[‡]. Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast, UK

[§]. Centre for Environment Fisheries and Aquaculture Science, Cefas Weymouth Laboratory, Weymouth, Dorset, UK

Running title: Combined SNP array for oysters

Keywords: Pacific oyster, Flat oyster, Single nucleotide polymorphism (SNP), Array, Aquaculture.

Corresponding author: The Roslin Institute and Royal (Dick) School of Veterinary Studies,

University of Edinburgh, Midlothian EH25 9RG, UK. e-mail: ross.houston@roslin.ed.ac.uk

Abstract

SNP arrays are enabling tools for high-resolution studies of the genetic basis of complex traits in farmed and wild animals. Oysters are of critical importance in many regions from both an ecological and economic perspective, and oyster aquaculture forms a key component of global food security. The aim of our study was to design a combined-species medium density SNP array for Pacific oyster (*C. gigas)* and European flat oyster (*O. edulis)*, and to test the performance of this array on farmed and wild populations from multiple locations, with a focus on European populations. SNP discovery was carried out by whole genome sequencing of pooled genomic DNA samples from eight *C. gigas* populations, and RAD Sequencing of 11 geographically diverse *O. edulis* populations. Nearly 12 million candidate SNPs were discovered and filtered based on several criteria including preference for SNPs segregating in multiple populations and SNPs with monomorphic flanking regions. An Affymetrix Axiom® Custom Array was created and tested on a diverse set of samples (n = 219) showing ~ 27 K high quality SNPs for *C. gigas* and ~ 11 K high quality SNPs for *O. edulis* segregating in these populations. A high proportion of SNPs were segregating in each of the populations, and the array was used to detect population structure and levels of linkage disequilibrium. Further testing of the array on three *C. gigas* nuclear families (n = 165) revealed that the array can be used to clearly distinguish between families both based on identity-by-state clustering parental assignment software. This medium-density, combined-species array will be publicly available through Affymetrix, and will be applied for genome-wide association and evolutionary genetic studies, and for genomic selection in oyster breeding programs.

Background

Oyster farming is one of the most important aquaculture activities worldwide, providing a socioeconomic contribution to many coastal communities. Among the numerous farmed oyster species, the Pacific oyster (*Crassostrea gigas)* is one of the most widely cultivated with a global annual production estimated at 626 K tonnes in 2014 (FAO 2015). Starting in the 1960s, *C. gigas* was successfully introduced from Japan to all continents for cultivation (Troost 2010) due to its high acclimation ability, rapid growth and high production, and as an alternative to replace the flat oyster farms affected by persistent disease outbreaks (Pernet *et al.* 2016). Accordingly, the European flat oyster (*Ostrea edulis*), an endemic species to Europe has suffered a decrease in global production from 30 K tonnes in 1960 to 3 K tonnes produced in 2014. *O. edulis* is now a target for conservation efforts to help restore native populations (Lallias *et al.* 2010), and is also a niche aquaculture product, particularly in Europe and the USA.

In the past decade there has been increasing interest from researchers and industry in the development of genomic resources for oysters, mainly because of the economic and ecological importance of both *C. gigas* and *O. edulis.* The genomic toolbox for *C. gigas* includes a moderate number of genetic markers, such as microsatellites (Li *et al.* 2003; Sekino *et al.* 2003) and SNPs (Fleury *et al.* 2009; Sauvage *et al.* 2007; Wang *et al.* 2015). Low density linkage maps have been developed, containing both microsatellites and SNPs (Hedgecock *et al.* 2015; Hubert and Hedgecock 2004). In addition, quantitative trait loci (QTL) analyses have been carried out to identify genomic regions associated with desirable traits for aquaculture (Sauvage *et al.* 2010; Guo *et al.* 2012; Zhong *et al.* 2014). In addition, a reference genome sequence assembly is available for *C. gigas* (Zhang *et al.* 2012), albeit a number of putative assembly errors have been identified (Hedgecock *et al.* 2015). In contrast, genomic tools and resources are scarce for *O. edulis*, and only a limited number markers, mostly microsatellites and amplified fragment length polymorphism (AFLP) have been utilised for the development of a linkage map (Lallias *et al.* 2009; Lallias *et al.* 2007). Recently, the generation of genomic resources led to the development of a database containing genomic and transcriptome resources *for O. edulis* (Pardo *et al.* 2016; Vera *et al.* 2016).

SNPs have become the marker of choice in genetics research due to their high abundance, co-dominant mode of inheritance, ease of high-throughput discovery and low cost of genotyping per locus. Next-generation sequencing technologies enable efficient identification of many thousands of SNPs in a single experiment using either Whole-Genome Sequencing (WGS) or reduced representation approaches such as Restriction Site-associated DNA (RAD) sequencing (Baird *et al.* 2008; Davey *et al.* 2011). While the medium density SNP arrays typically generated by direct genotyping by sequencing approaches has been widely applied in aquaculture species (Robledo *et al.* 2017), SNP arrays can offer a higher density genotyping platform that is simpler to use. SNP arrays have been developed for most terrestrial livestock species such as cattle, pig and chicken (Matukumalli *et al.* 2009; Ramos *et al.* 2009; Kranis *et al.* 2013), and also for farmed finfish species such as Atlantic salmon, rainbow trout, catfish, carp among others (Houston *et al.* 2014; Yáñez *et al.* 2016; Palti *et al.* 2015; Liu *et al.* 2014; Xu *et al.* 2014). These arrays have formed the basis of genome-wide association studies for traits of economic importance such as resistance to pathogens (Geng *et al.* 2015; Correa *et al.* 2015; Tsai *et al.* 2016) and the application of genomic selection in aquaculture breeding (Ødegård and Meuwissen 2014; Tsai *et al.* 2015; Tsai *et al.* 2016; Vallejo *et al.* 2016).

For oyster species, low density SNP arrays for *C. gigas* and *O. edulis* have been developed, with 384 markers per species (Lapègue *et al.* 2014), and these have been applied for parentage assignment. In addition, a *C. gigas* specific high density array was recently developed, which contains approximately 134 K SNP markers shown to be polymorphic across populations sampled from China, Japan, Korea and Canada (Qi *et al.* 2017). However, a medium density combined-species platform is a worthy addition to the genomic toolbox for oysters because (i) the performance of the higher density (133K) array in farmed *C. gigas* populations from other global regions (e.g. Europe) is not known, (ii) medium density arrays are adequate for many genetics and breeding studies at substantially lower cost than high density arrays, and (iii) there is not yet a medium or high density genotyping platform for *O. edulis*. The major aim of the current study was to design and test a medium density combined-species SNP array for two key oyster species; *C. gigas and O. edulis*, and to test the performance of the array on hatchery and wild populations from multiple locations, as well as nuclear families from pair-crosses.

Methods

Sample collection and sequencing

The DNA sequencing protocols for SNP discovery were tailored to the status of genomic tools available for the two species. Since *C. gigas* has a reference genome sequence (Zhang *et al.* 2012), a whole genome resequencing approach was taken with reads subsequently aligned to the reference assembly as described below. There was no reference sequence available for *O. edulis*, so a RAD Sequencing approach was taken since this is suitable for *de novo* assembly and discovery of SNPs within RAD loci (Baird *et al.* 2008).

Samples from eight *C. gigas* populations from different geographical locations (primarily from hatcheries in the UK and France) were obtained, each comprising 13 to 47 individuals (Table 1). These included a population of 16 samples from lines of oysters which had been selected for resistance to Oyster Herpes Virus by Ifremer (France). Genomic DNA from all individuals was extracted the CTAB (cetyl trimethylammonium bromide) protocol described by Richards *et al.* (2013). Briefly, oyster tissue was incubated at 56 °C in lysis solution (3% CTAB, 100 mM Tris-HCl, pH 7.5, 25 mM EDTA, 2 mM NaCl) with 0.2 mg/mL proteinase K and 5ul of RNase (10mg/mL). After lysis, a chloroform extraction was performed twice and three volumes of CTAB dilution solution were added (1% CTAB, 50 mM Tris-HCl, pH 7.5, 10 mM EDTA, pH 8). The pellet was then washed in 0.4 M NaCl in TE, re-suspended in 1.42 M NaCl in TE and finally precipitated overnight in 1mL ethanol (99%) at -4 C. Within each population, DNA samples were then pooled in equimolar concentrations, and these pools were prepared for whole-genome sequencing (WGS) using the TruSeq Nano DNA Library Prep kit (Illumina, San Diego). Libraries were sequenced across five lanes of Illumina Hiseq 2500 to produce 125 bp paired end reads.

Samples from eleven *O.edulis* wild populations from diverse geographical locations were obtained (Table 1). Each population sample comprised 13 to 15 individuals, and genomic DNA had previously been extracted from these samples using a phenol-chloroform method. Equimolar pools of genomic DNA were generated for each population and the pooled genomic DNA was digested using the endonuclease PstI. Standard RAD libraries were constructed in three replicates following the standard protocol described by Baird *et al.* (2008). Equimolar amounts of all libraries were combined and sequenced on a single Illumina Hiseq 2500 lane to produce 125 bp paired end reads.

Table 1. Detail of populations sampled for sequencing and SNP discovery.

C. gigas			*O. edulis*		
Population	Location (Lat, Long)	N	Population	Location (Lat, Long)	N
Guernsey, England	49.497, -2.502	47	Croatia	42.855, 17.688	14
Maldon, England	51.724, 0.710	15	Lough Foyle, Ireland	55.130, -7.087	15
Sea Salter, England	51.378, 1.212	13	Lake Grevelingen, Neth.	51.709, 4.017	15
Ifremer, France	n/a	16	Larne, N. Ireland	54.817, -5.751	14
Hatchery 1 (Marinove), Fr	46.987, -2.238	29	Mersea, England	51.776, 0.9646	15
Hatchery 2 (SATMAR), Fr	46.948, -2.052	26	Baie de Quiberon, France	47.548, -2.996	15
Hatchery 3 (France Naissain), Fr	47.514, -2.666	29	Rossmore (Cork), Ireland	51.883, -8. 247	15
Hatchery 4 (Novostrea), Fr	46.954, -2.044	28	Sveio, Norway	59.519, 5.227	15
			Swansea Bay, England	51.604,-3.981	15
			Tralee, Ireland	52.316, -10. 028	13
			Damariscotta, Maine. USA	44.028, -69.534	14

SNP identification and filtering

C. gigas WGS reads were aligned to the *C. gigas* genome (GCA_000297895.1) using BWA-mem (v0.7.10) (Li and Durbin 2009) with the -M flag. Potential duplicated reads originating from PCR were then removed using Picard Tools (v1.69) MarkDuplicates and Samtools (v1.2) (Li *et al.* 2009). Local realignment around indels was performed using the GATK (v3.4.0) (McKenna *et al.* 2010) and alignments with a quality phred score >20 were retained. SNP calling was performed using Popoolation2 (Kofler *et al.* 2011), filtering to discard bases with a call quality phred score of <30.

O.edulis RAD-Seq reads were trimmed with Cutadapt (v1.7.1) (Martin 2011). Data from each of the three replicates described above were combined. Read 1 reads were clustered using ustacks (v1.30) with the parameters (-m 2 -M 5 -H”, followed by cstacks (Catchen *et al.* 2013) with the parameter “-n 2”, to create consensus sequences for each locus. RAD loci absent from ≥8 of the 11 pooled samples were discarded. Read 1 trimmed reads from each of the samples were then aligned to the set of RAD consensus sequences using BWA (v.0.7.9a) (Li and Durbin 2009) (Step 1). Reads mapping to each separate consensus sequence were then identified, and the corresponding read 2 sequences extracted from the trimmed data. These read 2 sequences for each locus were then assembled using IBDA-UD (Peng *et al.* 2012) (Step 2). The read 1 consensus sequences and the associated assembled read 2 sequences for each locus were merged using flash (v1.2.2) (Magoč and Salzberg 2011). For SNP discovery, the trimmed sequences corresponding to each locus were then mapped to the merged consensus sequence using smalt (v0.7.6). Duplicate reads were marked using Picard tools (v1.115) and realignments around indels performed using GATK indel realigner (v 3.4.0) (McKenna *et al.* 2010).

SNPs were identified and genotyped using PoPoolation2 and samtools (v1.3) pileup. Reads with a mapping quality phred score of <20 and bases with a call quality phred score < 20 were discarded.

SNP selection for Axiom array design

A list of candidate SNPs from both species (containing 1,691,005 and 117,235 priority SNPs from *C.gigas* and *O. edulis* respectively), was provided to Affymetrix as 71-mer nucleotide sequences from the forward strand with the alleles at the target SNP highlighted at position 36. A 'p-convert' value (representing the probability of a given SNP converting to a reliable SNP assay on the Axiom array system) was computed by Affymetrix for each submitted SNP sequence. Probes are assessed for each SNP in both the forward and reverse direction, in return each strand is designated as 'recommended', 'neutral', or 'not recommended' based on p-convert values.

The list of recommended markers (1,316,870 SNPs for *C.gigas* and *O. edulis* combined) was much greater than the total capacity of the Axiom MyDesign custom array. Therefore, additional filtering steps were carried out. For *C. gigas*, starting from the 1,216,467 Affymetrix-recommended SNPs, those with evidence for a 20 bp flanking monomorphic region covered by at least 36 reads from each pooled sample were retained (n = 186,948). For *O. edulis*, the Affymetrix-recommended SNPs (n = 100,403) were filtered so that each RAD locus contained a maximum of one SNP. When a RAD locus had multiple recommended SNPs, only the best SNP (based on the p-convert scores) was included (resulting in 59,976 candidate SNPs). Subsequently, to filter the SNPs to the required number for the array, SNPs for both species were selected according to the following additional filtering criteria: (i) highest p-convert values, (ii) even distribution across the reference genome (with at least 1000bp distance between pairs of SNPs for *C. gigas*), (iii) preference for those with a positive hit (minimum e-value 10E^{-4}) against the BLASTx NCBI NR database or against the *C. gigas* genome (for *O. edulis)*. In addition, most A/T and C/G SNPs transversions were discarded since these require double the space on the Affymetrix Axiom array platform. Additionally, 463 SNPs identified and validated by Hedgecock *et al.* (2015) passed the SNP filtering and scoring process and were included in the final array design.

SNP array validation

A plate of 384 individual genomic DNA samples (274 *C. gigas* and 110 *O. edulis*) was sent to Edinburgh Genomics (Edinburgh, UK) for genotyping using the array. Of these 384 samples, 219 were used for testing and validating the array's performance and quantifying the number of segregating SNPs in the various sampled populations. These 'included 109 *C. gigas* samples of individuals of unknown relatedness from eight populations (the same eight populations used for SNP discovery, plus an additional set of 28 broodstock oysters from Guernsey Sea Farms (Guernsey, UK)). The validation samples also included 110 *O. edulis* samples corresponding to the 11 population samples used for SNP discovery (Table 1), with n = 10 from each population. The remaining 165 samples were offspring of three nuclear families derived from parents from Guernsey Sea Farms, reared at the Centre for Environment, Fisheries and Aquaculture Science (Cefas, UK). These were analysed separately to test parentage assignment, genetic structure and within-family linkage disequilibrium levels (see below).

Raw data containing the results of the intensity calculations (CEL files) was imported into the Axiom Analysis Suite (v2.0.035. Affymetrix) for quality control analysis and genotype calling. Samples with a dish quality control (DQC) value > 0.82 and QC call rate > 0.97 threshold (following the “Best Practices Workflow” recommended by Affymetrix), were considered to have passed the quality control assessment. The quality control analysis classifies the SNPs into categories according to their clustering performance with respect to various Axiom-generated quality-control criteria; (i) ‘polymorphic high resolution’ where the SNP passes all QC, (ii) ‘monomorphic high resolution’ where the SNP passes all QC except the presence of a minor allele in two or more samples, (iii) ‘call rate below threshold’ where genotype call rate is under 97%, (iv) ‘no minor homozygote’ where the SNP passes all QC but only two clusters are observed, (v) ‘off-target variant’ (OTV) where atypical cluster properties arise from variants in the SNP flanking region, and (vi) ‘other’ where the SNP does not fall into any of the previous categories. For further analyses, only SNPs from categories (i) and (iv) were included and classified as “good quality”, as they are most likely to be reliable and informative SNPs.

Descriptive statistics and family assignment

Calculations of minor allele frequencies (MAF), levels of heterozygosity, discriminant analysis of principal components (DAPC), linkage disequilibrium and identity-by-state

(IBS) followed by multi-dimensional scaling (MDS) were carried out using Plink (Purcell *et al.* 2007), adegenet 1.3-1 package in R (Jombart and Ahmed 2011) and Genepop (Rousset 2008). Family assignment for the *C. gigas* families was performed using Cervus 3.07 (Kalinowski *et al.* 2007). Cervus assigns offspring to their parent pairs based on the pair-wise likelihood comparison approach generating locus-by-locus likelihood scores for each candidate parent for each offspring and assigns parentage to a candidate parent with the highest LOD score.

Data Availability

The Illumina sequencing data for the pooled *C. gigas* and *O. edulis* samples have been deposited into the European nucleotide archive (ENA) under accession number PRJEB20253 (http://www.ebi.ac.uk/ena/data/view/PRJEB20253). The details of the SNP markers on the array are given in File S1. *O. edulis* markers with significant alignment to the *C. gigas* genome (e-value 1E^{-4}) are given in File S2.

Results and discussion

Sequencing and SNP selection

To discover and prioritise SNPs for inclusion on the combined-species oyster SNP array, species-specific DNA sequencing, SNP discovery and filtering strategies were followed.

For *C. gigas*, WGS data aligned to the oyster genome identified 12.4 million putative SNPs across all populations. The 1,216,467 putative SNPs that passed the Affymetrix evaluation were subsequently filtered using the criteria described above to 40,625 putative SNPs that were submitted for the final Axiom MyDesign array. For *O. edulis*, 588,266 putative SNPs were identified, of which 100,403 putative SNPs were recommended at least for one strand by Affymetrix. Further filtering based on the criteria described above reduced the set to 19,215 putative SNPs that were submitted for array design and production.

The final array contained 40,625 putative SNPs from *C. gigas* and 14,950 putative SNPs from *O. edulis* to give a total of 55,575 putative SNPs assayed by a total of 111,360 probes. There were a greater number of *C. gigas* SNPs placed on the array than *O. edulis* due to the anticipated greater future use of the array for genome-wide association studies and genomic prediction for economically important traits in breeding programmes in this species. This includes an ongoing project to study host resistance to Oyster Herpes Virus based on genotyping samples collected from a large challenge experiment on oysters derived from Guernsey Sea Farm stocks. Nonetheless, it is anticipated that the ~15 K putative *O. edulis* SNPs will be widely applied for population and conservation genetics in future studies of this species.

Evaluation of the SNP array in C. gigas and O. edulis

The oyster array was evaluated in *C. gigas* by analysing the “validation populations” of 109 samples corresponding to eight distinct populations from France and UK (Table 2). All but one sample passed DQC and genotype call rate ≥ 97% threshold. The classification of SNPs according to their quality showed that 68.2 % of (n = 27,697) had probes classified as good quality (either ‘Poly High Resolution’ or ‘No Minor Hom’), which is similar to the percentage of informative markers obtained by the recently published *C. gigas* 134 K array (Qi *et al.* 2017). The MAF of these good quality SNPs (MAF > 0) in the combined 108 samples varied between 0.005 and 0.5 with a median of 0.18 (Table 2). From the 110 *O. edulis* samples genotyped (Table 3), two samples failed the DQC and genotype call rate ≥ 97% threshold, resulting in genotypes for 108 samples. A total of 74.6% of SNPs (n = 11,151) were classified as good quality as described above. The MAF of these good quality SNPs (combining all the 108 samples and SNPs with a MAF > 0) also varied between 0.005 and 0.5 with a median of 0.21 (Table 3).

Within-Population segregation of SNPs

The segregation of the SNPs was evaluated within each of the eight genotyped *C. gigas* population samples. From the 27,697 high quality SNPs defined across all population samples, the majority of SNPs (MAF > 0) were segregating within each of the populations (Figure S1), with an average of 22,486 SNPs segregating within each population, ranging from 20,141 (Hatchery 2) to 26,549(Guernsey) (Table 2). Among the UK populations (sampled from Guernsey, Maldon and Sea Salter), 19,613 SNPs were shared, while Guernsey had the highest number of exclusive SNPs (n = 2,373) (Figure S2). This is likely to be due to the fact that the Guernsey population was the most highly represented within the sequenced populations used for SNP discovery (Table 1) and the validation samples (Table 2), giving a greater chance of detecting rare minor alleles. Among all the five French populations, 13,855 SNPs were shared, with few SNPs segregating exclusively in particular populations (Figure S3). Finally, 11,997 common SNPs were segregating in all the eight populations from both France and the UK (Figure S4). The average MAF (for markers showing a MAF > 0) was 0.207 across all UK populations, while 0.214 across all French populations. Analysis of the distribution of MAF values for polymorphic SNPs (MAF > 0) showed that the highest number of SNPs are located within a MAF value range between 0.01 and 0.2 in all populations and decreasing in frequency when the MAF approaches 0.5 (Figure S5). A similar situation was observed by Lapègue *et al.* (2014), who found a high proportion of low MAF SNPs within *C. gigas* populations. Based on an additional test of the array on a small number of Australian *C. gigas* samples (data not shown), the number of segregating SNPs was similar, indicating that the array is likely to perform comparably for geographically diverse populations.

Table 2. Descriptive population genetic estimates for the sampled *C. gigas* populations included in the validation of the array.

		MAF > 0			
	sample N	*# SNPs*	*Average MAF*	*Ho*	*He*
UK (Combined)	**56**	**27,313**	**0.186**	**0.294**	**0.298**
GSF+Parents	38	26,549	0.19	0.308	0.304
Maldon	9	22,079	0.216	0.308	0.303
Sea Salter	9	22,821	0.214	0.317	0.302
Average within UK populations		*23,816*	*0.207*	*0.311*	*0.303*
France (Combined)	**52**	**26,891**	**0.182**	**0.240**	**0.254**
Ifremer	13	23,010	0.203	0.312	0.328
Hatchery 1	10	21,479	0.217	0.321	0.303
Hatchery 2	10	20,141	0.221	0.322	0.307
Hatchery 3	10	21,730	0.215	0.302	0.302
Hatchery 4	9	22,052	0.214	0.317	0.301
Average within French populations		*21,682*	*0.214*	*0.315*	*0.308*
All populations (Combined)	**108**	**27,697**	**0.182**	**0.268**	**0.283**

*Values in **bold** were obtained by the analysis of the combined dataset, not the average of the individual populations. Values in *italics* represent the within-population average,

From the 11,151 high quality SNPs segregating in the *O. edulis* populations, the average number of SNPs segregating (MAF > 0) in each population was 9,597. The samples from Croatia showed the lowest number of segregating SNPs (n = 8,474), while those from Foyle (IRL) showed the highest (n = 10,013) (see Table 3 & Figure S6). A total of 4,912 SNPs were shared between all (11) populations, with no particular population showing a high number of unique segregating SNPs. The average MAF value across the populations was 0.225, with Croatia showing the highest value of 0.234. Analysis of the distribution of MAF values for polymorphic SNPs (MAF > 0) showed that most populations have a large number of SNPs within a MAF value range between 0.05 and 0.2 with the exception of Croatia and Swansea that show a greater number of SNPs with a MAF higher than 0.1 (Figure S7).

The levels of genetic variability in terms of observed (Ho) and expected (He) heterozygosity (according to HWE) showed that most populations *(C. gigas and O. edulis)* had higher observed levels of heterozygosity than expected. Overall, no strong evidence of heterozygous deficiency was detected, in contrast to some previous studies that have described heterozygous deficiency in oysters and bivalves in general, albeit typically using a much lower number of microsatellites, SNPs, and allozymes (Appleyard and Ward 2006; English *et al.* 2000; Li *et al.* 2003; Sekino *et al.* 2003; Lapègue *et al.* 2014; Yu and Li 2007; Sobolewska and Beaumont 2005; Vercaemer *et al.* 2006). This discrepancy may be due to the fact that genome-wide SNP markers were used in the current study at a density not previously tested. In a larger-scale SNP-assay-based evaluation of the bivalve mollusc *Chlamys farreri*, no evidence for heterozygote deficiency was detected (Jiao *et al.* 2014). It is also possible that the strict filtering process led to SNPs on the array being enriched for stable genomic regions with lower levels of variation, while genomic regions with higher variability (and potentially more prone to null alleles) might have been discarded.

Table 3. Descriptive population genetic estimates for the sampled *O.edulis* populations included in the validation of the array.

		MAF > 0			
	sample N	*#SNPs*	*Average MAF*	*Ho*	*He*
Croatia	9	8,474	0.234	0.323	0.320
Foyle_IRL	10	10,013	0.224	0.319	0.311
Grevelingen_NLD	10	9,946	0.224	0.319	0.310
Larne_NIRL	10	8,927	0.231	0.354	0.316
Mersea_UK	10	9,980	0.224	0.318	0.310
Quiberon_FR	10	9,973	0.226	0.315	0.312
Rossmore_IRL	10	9,846	0.228	0.327	0.314
Sveio_NOR	10	9,118	0.226	0.322	0.313
Swansea_UK	9	9,696	0.224	0.319	0.311
Tralee_IRL	10	9,980	0.219	0.317	0.306
Maine_USA	10	9,614	0.221	0.317	0.305
Average within population		*9,597*	*0.225*	*0.323*	*0.312*
All populations (Combined)	**108**	**11,151**	**0.210**	**0.292**	**0.311**

Values in **bold** were obtained by the analysis of the combined dataset, not the average of the individual populations. Values in *italics* represent the within-population average.

Assessing population structure using Identify-by-state

The overall genetic similarity of any two samples can be evaluated by calculating average measures of identity-by-state (IBS) of the marker loci, which was then summarised using multidimensional scaling (MDS) to give indications of population (sub)structure (IBS clustering was also confirmed by DAPC analysis (data not shown)). There was some evidence of *C. gigas* samples according to their hatchery origin, and French hatchery populations tended to cluster separately to UK hatchery populations (Figure S8). The *O. edulis* samples were typically from 'wild' stocks from more diverse geographical locations than for the *C. gigas* samples. Accordingly, certain populations did show evidence of genetic differentiation, notably Croatia, Larne (Northern Ireland) and Sveio_(Norway) which are geographical outgroups (Figure 1 & Figure S10).Our results show evidence of a strong genetic similarity between Maine (USA), Sveio (Norway) and Grevelingen_(Netherlands) populations. Similarly, the origin of the Maine population has been linked to Netherlands (Loosanoff 1955; Vercaemer *et al.* 2006), Netherlands populations have been linked to Denmark's (Vera *et al.* 2016) and the genetic similarity between the Maine, Norway, Denmark and Netherland samples has also been observed using microsatellite markers (Mark McCullough, pers comm). A lack of population structure according to geographical original was observed in the other *O. edulis* population samples tested, for example the majority of samples from the coast of the UK and Ireland. This is consistent with existing evidence that suggests that marine organisms with larval stages (such as bivalves) often show low genetic differentiation (Li *et al.* 2015; Shabtay *et al.* 2014; Rohfritsch *et al.* 2013; Giantsis *et al.* 2014), with temporal factors rather than geographical factors often playing the major role in population structure. It is also possible that historical stock translocations might have also played an important role in the lack of genetic structure and admixture of the *O. edulis* populations (Bromley *et al.* 2016).

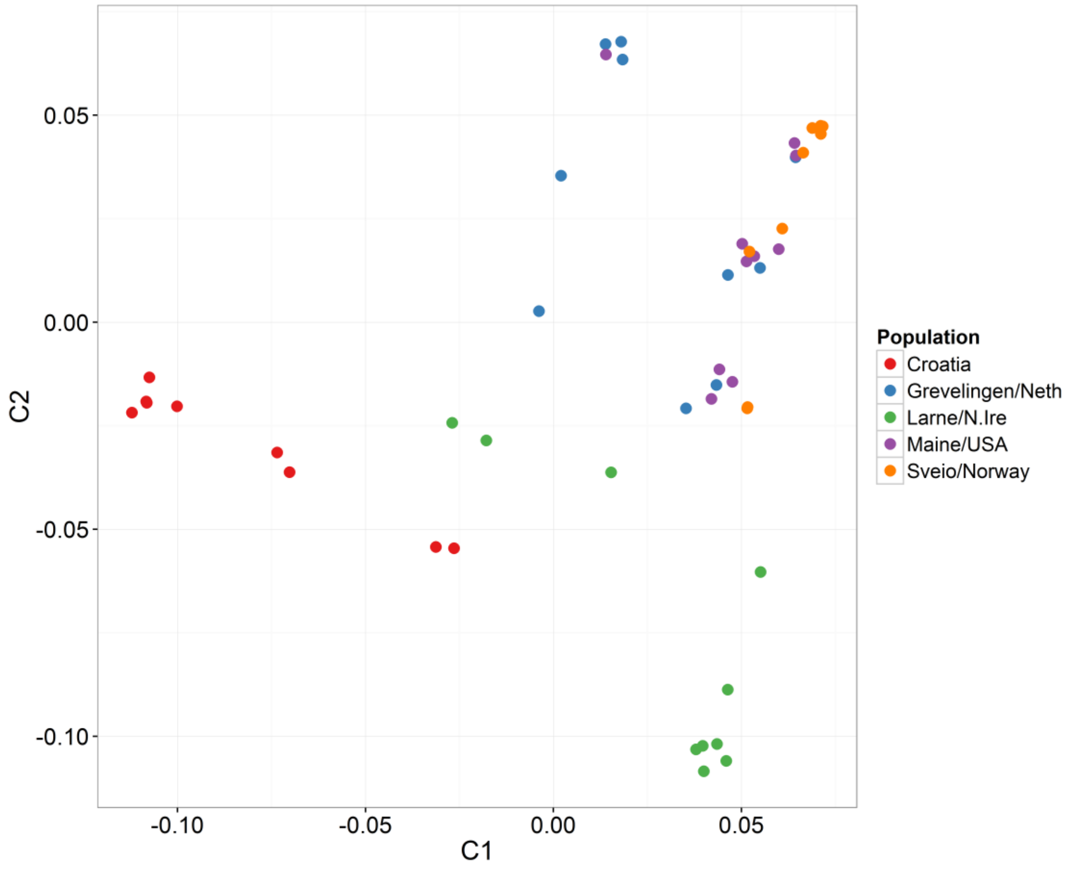

Figure 1. IBS clustering of selected *O. edulis* populations

Evaluation of the SNP array in pair crosses of C. gigas

Three pair crosses between Guernsey Sea Farms parents were created, reared separately and genotyped using the SNP array. Two of these nuclear families were half-siblings sharing a dam (F29 & F30). A total of 165 samples (161 offspring and their five parents) were genotyped. These families were analysed separately from the population samples used to validate the array described above. In part, this was due to the difficulty in obtaining high quality genomic DNA from the juvenile oysters. From the 165 samples, 139 passed the DQC and genotype call rate ≥ 97% threshold, resulting in a total of 25,629 SNPs which were classified as good quality in these families. The vast majority of SNPs showed stable Mendelian inheritance in all samples, although there was an average of 395 SNPs (~2% of total informative SNPs) with evidence for a Mendelian error per individual.

Since the offspring from each nuclear family were physically tracked throughout the experiment, such that their family structure was known *a priori*, the utility of the SNP array to differentiate between families was assessed using IBS clustering with MDS scaling. The MDS scaling plot based on IBS clustering clearly shows a clear separate cluster for each of the families, as shown in Figure 2. Interestingly, the clustering and separation of the three nuclear families was more obvious than for the population samples, even for populations from very distant geographical locations. Four individuals were distant to any of the family clusters, which may suggest incorrect pedigree assignment according to the physical animal tracking. Family assignment successfully assigned all the individuals to their correspondent parents using 3,000 randomly chosen SNPs, and confirmed that the four aforementioned individuals were not members of any of these three families. Microsatellites and SNP panels for parentage assignment have been described previously for oysters (Wang *et al.* 2010; Li *et al.* 2010; Lapègue *et al.* 2014; Jin *et al.* 2014). However, the successful parentage assignment in these physically tracked nuclear families, and the clear IBS-based differentiation of these families bodes well for the utility of this SNP array for high resolution genetic mapping studies and selective breeding programmes for oysters.

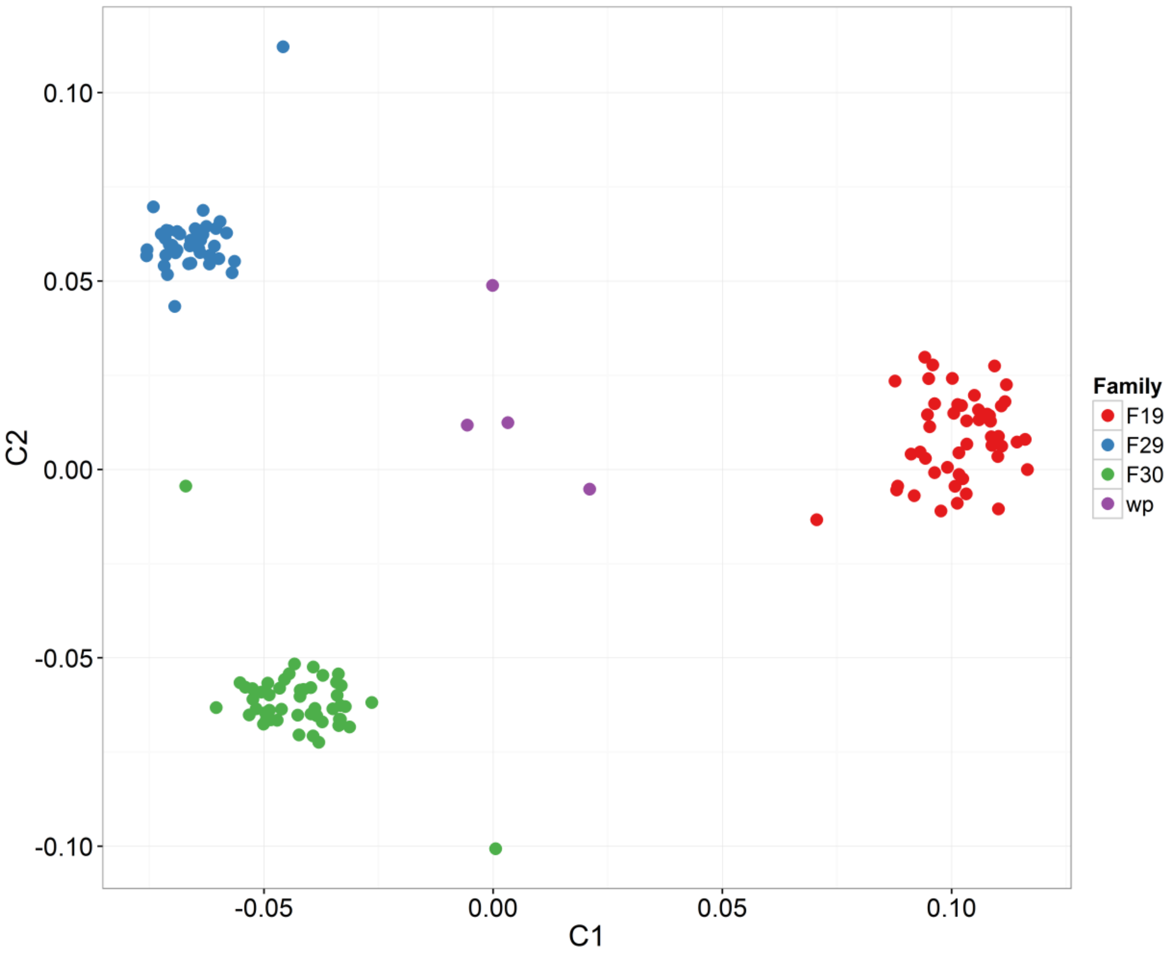

Figure 2. IBS-based clustering of the three nuclear *C. gigas* families. Samples in purple (wrong pedigree "wp") were not assigned to any of the three families.

Distribution of SNPs in the Pacific oyster genome

To assess the distribution of SNPs in the *C. gigas* genome (Zhang *et al.* 2012), SNPs were annotated according to the publicly available Ensembl oyster genome assembly (NCBI accession number: GCA_000297895.1). The oyster genome contains 7,658 scaffolds (N50 = 401,585) and 30,459 contigs (N50 = 31,239) and a total of ~ 558 Mb of assembled sequence. All 27,697 SNPs are mapped to the oyster genome according to BLAST alignment using their flanking region(s), with at least one SNP on 2,007 of the scaffolds, which in total covered 501 Mb (89.6 % of the total assembled genome sequence). The number of SNPs per scaffold was positively associated with scaffold length (Figure 3), with approximately one fifth of the scaffolds containing only one SNP. Additionally, harnessing the publicly-available oyster genome annotation (GCA_000297895.133), the SNPs on the array were grouped into putative positional and functional categories using SNPeff (Cingolani *et al.* 2012). A total of 14.6%, 13.1%, 18.7%, 17.6%, and 2.8% of the SNPs were located in intergenic, intron, downstream, upstream, and exon regions, respectively. The remaining SNPs (33%) were identified as transcript, splice site donor, splice site acceptor and splice site region.

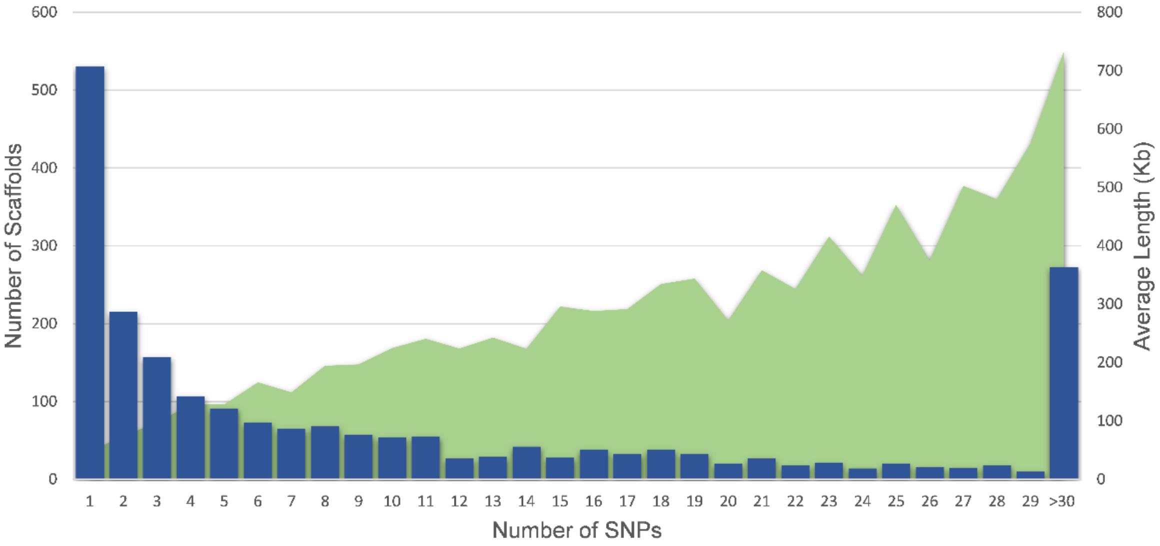

Figure 3. Distribution of SNPs on the *C. gigas* genome. Number of scaffolds containing SNPs (primary axis) and the average length of the scaffolds holding an increasing number of SNPs (secondary axis).

The extent of linkage disequilibrium (LD) between SNP pairs was assessed relative to their physical distance for the *C. gigas* populations. Pairwise r^2 was calculated using polymorphic SNPS with MAF ≥ 0.05 as shown in Table 2. The mean r^2 was calculated for every kilobase (Kb) and covering up to 500 Kb, according to the physical distance on the oyster genome assembly, as shown in Figure 4. In general, low levels of LD with slow decay with increasing physical distance were observed. The Guernsey and

Ifremer populations had lower levels of LD than the other populations. Although these LD levels are low compared to other aquaculture species such as carp or tilapia (Hong Xia *et al.* 2015; Xu *et al.* 2014), they are in accordance to recent reports describing low levels and short extent of LD in wild *C. gigas* populations (Zhong *et al.* 2017). Moreover, differences in LD levels between populations can be related to the divergence of these populations and the number of generations they have been bred in isolation, as observed in cattle (de Roos *et al.* 2008).

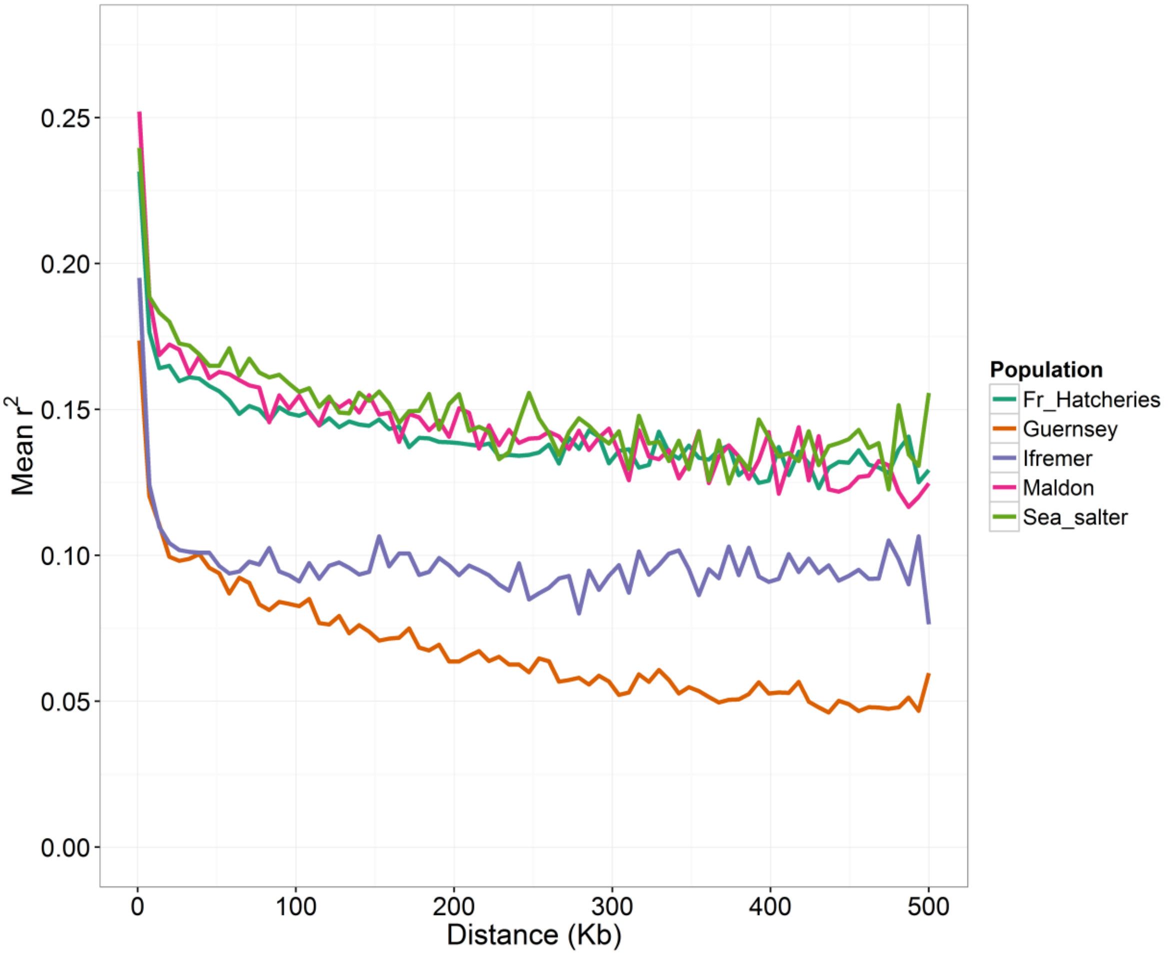

Figure 4. Decay of linkage disequilibrium (LD) with physical distance between markers among all the sampled *C. gigas* populations.

There was a higher extent and slower decay of LD in the three nuclear families, and LD levels were substantially higher than those observed in the (presumably unrelated) validation populations, as would be expected (Figure 4 & Figure 5). A lower effective population size (Ne) brings higher levels of kinship between individuals and therefore higher extent of LD (Sved 1971; Falconer and Mackay 1996).

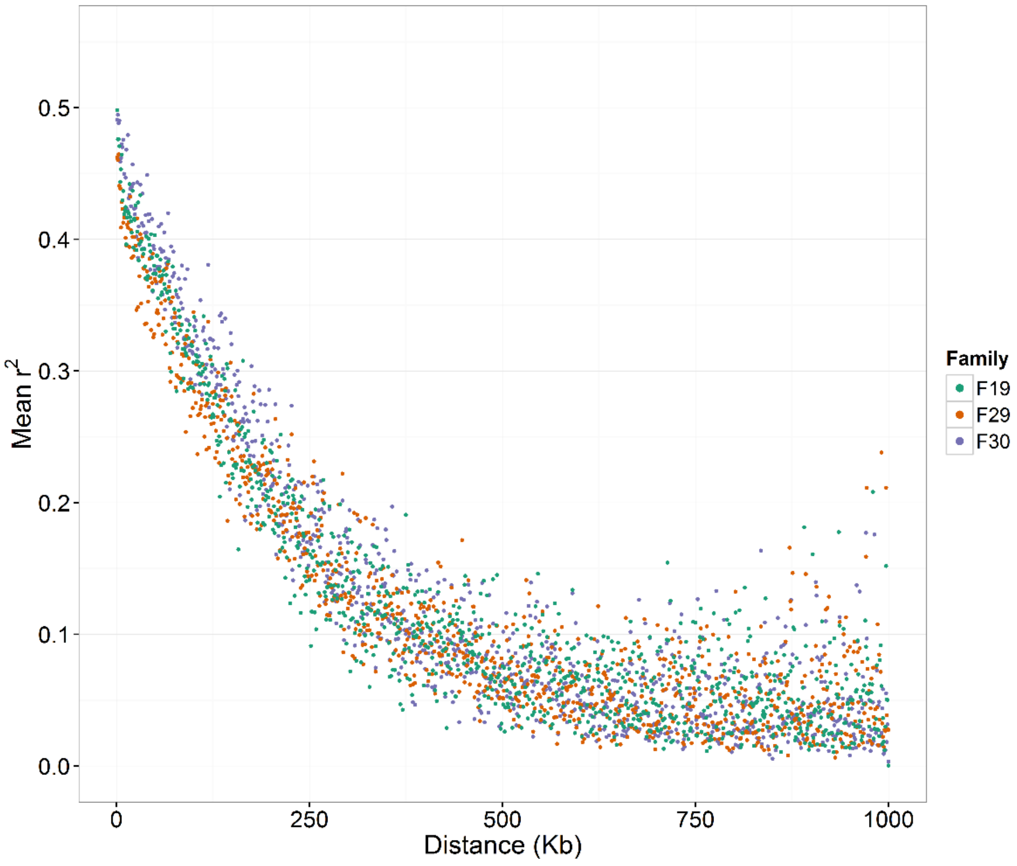

Figure 5. Decay of linkage disequilibrium (LD) among the three *C. gigas* families

Conclusions

This manuscript describes the development and analysis of a high density SNP array for two oyster species. A very large database of SNP markers was developed for both *C. gigas* using WGS, and *O. edulis* using RAD-Seq. Following extensive filtering, SNP assays for these two oyster species were combined on the array with 40,625 high quality SNPs for *C. gigas* and 14,950 for *O. edulis*. Testing of the array on genomic DNA samples from diverse locations revealed that the array contains a high number of SNPs that are shared between populations, and that the array can be applied to detect population and family structure. This oyster SNP array will be publicly available and will facilitate the study of important economic and ecological traits for these two oyster species, with possible applications for genomic selection, QTL mapping, evolutionary genetics and conservation programs.

Acknowledgements

The authors acknowledge Dr. Pierrick Haffray, Anastasia Bestin, Florian Enez and Anne-Sophie Tyran from SYSAAF, France for co-ordinating the provision of the French hatchery samples used for SNP discovery and genotyping and also to Lucie Buockellyoen from Marinove, Emilie Vetois from Satmar, Adeline Lange from France Naissain and Fiz Dacosta from Novostrea for gently agreeing to provide samples, Dr. Lionel Degremont from Ifremer, France for co-ordinating the provision of the Ifremer selection line samples, and Professor Dennis Hedgecock from the University of California, Davis for providing details of validated SNPs to include on the array. PAP and MMc were supported by funding from the European Union's INTERREG IVA Programme (project 2859 'IBIS') managed by the Special EU Programmes Body. The authors acknowledge funding for this work under the BBSRC-NERC UK Aquaculture Initiative (Grant Numbers: BB/M026140/1 and NE/P010695/1), and via BBSRC Institute Strategic Funding Grants to The Roslin Institute (BB/J004235/1, BB/J004324/1, BB/J004243/1), and via the Cefas Seedcorn fund.

References

Appleyard, S.A., and R.D. Ward, 2006 Genetic diversity and effective population size in mass selection lines of Pacific oyster (*Crassostrea gigas*). Aquaculture 254 (1–4):148-159.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3 (10):e3376.

Bromley, C., C. McGonigle, E.C. Ashton, and D. Roberts, 2016 Bad moves: Pros and cons of moving oysters – A case study of global translocations of *Ostrea edulis* Linnaeus, 1758 (Mollusca: Bivalvia). Ocean Coast. Manage. 122:103-115.

Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko, 2013 Stacks: an analysis tool set for population genomics. Mol. Ecol. 22 (11):3124-3140.

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6 (2):80-92.

Correa, K., J.P. Lhorente, M.E. López, L. Bassini, S. Naswa *et al.*, 2015 Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. BMC Genomics 16 (1):854.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet.12 (7):499-510.

de Roos, A.P.W., B.J. Hayes, R.J. Spelman, and M.E. Goddard, 2008 Linkage Disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics 179 (3):1503-1512.

English, L.J., G.B. Maguire, and R.D. Ward, 2000 Genetic variation of wild and hatchery populations of the Pacific oyster, *Crassostrea gigas* (Thunberg), in Australia. Aquaculture 187 (3–4):283-298.

Falconer, D., and T. Mackay, 1996 *Introduction to quantitative genetics*. Pearson, Harlow, UK.

FAO, 2015 *Food and Agriculture Organization Statistical Yearbook*.

Fleury, E., A. Huvet, C. Lelong, J. de Lorgeril, V. Boulo *et al.*, 2009 Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. BMC Genomics 10 (1):341.

Geng, X., J. Sha, S. Liu, L. Bao, J. Zhang *et al.*, 2015 A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. BMC Genomics 16 (1):196.

Giantsis, I.A., N. Mucci, E. Randi, T.J. Abatzopoulos, and A.P. Apostolidis, 2014 Microsatellite variation of mussels (*Mytilus galloprovincialis*) in central and eastern Mediterranean: genetic panmixia in the Aegean and the Ionian seas. J. Mar. Biol. Assoc. UK 94 (4):797-809.

Guo, X., Q. Li, Q.Z. Wang, and L.F. Kong, 2012 Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. Mar. Biotechnol. 14 (2):218-226.

Hedgecock, D., G. Shin, A.Y. Gracey, D.V. Den Berg, and M.P. Samanta, 2015 Second-generation linkage maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. G3 Genes Genom. Genet. 5 (10):2007-2019.

Hong Xia, J., Z. Bai, Z. Meng, Y. Zhang, L. Wang *et al.*, 2015 Signatures of selection in tilapia revealed by whole genome resequencing. Sci. Rep. 5:14168.

Houston, R.D., J.B. Taggart, T. Cézard, M. Bekaert, N.R. Lowe *et al.*, 2014 Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics 15 (1):90.

Hubert, S., and D. Hedgecock, 2004 Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics 168 (1):351.

Jiao, W., X. Fu, J. Li, L. Li, L. Feng *et al.*, 2014 Large-scale development of gene-associated single-nucleotide polymorphism markers for molluscan population genomic, comparative genomic, and genome-wide sssociation studies. DNA Res. 21 (2):183-193.

Jin, Y.-L., L.-F. Kong, H. Yu, and Q. Li, 2014 Development, inheritance and evaluation of 55 novel single nucleotide polymorphism markers for parentage assignment in the Pacific oyster (*Crassostrea gigas*). Genes Genom. 36 (2):129-141.

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

Kalinowski, S.T., M.L. Taper, and T.C. Marshall, 2007 Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16 (5):1099-1106.

Kofler, R., R.V. Pandey, and C. Schlötterer, 2011 PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27 (24):3435-3436.

Kranis, A., A.A. Gheyas, C. Boschiero, F. Turner, L. Yu *et al.*, 2013 Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14 (1):59.

Lallias, D., A.R. Beaumont, C.S. Haley, P. Boudry, S. Heurtebise *et al.*, 2007 A first-generation genetic linkage map of the European flat oyster *Ostrea edulis* (L.) based on AFLP and microsatellite markers. Anim. Genet. 38 (6):560-568.

Lallias, D., P. Boudry, S. Lapègue, J.W. King, and A.R. Beaumont, 2010 Strategies for the retention of high genetic variability in European flat oyster (*Ostrea edulis*) restoration programmes. Conserv. Genet. 11 (5):1899-1910.

Lallias, D., R. Stockdale, P. Boudry, A.R. Beaumont, and S. Lapègue, 2009 Characterization of 27 microsatellite loci in the European flat oyster *Ostrea edulis*. Mol. Ecol. Resour. 9 (3):960-963.

Lapègue, S., E. Harrang, S. Heurtebise, E. Flahauw, C. Donnadieu *et al.*, 2014 Development of SNP-genotyping arrays in two shellfish species. Mol. Ecol. Resour. 14 (4):820-830.

Li, G., S. Hubert, K. Bucklin, V. Ribes, and D. Hedgecock, 2003 Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. Mol. Ecol. Notes 3 (2):228-232.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25 (14):1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. Bioinformatics 25 (16):2078-2079.

Li, R., Q. Li, F. Cornette, L. Dégremont, and S. Lapègue, 2010 Development of four EST-SSR multiplex PCRs in the Pacific oyster (*Crassostrea gigas*) and their validation in parentage assignment. Aquaculture 310 (1–2):234-239.

Li, S., Q. Li, H. Yu, L. Kong, and S. Liu, 2015 Genetic variation and population structure of the Pacific oyster *Crassostrea gigas* in the northwestern Pacific inferred from mitochondrial COI sequences. Fisheries Sci. 81 (6):1071-1082.

Liu, S., L. Sun, Y. Li, F. Sun, Y. Jiang *et al.*, 2014 Development of the catfish 250K SNP array for genome-wide association studies. BMC Res. Notes 7 (1):135.

Loosanoff, V.L., 1955 The European oyster in American waters. Science 121 (3135):119-121.

Magoč, T., and S.L. Salzberg, 2011 FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27 (21):2957-2963.

Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1).

Matukumalli, L.K., C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan *et al.*, 2009 Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4 (4):e5350.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20 (9):1297-1303.

Ødegård, J., and T.H. Meuwissen, 2014 Identity-by-descent genomic selection using selective and sparse genotyping. Genet. Sel. Evol. 46 (1):3.

Palti, Y., G. Gao, S. Liu, M.P. Kent, S. Lien *et al.*, 2015 The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol. Ecol. Resour. 15 (3):662-672.

Pardo, B.G., J.A. Álvarez-Dios, A. Cao, A. Ramilo, A. Gómez-Tato *et al.*, 2016 Construction of an *Ostrea edulis* database from genomic and expressed sequence tags (ESTs) obtained from *Bonamia ostreae* infected haemocytes: Development of an immune-enriched oligo-microarray. Fish Shellfish Immunol. 59:331-344.

Peng, Y., H.C.M. Leung, S.M. Yiu, and F.Y.L. Chin, 2012 IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28 (11):1420-1428.

Pernet, F., C. Lupo, C. Bacher, and R.J. Whittington, 2016 Infectious diseases in oyster aquaculture require a new integrated approach. Philos. Trans. R. Soc. B Biol. Sci. 371 (1689).

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira *et al.*, 2007 PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81 (3):559-575.

Qi, H., K. Song, C. Li, W. Wang, B. Li *et al.*, 2017 Construction and evaluation of a high-density SNP array for the Pacific oyster *(Crassostrea gigas*). PLoS One 12 (3):e0174007.

Ramos, A.M., R.P.M.A. Crooijmans, N.A. Affara, A.J. Amaral, A.L. Archibald *et al.*, 2009 Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4 (8):e6524.

Richards, P.M., M.M. Liu, N. Lowe, J.W. Davey, M.L. Blaxter *et al.*, 2013 RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. Mol. Ecol. 22 (11):3077-3089.

Robledo, D., C. Palaiokostas, L. Bargelloni, P. Martínez, and R. Houston, 2017 Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquacult. 10.1111/raq.12193.

Rohfritsch, A., N. Bierne, P. Boudry, S. Heurtebise, F. Cornette *et al.*, 2013 Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. Evol. Appl. 6 (7):1064-1078.

Rousset, F., 2008 genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8 (1):103-106.

Sauvage, C., N. Bierne, S. Lapègue, and P. Boudry, 2007 Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. Gene 406 (1–2):13-22.

Sauvage, C., P. Boudry, D.J. De Koning, C.S. Haley, S. Heurtebise *et al.*, 2010 QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). Anim. Genet. 41 (4):390-399.

Sekino, M., M. Hamaguchi, F. Aranishi, and K. Okoshi, 2003 Development of novel microsatellite DNA markers from the Pacific oyster *Crassostrea gigas*. Mar. Biotechnol. 5 (3):227-233.

Shabtay, A., Y. Tikochinski, Y. Benayahu, and G. Rilov, 2014 Preliminary data on the genetic structure of a highly successful invading population of oyster suggesting its establishment dynamics in the Levant. Mar. Biol. Res. 10 (4):407-415.

Sobolewska, H., and A.R. Beaumont, 2005 Genetic variation at microsatellite loci in northern populations of the European flat oyster (*Ostrea edulis*). J. Mar. Biol. Assoc. UK 85 (04):955-960.

Sved, J.A., 1971 Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2 (2):125-141.

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. J. Sea Res. 64 (3):145-165.

Tsai, H.-Y., A. Hamilton, A.E. Tinch, D.R. Guy, J.E. Bron *et al.*, 2016 Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genet. Sel. Evol. 48 (1):47.

Tsai, H.-Y., A. Hamilton, A.E. Tinch, D.R. Guy, K. Gharbi *et al.*, 2015 Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. BMC Genomics 16 (1):969.

Vallejo, R.L., T.D. Leeds, B.O. Fragomeni, G. Gao, A.G. Hernandez *et al.*, 2016 Evaluation of Genome-Enabled Selection for Bacterial Cold Water Disease Resistance Using Progeny Performance Data in Rainbow Trout: Insights on Genotyping Methods and Genomic Prediction Models. Front. Genet. 7 (96).

Vera, M., J. Carlsson, J.E. Carlsson, T. Cross, S. Lynch *et al.*, 2016 Current genetic status, temporal stability and structure of the remnant wild European flat oyster populations: conservation and restoring implications. Mar. Biol. 163 (12):239.

Vercaemer, B., K.R. Spence, C.M. Herbinger, S. Lapègue, and E.L. Kenchington, 2006 Genetic diversity of the european oyster (*Ostrea edulis* L.) in nova scotia: Comparison with other parts of canada, maine and europe and implications for broodstock management. J. Shellfish Res. 25 (2):543-551.

Wang, J., H. Qi, L. Li, H. Que, D. Wang *et al.*, 2015 Discovery and validation of genic single nucleotide polymorphisms in the Pacific oyster *Crassostrea gigas*. Mol. Ecol. Resour. 15 (1):123-135.

Wang, Y., X. Wang, A. Wang, and X. Guo, 2010 A 16-microsatellite multiplex assay for parentage assignment in the eastern oyster (*Crassostrea virginica* Gmelin). Aquaculture 308, Supplement 1:S28-S33.

Xu, J., Z. Zhao, X. Zhang, X. Zheng, J. Li *et al.*, 2014 Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). BMC Genomics 15 (1):307.

Yáñez, J.M., S. Naswa, M.E. López, L. Bassini, K. Correa *et al.*, 2016 Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. Mol. Ecol. Resour. 16 (4):1002-1011.

Yu, H., and Q. Li, 2007 Genetic variation of wild and hatchery populations of the Pacific oyster Crassostrea gigas assessed by microsatellite markers. J. Genet. Genomics 34 (12):1114-1122.

Zhang, G., X. Fang, X. Guo, L. Li, R. Luo *et al.*, 2012 The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490 (7418):49-54.

Zhong, X., Q. Li, X. Guo, H. Yu, and L. Kong, 2014 QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. Aquac. Int. 22 (6):1877-1889.

Zhong, X., Q. Li, L. Kong, and H. Yu, 2017 Estimates of linkage disequilibrium and effective population size in wild and selected populations of the Pacific oyster using single-nucleotide polymorphism markers. J. World. Aquac. Soc. 10.1111/jwas.12393.