Aquaculture Reports 20 (2021) 100687

Contents lists available at ScienceDirect

Aquaculture Reports

journal homepage: www.elsevier.com/locate/aqrep

A SNP parentage assignment panel for the silver lipped pearl oyster (*Pinctada maxima*)

C. Massault [a,*], D.B. Jones [a], K.R. Zenger [a], J.M. Strugnell [a], R. Barnard [b], D.R. Jerry [a,c]

[a] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville, Australia
[b] RMB Aqua Pty Ltd, Perth, Australia
[c] Tropical Futures Institute, James Cook University, Singapore

ARTICLE INFO

Keywords:
Silver-lipped pearl oyster (*Pinctada maxima*)
Pedigree reconstruction
SNP panel
Breeding program

ABSTRACT

The silver-lipped pearl oyster, *Pinctada maxima*, is an important aquaculture species extensively farmed in tropical Australia and Southeast Asia to produce "South Sea" pearls. The pearling industry in Australia is in the process of implementing breeding programs targeting improved disease resistance, growth and pearl quality, and as a result genetic tools that can be readily implemented into breeding programs have been developed. However, to date there is no routine single nucleotide polymorphism (SNP) pedigree reconstruction tool publicly available.

In this study, we isolated and assessed a panel of 934 genome-wide DArTcap™ SNPs for their utility and discriminatory power in resolving pedigree relationships in *P. maxima*. The panel was assessed for its power to assign parentage through *in silico* simulations of 1,000 progeny in a mass-spawning scenario, based on genotypes of 100 broodstock (1:1 sex ratio) and allowing for random missing genotypes in individuals (0–40%). Assignment success was then assessed across various pedigree panel sizes (25–934 SNPs ranked from the highest minor allele frequency). Simulations showed that a SNP panel comprising 50 or more SNPs had high pedigree resolution and the power to correctly assign progeny to a parent-progeny trio with 95 % confidence. In the simulation examining the effect of missing genotypes on assignment success, even where 40 % of alleles were missing in individuals only 150 SNPs were required in the panel to accurately assign parentage. Progeny from a commercial hatchery mass spawn involving 184 candidate parents were also assigned to their parents using the SNP panel and identified 79 full-sib families and genetic contributions from 33 sires and 11 dams. A highly skewed representation of family distributions in the cohort was observed which highlights the challenge of using a mass-spawning approach to the capture and identification of genetic diversity in *P. maxima* breeding programs.

1. Introduction

The Australian South Sea pearl, produced by the silver-lipped pearl oyster (*Pinctada maxima*) (Jameson, 1901), is a highly coveted luxury item due to its large size (>15 mm), high quality and silver-white colouration. While once Australia's most valuable aquaculture export sector, the South Sea pearling industry has been in decline over the last decade due to both economic and disease impacts. The worldwide demand for pearls dropped dramatically during the Global Financial Crisis (GFC) in 2008 and has struggled to regain its popularity since (DPIRD, 2016). Additionally, the Australian industry has been exposed to periodic and severe mass-mortality events caused by an unknown agent which has the potential to extirpate entire cohorts of juvenile oysters (such as juvenile pearl oyster mortality syndrome (JPOMS); Massault et al., 2019, or Oyster Oedema disease (OOD); Goncalves et al., 2017). As a result, the Australian pearling industry has been in decline over the last decade with its farm-gate value dropping from ~$122 million in 2006–2007, to $70 million in 2017–2018 (ABARES, 2018).

To increase productivity and lower the risk of disease, several companies within the Australian pearling industry have turned to implementation of selective breeding programs targeting tolerance to JPOMS, improved pearl quality and increased oyster growth. While in many cases these breeding programs are nascent, they offer the industry the potential to transition from relying on unimproved stocks to oysters with better commercial characteristics for farming and that align the pearl to market trends. Consequently, in recent years, there have been several studies conducted on *P. maxima* to understand the genetics behind pearl production. These included the development of microsatellite DNA

* Corresponding author.
E-mail address: cecile.massault@jcu.edu.au (C. Massault).

https://doi.org/10.1016/j.aqrep.2021.100687
Received 17 December 2020; Received in revised form 31 March 2021; Accepted 1 April 2021
Available online 8 April 2021

markers for parentage determination and population genetic analyses (Evans et al., 2006; Lind et al., 2007, 2009), investigations of family by environment and genotype by environment interactions for growth (Kvingedal et al., 2008, 2010), use of xenografts to understand transcriptomic control of biomineralisation and pearl colour (McGinty et al., 2010, 2011, 2012), heritability of growth and pearl quality traits (Jerry et al., 2012; Kvingedal et al., 2010), and using dense single nucleotide polymorphisms (SNPs) to construct genetic linkage maps, quantitative trait loci (QTL) mapping and genome wide association studies (GWAS; Jones et al., 2013a, 2013b, 2014a, 2014b).

While some of the genetic tools are now available to implement selective breeding in the species, one very important tool that has not yet been refined is a routine, cost-effective and industrial-scale relevant SNP panel for parentage assignment. The traditional silver-lipped pearl oyster mode of reproduction is mass-spawning, necessitating a group of males and females to be put into the one tank to ensure many families are produced (Southgate and Lucas, 2008; Lind et al., 2009). While this mass-spawn approach yields commercial numbers of larvae for stocking it results in larvae not being able to be traced to their individual parents without DNA parentage analysis. Single-pair spawning has been achieved and also practiced (Jones et al., 2014a, 2014b) in an effort to try to retain pedigree of larvae; however, it is inherently unreliable and thus the use of traditional mass-spawning is still the main reproductive strategy used in the industry to create commercial seedstock. Additionally, due to infrastructure constraints, oysters from different spawns, or family cohorts, are often pooled for ease of culture, or their pedigree lost due to shuffling oysters on long-lines during cleaning for biofouling control or reallocating to empty panel nets. This results at the time of recording traits and subsequent spawning in uncertainties of the pedigree of individual oysters. Knowing the parentage of measured individuals is necessary information to perform estimation of genetic parameters of important economical traits, such as growth or pearl quality, and in the calculation of estimated breeding values (EBVs), two important steps in breeding programs.

A common practice in aquaculture to obtain parentage information (pedigree) is to use molecular markers such as SNPs. The advantage of SNPs is the availability of large number of markers scattered throughout the genome and the relatively low number required to perform reliable parentage assignment (Yue and Xia, 2014). SNPs have been used successfully for pedigree reconstruction in a number of bivalves: Pacific oysters (Jin et al., 2014), eastern oysters (Thonga et al. 2018), European abalone (Harney et al., 2018) and blue mussel (Nguyen et al., 2014). In this study, we developed a robust SNP marker panel using DArTcap™ genotype-by-sequencing (GBS) for the purpose of routine pedigree assignment in the context of selective breeding programs for *P. maxima*. The panel was designed using a refined set of SNPs developed from a prior DArTseq™ genotype-by-sequencing effort using *P. maxima* originating from Indonesian populations (unpublished data). This set of SNPs was validated as useful for Australian *P. maxima* and its power to assign pedigree validated using the parentage analysis software program CERVUS 3.0 (Kalinowski et al., 2007).

2. Materials and methods

2.1. SNP discovery through DArTseq™ and DArTcap™ genotype by sequencing

The initial discovery of genome-wide SNPs was undertaken using 156 wild adult individuals collected from three locations across Indonesia [Aru (N = 63; 6.43′S, 134.63′E); Bali (N = 59; 8.32′S, 114.92′E); and West Papua (N = 34; 1.13′N, 130.54′E], as well as 756 F1 broodstock sourced from the same locations within Indonesia (as described in Jerry et al., 2012 and Jones et al., 2014a). Mantle or foot tissue was collected from individuals and stored in 70 % ethanol until DNA was extracted using a modified CTAB-chloroform DNA extraction protocol. DNA was standardised to 50 ng/μl and visualised on a 0.8 % agarose gel to ensure it was of sufficient quality. All 912 DNA samples were submitted for DArTseq™ genotyping through Diversity Arrays Technologies, Canberra, following procedures outlined in Sansaloni et al. (2011). Briefly, DArTseq™ employed two enzymes (PstI and SphI for *P. maxima*) in a restriction digest complexity reduction approach before libraries were pooled, sequenced, de-multiplexed and genotypes were called within KDCompute™ (Kilian et al., 2012). This resulted in the identification of 16,205 novel SNPs.

As an extension of the DArTseq™ GBS platform, DArTcap™ allows capture of specific SNPs using baits probes (e.g. MYbaits®) from DNA samples that have been through DArTseq™ complexity reduction steps (unpublished modifications to Sansaloni et al., 2011), resulting in a more targeted and cost-effective GBS method (Guppy et al., 2020). To develop a DArTcap™ SNP panel for *P. maxima*, a total of 3,300 high-quality and genome-wide SNPs were selected from the 16,205 DArTseq™ SNPs using a custom filtering script (https://github.com/esteinig/dartqc). Selection was based on the following filtering parameters (Table 1.A): individual genotype read count of > 5, minor allele frequency (MAF) of > 0.02, average repeatability of > 0.9, SNP call rate of > 0.8, and finally, CD-HIT clustering (Fu et al., 2012) of clone sequences were conducted to select only one SNP per sequence cluster based on highest MAF.

2.2. DArTcap™ genotyping of Australian broodstock

The objective of this study was to develop a SNP panel for the silver-lipped pearl oyster, *P. maxima,* that can be routinely applied to assign pedigree in breeding programs within Australia. To validate the DArTcap™ SNPs within Australian populations, 1,185 individual Australian adult oysters were genotyped on the developed DArTcap™ array to ensure suitability of SNPs for parentage assignment (source of samples were 231 wild oysters originally collected for pearl implantation from Eighty Mile Beach, Western Australia (-19. 05 S, 121.52 E 134.63′E); 876 1st generation hatchery-bred; and 78 oysters with either farmed or wild, but with unknown heritage) using extracted DNA from foot tissue stored in 100 % ethanol. By utilising the DArTcap™ array, a total of 5,874 SNPs were successfully genotyped across samples. This included 2,304 SNPs (69.8 %) from the 3,300 SNPs submitted for the design of the DArTcap™ array, plus an additional 3,570 non-target SNPs within the captured sequence which were detected due to the nature of the DArTcap™ technology and probe-based capture techniques (Guppy et al., 2020).

To ensure the high-quality and repeatability of the SNPs selected for the parentage panel in Australian *P. maxima*, SNP data were again filtered based on their read counts, number of individuals successfully genotyped for a SNP, clustering (retaining only one SNP per 72 bp sequence fragment), minor allele frequency (MAF) and average repeatability (removing SNPs that are not identical between replicated samples in at least 99 % of the replicated samples (327 (28 %) of the 1185 individuals were genotyped twice to ensure high data integrity) (see Table 1.B for description of filtering metrics of DArTcap™). To select the most high-valued SNPs, we excluded SNPs that did not conform to HWE (p value <0.001). The selected SNPs would therefore be unlikely to be under putative selection. Finally, remaining SNPs were ranked by MAF from highest to lowest.

2.3. Spawning and larvae sampling

In March 2018, a commercial mass-spawning was conducted on farm (Elizabeth Bay - 11.90 °S, 136.55 °E) based on a pool of 87 male and 97 female broodstock which were allowed to spawn naturally together in a single spawning tank. Larvae from the mass-spawn were reared in the hatchery via standard industry practices and allowed to settle on frames as spat, before deployment in the ocean at approximately 1.5 months of age. At the time of ocean deployment, 923 spat were collected, preserved in 100 % ethanol and then sent to Diversity Arrays Technologies to be genotyped using the DArTcap™ platform as described above. Foot

Table 1
Quality control process used A – to filter DartSeqTM SNPs from 16,205 SNPs to 3,300 high quality SNPs and B- to filter 5,874 SNPs derived from DArTcap™to obtain 934 high-quality SNPs suitable for parentage assignment.

A							
Step		1	2	3	4	5	6
Filter		Read count[1] > 5	MAF[4] > 0.02	Replication average[5] > 0.9	SNP call rate[2] > 0.8	Cluster CD HIT	Top ranked MAF
# SNP remaining	16,205	16,205	14,216	14,210	8,285	5,543	3,300

B							
Step		1	2	3	4	5	6
Filter		Read count[1] > 7	SNP call rate[2] > 0.7	Cluster[3]	MAF[4] > 0.01	Replication average[5] > 0.99	HWE[6] > 0.0001
# SNP remaining	5,874	5,874	3,597	2,247	2,020	2,004	934

[1] Retain individuals genotype calls if more than 5 (DartSeqTM) and more than 7 (DartCapTM).
[2] Retain SNP if proportion of individuals successfully genotyped for that SNP is larger than 0.8 for DartSeqTM and 0.7 for DartCapTM.
[3] Only retain 72 bp sequence.
[4] Retain SNP with MAF > 0.02 for DartSeqTM and 0.01 for DartCapTM.
[5] Retain SNP that have the same genotype for duplicated individuals in at least 90 % of duplicated individuals in DartSeqTM and at least 99 % duplicated individuals in DartCapTM.
[6] Retain SNP that have a Hardy Weinberg equilibrium test p-value over 0.0001.

muscle tissue was also excised from each potential broodstock oyster involved in the mass spawn, preserved in 80 % ethanol, and also sent to Diversity Arrays Technologies for DArTcap™ genotyping. Due to uncertainties on which parents had actually spawned all possible parents were genotyped and included in the parentage analysis. Prior to parentage assignment, quality control on the progeny SNP's was performed as outlined in 2.2 above.

2.4. Simulated power analysis and parentage assignment

To evaluate the power of the DArTcap™ SNPs to assign parentage a three-step validation process was undertaken; i) progeny genotypes were simulated against increasing numbers of SNPs derived from the DArTcap™ panel to evaluate parentage assignment power of the marker suite based on the number of SNP's used when true parent-progeny relationships are known – this simulation was undertaken to explore the potential to design a smaller, direct-targeted SNP panel for parentage determination and what would be the minimum numbers of SNPs required to ensure successful assignment; ii) GBS approaches such as DArTcap™often result in missing SNP genotypes among runs due to quality of DNA, filtering processes and/or sequencing technical implications; thus the simulated impacts of various proportions of missing genotypes on parentage assignment success were evaluated, and; iii) a *combined* simulation was conducted involving; a) *in silico* generated progeny with known broodstock parents and b) field genotyping of pearl oyster spat based on a hatchery cohort produced from a commercial mass-spawn where actual true parent-progeny relationships were unknown. In this last component a pool of progeny based on the same broodstock genotypes as involved in the spawn were created *in silico* as a benchmark to determine if parentage assignments of the commercially hatchery bred spat were likely correct. The methodologies for each of these components were as indicated below;

i *Simulated power of parentage assignment based on an increasing number of SNPs in each panel.* A pool of 1,000 simulated offspring were created by mating 50 randomly chosen sire and dam parents (sex ratio 1:1) *in silico* based on parental genotypes (Section 2.3). Each sire was mated to a single dam and produced a family of 20 progeny, each progeny with a recorded pedigree. There were no missing genotypes in this scenario. Parentage assignment was undertaken for various SNP panel sizes ranging from 25 to 934 SNPs (25, 50, 75, 100, 150, 200, 300, 400, 500 and 934 SNPs, ranked from highest MAF to lowest as panel size increased). To ensure results were not driven by a single simulation, creation of *in silico* offspring genotypes and subsequent parentage assignment was replicated five times for each panel size. To assess the power of parentage assignment for each panel size, the percentage of offspring that were assigned by software (CERVUS 3.0; Kalinowski et al., 2007) to candidate parents with 95 % confidence (described below) and for which the assigned parents corresponded to the actual parents in the *in silico* generated pedigree were considered as being a correct parentage assignment.

ii *Simulated power of parentage assignment dependent on missing genotype proportion.* To assess the impact of missing genotypes on parentage assignment, given that GBS genotype data often involves analyses with missing SNP calls, a second parentage assignment simulation was undertaken where the progeny were simulated in the same way as step (i), but the offspring were divided into five groups, equally represented within each family (four progeny of each group per family), with different levels of missing genotypes (0%, 10 %, 20 %, 30 % and 40 % randomly missing genotypes among individuals). Parentage assignment involving missing data was conducted for the various panel sizes from 25 to 934 SNPs (same as in step (i)). The minimum number of SNPs that were necessary to correctly assign at least 95 % of progeny to a parent-progeny trio, sire and/or dam correctly for each missing genotype category was recorded.

iii *Assignment of hatchery produced progeny* - To ascertain how the parentage panel may perform to assign offspring originating from a real mass-spawning event, parentage assignment was conducted for the 923 spat produced from the commercial mass-spawning described in Section 2.3. After quality-control filtering of the DArTcap™ data, 648 SNPs passed quality control filtering and were polymorphic in the spat genotyped in this batch of progeny; therefore 648 SNPs (not 924 SNPs as above) were used as the maximum panel size for parentage assignment in this third analysis. As the true parents to any of the mass-spawned hatchery spat were unknown, 250 *in silico* generated offspring from randomly chosen broodstock involved in the mass spawn were added to the offspring analysis pool as a quality assurance step to evaluate the confidence that broodstock assigned to the spat were correctly assigned (i.e., simulated offspring were included in the parentage analysis to serve as internal controls to evaluate correct assignment success of progeny to broodstock at a level of 95 % confidence). Parentage analysis was performed based on a step-wise process whereby in each analysis the number of SNPs used increased from a panel size of 25–648 SNPs (ranked highest MAF to lowest). Success of parentage assignment of spat was determined solely based on the 95 % confidence level given by Cervus 3.0, as pedigree for individual spat was not recorded.

CERVUS 3.0 (Kalinowski et al., 2007) (referred hereon as CERVUS) was used to perform all parentage analyses. CERVUS calculates the pairwise parentage likelihood of each offspring with each possible dam, sire and parent-pair trio at each SNP and gives an overall log order

difference (LOD) score against these three parental types. It then assigns the sire, dam and parent-progeny trio with the highest LOD score to the offspring. CERVUS also provides a level of statistical confidence for each assignment (i.e. 95 % confidence, 80 % confidence, most likely parents, or no label) by simulating parent-progeny relationships and determining LOD scores between the two most likely parents to determine confidence level of assignment (Kalinowski et al., 2007). For analyses in the current study, progeny were only considered correctly assigned to broodstock parents when the confidence of assignment was 95 %, or greater. For all simulations, it was expected that all parents that have contributed to the offspring were genotyped and included in parentage assignment. As reconstructed pedigrees obtained with 1% and 5% genotyping error were practically the same (unpublished data), the more conservative 1% genotyping error rate was applied to all analyses. Additionally, a simulation of parentage assignment using the 87 sires and 97 dams and 10,000 offspring was conducted in CERVUS to determine the acceptable threshold for statistical confidence of LOD scores. Finally, SNPs with genotype data for less than 70 % of the offspring were removed, along with individuals with less than 50 % of SNPs genotyped.

2.5. Utility of SNPs

To evaluate if the SNPs chosen for the parentage assignment panel were also suitable to elucidate fine-scale family-level population genetic structure, pairwise genetic distance between all hatchery-bred spat were calculated. The NetView framework described by Neuditschko et al. (2012), first implemented in Python (Steinig et al., 2016) and now available in R (https://github.com/esteinig/netview) was then used to construct a relationship network using the k-nearest neighbour (knn) algorithm. Using a specified knn, Netview displays the connectivity between individuals by clustering them according to their genetic similarity in genotype and helps visualise fine-scale relationships among pairs of individuals.

3. Results

3.1. DArTseq™ and DArTcap™ SNP filtering

The number of SNPs remaining after each quality control filtering step for both DArTseq™ and DArTcap™ are provided in Fig. 1. Briefly, for the DArTseq™ dataset, filtering for low read counts less than five removed 2,583,670 (17.26 %) of genotype calls. A series of filters were applied to ensure the resulting SNPs were of higher quality. A total of 5,543 SNPs were identified with MAF > 0.02, replication average > 0.9, call rates > 0.8 and did not originate from sequences that clustered together (Table 1.A). As stated in the methods, the top ranked 3,300 SNPs were put forward for DArTcap design.

For the DArTcap™ dataset consisting of 5,874 SNPs (2,304 of the 3,300 submitted SNPs detected, as well as an additional 3,570 non-target SNPs), filtering for low read counts less than seven removed

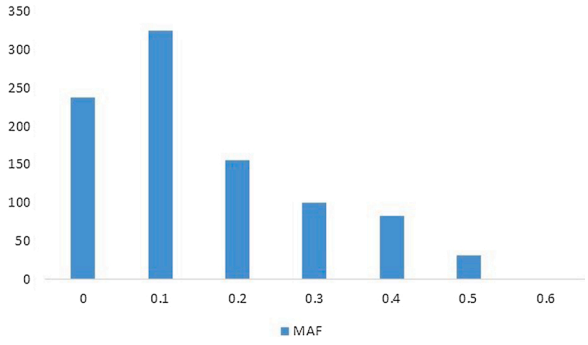

Fig. 1. Distribution of minor allele frequency (MAF) of 934 SNPs after quality control filtering and removal if not in Hardy Weinberg equilibrium.

745,746 genotype calls (1.7 %). Table 1.B described the filtering steps and the remaining step after each filtering procedure: 3597 SNPs remained after removing SNPs with call rate below 0.7. With the next step, keeping only one SNP within a 72 kb sequence, a further 1350 SNPs were removed. Then, only SNPs with MAF > 0.01 were kept – 2020 SNPs. Sixteen more SNPs were removed when applying the replication average filter (step 5). Therefore a total of 3,870 SNPs were removed, as they did not pass the quality control in steps 2–5, with 2,004 SNPs remaining. Exclusion of SNPs that deviated significantly from HWE (step 6) returned a total of 934 high-quality SNPs present in genotyped oysters. MAF was calculated for all the 934 SNPs, which were ordered from highest to lowest (Fig. 1). No evidence of linkage disequilibrium was observed between the 934 retained SNPs (test linkage disequilibrium between 2 loci with p-value > 0.05 for all pairs, proposed by Weir (1996)).

3.2. Power of assignment with simulated offspring without and with missing genotypes

Assignment of *in silico* offspring to their parent-progeny trio, sire and dam, according to the number of SNPs used in the panel SNPs (25, 50, 75, 100, 150, 200, 300, 400, 500 and 934), is shown in Fig. 2. Simulations indicate that use of less than 50 of the DArTcap™ SNPs with highest MAF results in a high error rate of assignment (based on analyses looking at parent-progeny trios). When 25 SNPs are used a mean of 39.3 % of offspring were correctly assigned to their known parents, with wide ranges of assignment success evident among replicate analyses. When 50 SNPs were evaluated, assignment success increased to a mean of 95.6 % of progeny correctly assigned. As the number of SNPs used in the panel further increased, assignment success for parent-progeny trios only marginally increased from that based on 50 SNPs to achieve a mean maximum of 98 % (± 0.9–1.5% SE) of offspring correctly assigned (from 100 SNPs and greater). Compared to the power of assigning offspring to the correct parent-progeny trio, when assigning offspring to only sire or dam a larger number of SNPs were initially required, with SNP panels of ≥75 SNPs required to assign 95 % of offspring to their correct paternal or maternal parent with high confidence (Fig. 2).

Missing SNP genotypes can influence the power of parentage assignment. To test the effect of missing genotypes we simulated how varying numbers of randomly missing genotypes in offspring influenced assignment success. Fig. 3 indicates the minimum number of SNPs required in the panel in order to assign parent-progeny trio, sire and dam correctly with 95 % confidence to at least 95 % of offspring as a function of the proportion of the missing genotypes. The number of SNPs required increases as the proportion of missing genotypes in the progeny increases as expected; however, the number of SNPs required did not increase in the same way for parent-progeny trio, sire and dam. For instance, a panel of 50 SNPs are required to accurately assign parent-progeny trio for at least 95 % of the offspring with 0 and 10 % missing genotypes, and at least 75 SNPs for 20 %, 30 % and 40 % missing genotypes. However, a minimum of 75 SNPs were required to assign progeny to sire and dam for 0–20 % missing genotypes, and a panel size of 100–150 SNPs for 30–40 % missing SNPs, respectively.

3.3. Spat parentage assignment

From the 923 spat genotyped, three spat had more than 50 % missing SNP calls and were excluded from subsequent parentage analysis; Fig. 4 shows the mean percentage of assignment for the spat and the 250 simulated offspring for parent-progeny trio, sire and dam assignments over the five replicates from using 25 to 934 SNPs. When considering parent-progeny trio assignment, simulated progeny showed that a panel size of ≥100 SNPs provided 100 % correct parental assignment. However, interestingly assignment of the hatchery-bred spat at 95 % confidence was only 66 %. Increasing the number of SNPs used in the panel from 100 to 300 SNPs increased progeny assignment from 66 % to 80 %;

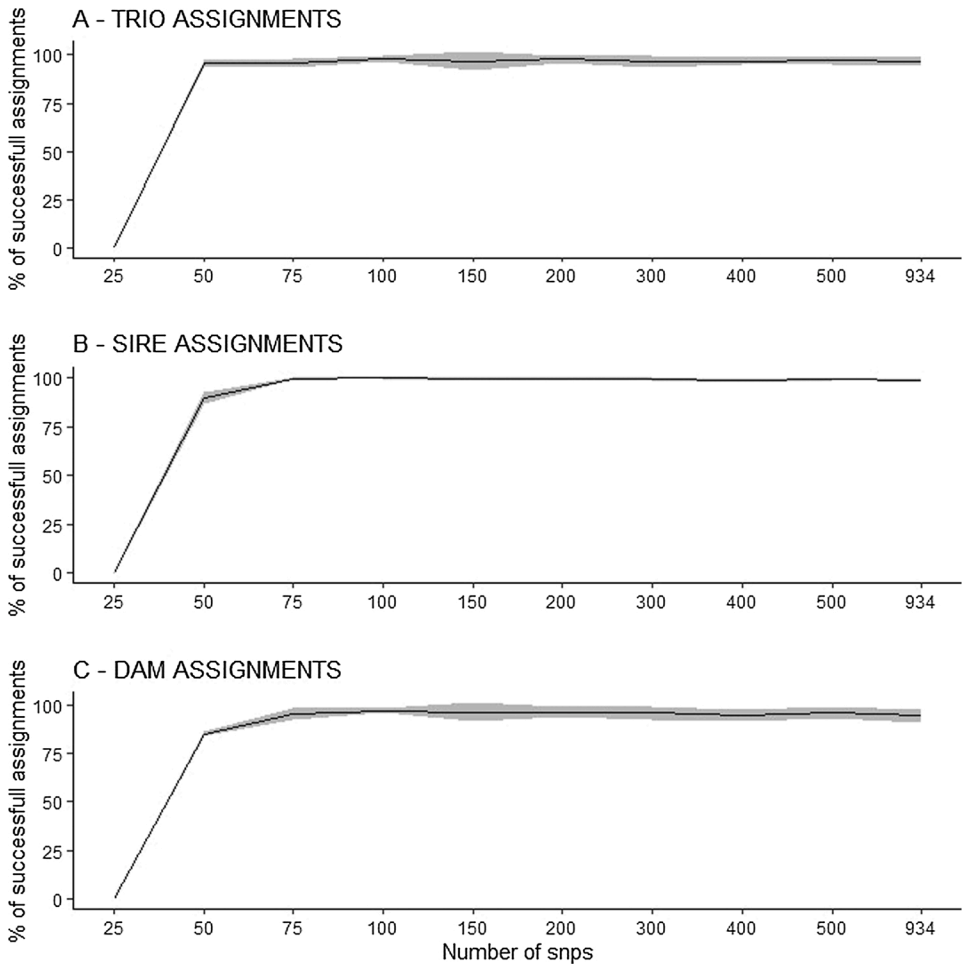

Fig. 2. Percentage of successful assignments (correct parent with 95 % confidence) of 1,000 simulated offspring for parent-progeny trio (A), sire (B) and dam (C), as a function of the number of SNPs used in the panel. Bolded continuous represent the mean assignment success of progeny to their respective correct parents over five replicated simulations for parent-progeny trio (A), sire (B) and dam (C). Grey shading represents the standard deviation around this mean.

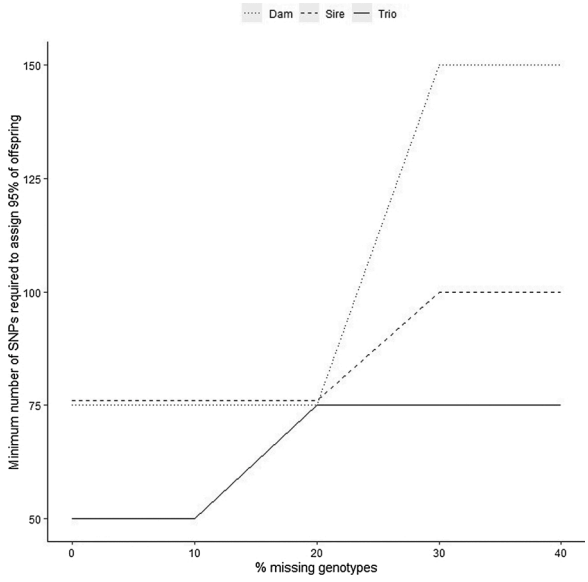

Fig. 3. Minimum number of SNPs required to assign 95 % or more individuals correctly (95 % confidence) for simulated offspring as a function of the percentage of randomly missing genotypes.

however, there was little improvement in assigning the remaining unassigned progeny when using larger SNP panels (i.e. only 83 % of correct progeny assignment was achieved with the panel of 648 SNPs).

3.4. Family distribution (with the 648 SNP-sized panel)

To investigate the distribution of families within the hatchery-spawned cohort of progeny, we used a reconstructed pedigree with only offspring that were assigned parents with 95 % confidence using the 648 SNP panel. The pedigree reconstructed from the parentage assignment from CERVUS, resulted in 768 offspring assigned (83 % of the offspring). The pedigree reconstruction detected 79 full-sib families with 35 families with more than one offspring (Fig. 5A), 11 maternal half-sib families with 8 families with more than one offspring (Fig. 5B) and 33 paternal half-sib families with 20 families with more than one offspring (Fig. 5C). The family size ranged from 1 to 134 offspring per full-sib family, from 1 to 300 offspring for maternal half-sib families and from 1 to 242 offspring for paternal half-sib families. As expected, parents contributed differentially to the progeny cohort resulting in a largely skewed full-sib and half-sib family distribution, with a few large families (two full-sib families with more than 100 offspring each accounts for 33 % of assigned offspring) and many smaller families. Four females out of 11 (80992DD, 80990DF, 8099208 and 8061821) contributed to 87 % of the offspring (39 %, 21 %, 16 % and 11 % respectively), while three males out of 33 (80617FC, 8061815 and 8061808) contributed to 81 % of the offspring (31 %, 30 % and 19 %,

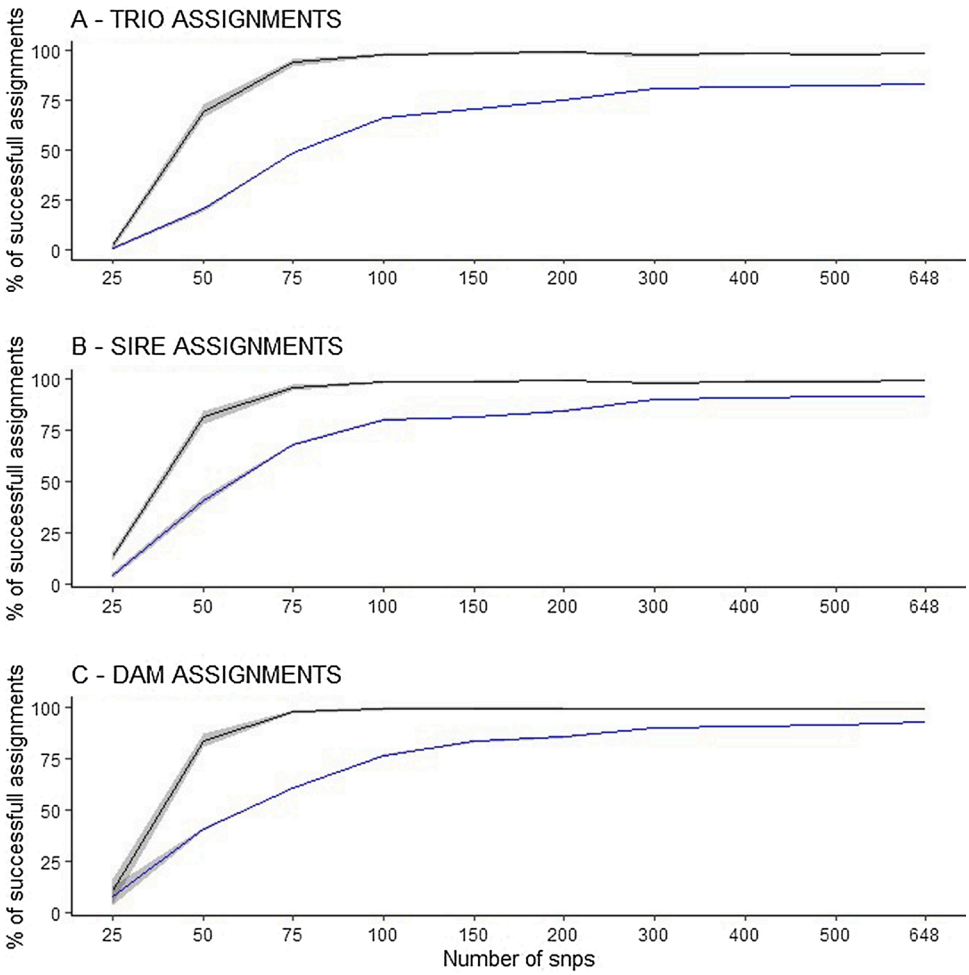

Fig. 4. Percentage of offspring assigned to parents with 95 % confidence using the software CERVUS 3.0 (Kalinowski et al., 2007) for both parents-progeny trio (A), sire (B) and dam (C), for the 920 hatchery-produced spat (blue line) and the 250 simulated offspring (black line). The grey shaded area around the lines corresponds to standard errors over 5 replicates.

respectively).

3.5. Utility of SNP to elucidate fine-scale family population structure

To further assess the utility of the SNPs to address questions related to fine-scale family genetic structure, a genetic distance network was produced and compared to the reconstructed pedigree. The nearest neighbour algorithm identified knn = 15 as optimal for the dataset. NetView analysis produced clusters of progeny that closely corresponded with the known pedigree and was able to resolve full-sib and maternal/paternal half-sib relationships among spat. Fig. 6 shows that most large full-sib families form independent clusters and validate the results found in parentage assignment. One cluster seems to be formed from small full-sib families (grey dots, top right cluster), but it is a distinct family with an identified dam but likely a non-sampled sire. This indicates that even in the absence of pedigree data, the SNP markers can aid in the delineation of family groupings.

4. Discussion

4.1. Comparison of parentage assignment with other molluscs

SNP-based panels have been recently developed for other molluscs and found to be powerful in resolving pedigree relationships. For instance, in the Pacific oyster (*Crassostrea gigas*) a panel of 40 SNPs was deemed adequate for parentage determination (Jin et al., 2014), while in Eastern oysters (*Crassostrea virginica*) and European abalone (*Haliotis tuberculate*) a panel of ~60 SNPs adequately resolved pedigree (Harney et al., 2018; Thongda et al., 2018). In the blue mussel (*Mytilus galloprovincialis*), a larger number of 179 SNPs were required to assign parents correctly in 92–100 % of assignment tests (Nguyen et al., 2014). The results from simulations conducted in the current study for *P. maxima* suggest that panels based on >50 SNPs are able to assign parent-progeny trios with 95 % confidence and that the full panel of 934 high quality SNPs will resolve parentage and/or elucidate fine-scale family relationships based on estimates of genetic relatedness, even in the case where there are a large number of possible candidate parents (i.e. > 150). However, based on the correct assignment of the *in silico* generated controls, the lower than expected assignment success of the hatchery bred progeny suggests that not all broodstock involved in the mass-spawn were actually tissue sampled and genotyped and thus the analysis was unable to assign parents-progeny trios from these pairings. This assumption that unassigned progeny result from a pairing where a broodstock was not genotyped is supported by the fact that the full-size panel of 648 SNPs was only able to assign progeny to a sire and dam in 91.8 % and 92.8 % of cases, respectively.

4.2. The use of SNPs versus microsatellites in parentage assignment success in Pinctada maxima

This study showed that 83 % of individuals had both parents assigned, 92 % of individuals had a sire assigned and 92 % of individuals

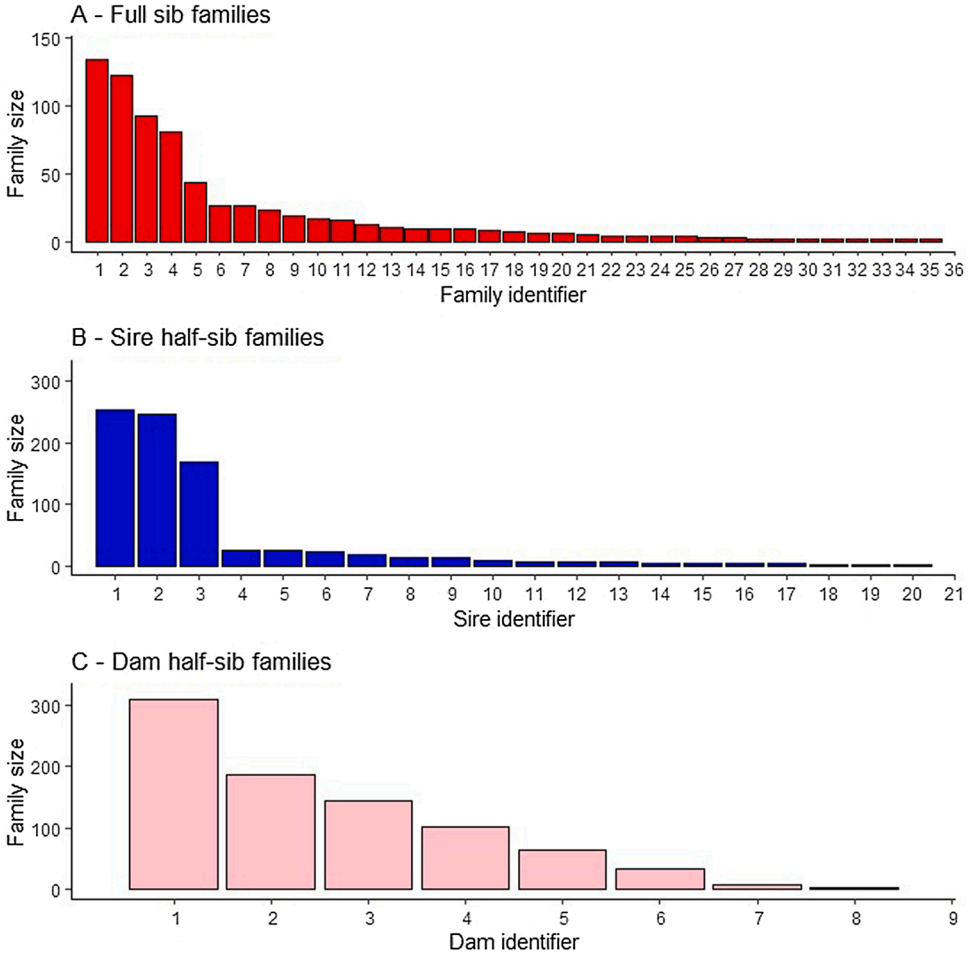

Fig. 5. Family distribution using the reconstructed pedigree of 768 offspring for full-sib (A), paternal half-sib (B) and maternal half-sib (C) families. Families identified with only one progeny have been omitted from the figure for clarity.

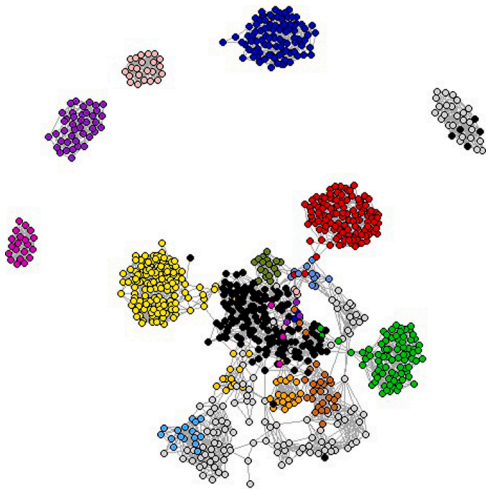

Fig. 6. NetView visualization of full-sib family distribution using SNPs with knn = 15 of Blue and pink dots represent the sires and dams, while other colours represent clusters of individuals indicative of full-sib families (The light grey dots represent spat from small families (less than 10 offspring) and the black dots represents the individuals with no parents assigned.

had a dam assigned using a panel of 648 SNPs. There are published microsatellite panels for the silver-lipped pearl oyster reported useful for DNA pedigree determination (Evans et al., 2005, 2006, Smith et al., 2003); however, while the microsatellite markers reported had good discrimination power when a low number of families were present in a cohort (i.e. < 10), their overall power to discriminate parentage accurately in the situation where there were potentially many candidate parents was only moderate. Evans et al. (2005) simulated the assignment power of a panel comprising eight microsatellites and showed that for a pearl oyster cohort comprising 50 full-sib families, only ~60 % of the offspring were able to be assigned with 95 % confidence using microsatellites to a parent-pair trio and ~80 % of offspring were assigned only to a single dam or sire. Therefore, the SNP panel and simulations reported in the present study shows a much higher level of pedigree resolving potential than the microsatellite panel of Evans et al. (2006), whereby 83 % of progeny from a potential 184 candidate parents had correct parent-progeny trio assignment, an improvement of ~20 % in overall assignment power.

4.3. Family distribution and implementation for breeding programs

DNA studies involving mass-spawning aquaculture species have highlighted the presence of skewed family distributions which can hinder the long-term capture and maintenance of genetic diversity in breeding programs (Blonk et al., 2010; Brown et al., 2005; Domingos et al., 2014; Foote et al., 2019; Frost et al., 2006) and specifically in oysters (i.e. flat oysters *O. edulis*, Lallias et al., 2010; *P. maxima*, Lind et al., 2009). Without an ability to resolve pedigree relationships through DNA parentage it is difficult to therefore identify and retain individuals from the lower represented families and/or to avoid consanguineous matings among progeny from the larger families that if

not avoided dramatically increase inbreeding levels. Previous research that has examined family distribution in hatchery-bred *P. maxima* are consistent with that seen for other mass-spawning species, where parentage assessment indicates that the progeny cohort was dominated by only 1–2 large full-sib families despite numerous broodstock potentially spawning (Lind et al., 2009). Our study supports the results of Lind et al. (2009) and even though the number of candidate broodstock potentially participating in the mass-spawn (n = 184) and full-sib families detected (n = 79) was the greatest yet reported in the literature for pearl oysters, family distribution was still observed to be highly skewed with three dams and sires contributing to ~70 % of the progeny. Massault et al. (2019) simulated a pearl oyster breeding program with a focus on selecting for JPOMS and suggested that ~200 half-sib families would be the optimum number of families to produce each generation to achieve a high genetic response and maximise control of inbreeding. In a mass spawning approach similar to the cohort we evaluated in this manuscript, it may be possible to generate this number of families within 2–3 mass spawning events; however, due to the differential contribution to the progeny by candidate parents, the results of our dataset suggest an increase in genotyping effort would be required to find sufficient individuals from the lower represented families and include them in the next round of spawning. Alternatively, hatchery spawning methodologies can be implemented resulting in more equal representation of families (i.e. single-pair spawning, strip-spawning, multiple mass-spawn involved only a few broodstock in each spawning tank, etc.).

Finally, although we did not simulate outcomes of parentage assignment under different incidences of family skewing, the power of assignment is not expected to decrease significantly, as CERVUS uses a likelihood approach to independently assign each progeny to a set of putative set of genotyped parents. As evidence of the power of the SNP panel to work in a real commercial situation, we were able to assign 83 % of progeny to parents with high confidence where there were 79 families present and two families dominated 33 % of the progeny cohort; this was even in the situation of where there were missing parents not genotyped that contributed to the progeny. Thus the markers perform adequately for parentage assignment under real-world commercial scenarios and will be applicable for use in selective breeding programs.

Author's statement

C. Massault: conceptualisation, SNP panel and parentage assignment (real and simulated), drafting the manuscript, editing

D.B Jones: conceptualisation, analysis linked to DartSeq and editing

K.R Zenger: conceptualisation and editing

J.M. Strugnell: conceptualisation and editing

R. Barnard: conceptualisation, supervising spawning on farm and data collection

D.R. Jerry: conceptualisation and editing

Funding

This research was funded by a Co-Operative Research Centre Program grant (CRC-P58609 "Breeding for resilience of JPOMS") and in collaboration with Ellies Pearls and Cygnet Bay Pearls.

Declaration of Competing Interest

The authors declare that there is no conflict of interest.

Acknowledgments

Special thanks to Julie Goldsbury for laboratory work at JCU involved in this study, Gary Firman and Liam Rawlins for collection of DNA of broodstock samples at Cygnet Bay and Tom Ackery for collection of DNA of broodstock at Ellies Pearling and laboratory work at JCU.

References

ABARES, 2018. Australian Bureau of Agricultural Resource, Economics and Science accessed 25/05/2020. https://www.agriculture.gov.au/abares.

Blonk, R.J.W., Komen, H., van, Kamstra A., Arendonk, J.A.M., 2010. Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations in common sole, *Solea Solea*. Genetics 184, 213–219.

Brown, C.R., Woolliams, J.A., McAndrew, B.J., 2005. Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. Aquaculture 247, 219–225.

Domingos, J.A., Smith-Keune, C., Jerry, D.R., 2014. Fate of genetic diversity within and between generations and implications for DNA parentage analysis in selective breeding of mass spawners: a case study of commercially farmed barramundi, *Lates calcarifer*. Aquaculture 424-425, 174–182.

DPIRD, 2016. Department of Primary Industry and Regional Development – Fisheries – Western Australia accessed 25/05/2020. http://www.fish.wa.gov.au/Pages/Home.aspx.

Evans, B.S., Knauer, J., Taylor, J.J., Jerry, D.R., 2005. A reliable DNA pedigree system for *Pinctada maxima* breeding studies. In: World Aquaculture Society Annual Conference. Bali, Indonesia.

Evans, B.S., Knauer, J., Taylor, J.J.U., Jerry, D.R., 2006. Development and characterization of six new microsatellites for the silver- or gold-lipped pearl oyster, *Pinctada maxima* (Pteriidae). Mol. Ecol. Notes 6, 835–837.

Foote, A.R., Simma, D.B., Khatkar, M.S., Raadsma, H., Guppy, J.L., Coman, G., Giardina, E., Jerry, D.R., Zenger, K.R., Wade, N., 2019. Considerations for maintaining family diversity in commercially mass spawned penaeid shrimp: a case study on *Penaeus monodon*. Front. Genet. 10 (1127).

Frost, L.A., Evans, B.S., Jerry, D.R., 2006. Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). Aquaculture 261, 1056–1064.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics 28 (23), 3150–3152.

Goncalves, P., Raftos, D., Jones, D., Anderson, K., Jones, B., Snow, M., 2017. Identifying the Cause of Oyster Oedema Disease (OOD) in Pearl Oysters (*Pinctada Maxima*), and Developing Diagnostic Tests for OOD, Sydney NSW Australia. ISBN 978-0-646-96779-0.

Guppy, J.L., Jones, D.B., Kjeldsen, S.R., Le Port, A., Khatkar, M.S., Wade, N.M., Sellars, M.J., Steinig, E.J., Raadsma, H.W., Jerry, D.R., Zenger, K.R., 2020. Development and validation of a RAD-Seq target-capture based genotyping assay for routine application in advanced black tiger shrimp (*Penaeus monodon*) breeding programs. BMC Genomics 21 (541).

Harney, E., Lachambre, S., Roussel, L., Huchette, S., Enez, F., Morvezen, R., Haffray, P., Boudry, P., 2018. Transcriptome based SNPs discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculate*. Aquaculture 491, 105–113.

Jerry, D.R., Kvingedal, R., Lind, C., Evans, B., Taylor, J.J.U., Safari, A., 2012. Donor-oyster derived heritability estimates and the effect of genotype X environment interaction on the production of pearl quality traits in the silver-lipped pearl oyster, *Pinctada maxima*. Aquaculture 338–341, 66–71.

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

Jones, D.B., Jerry, D.R., Khatkar, M.S., Raadsma, H.W., Zenger, K., 2013a. A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for a gene localisation and marker-assisted selection. BMC Genomics 14 (810).

Jones, D.B., Jerry, D.R., Foret, S., Konovalov, D., Zenger, K.R., 2013b. Genome-wide SNP validation and mantle tissue transcriptome analysis in the silver-lipped pearl oyster, *Pinctada maxima*. Mar. Biotechnol. 15, 647–658.

Jones, D.B., Jerry, D.R., Khatkar, M.S., Moser, G., Raadsma, H.W., Taylor, J.J.U., Zenger, K., 2014a. Determining genetic contributions to host oyster shell growth: quantitative trait loci and genetic association analysis for the silver-lipped pearl oyster, *Pinctada maxima*. Aquaculture 434, 367–375.

Jones, D.B., Jerry, D.R., Khatkar, M.S., Moser, G., Raadsma, H.W., Taylor, J.J.U., Zenger, K., 2014b. Quantitative trait loci and genetic association analysis reveals insights into complex pearl quality traits in donor silver-lipped pearl oyster. Aquaculture 434, 476–485.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099–1106.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., Unzynski, G., 2012. Diversity Arrays technology: a generic genome profiling technology on open platforms. Methods Mol. Biol. 888, 67–89.

Kvingedal, R., Evans, B.S., Taylor, J.J.U., Knauer, J., Jerry, D.R., 2008. Family by environment interactions in shell size of 43-day old silver-lip pearl oyster (Pinctada maxima), five families reared under different nursery conditions. Aquaculture 279, 23–28.

Kvingedal, R., Evans, B.S., Lind, C.E., Taylor, J.J.U., Dupont-Nivet, M., Jerry, D.R., 2010. Population and family growth response to different rearing location, heritability estimates and genotype×environment interaction in the silver-lip pearl oyster (Pinctada maxima). Aquaculture 304, 1–6.

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., Lapègue, S., 2010. Variance in the reproductive success of flat oyster *Ostrea edulis L.* Assessed by parentage analyses in natural and experimental conditions. Genet. Res. 92, 175–187.

Lind, C.E., Evans, B.S., Knauer, J., Taylor, J.J.U., Jerry, D.R., 2007. *Population genetics of a marine bivalve, Pinctada maxima, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits*. Mol. Ecol. 16 (24), 5193–5203.
Lind, C.E., Evans, B.S., Knauer, J., Taylor, J.J.U., Jerry, D.R., 2009. Decreased genetic diversity and a reduced effective population size in cultured silver-lipped pearl oysters (Pinctada maxima). Aquaculture 286, 12–19.
Massault, C., Zenger, K., Strugnell, J.M., Knauer, J., Firman, G., Barnard, R., Jerry, D.R., 2019. Selective breeding for improved survival to juvenile Pearl oyster Mortality Sydrome In Silver Lipped Pearl Oyster, *Pinctada Maxima* Proceedings of the Conference of the Association for the Advancement of Animal Breeding and Genetics. In: 23nd AAABG Conference 2019, 27/10–01/1/2019. Armidale, NSW, Australia.
McGinty, E.L., Evans, B.S., Taylor, J.U.U., Jerry, D.R., 2010. Xenografts and pearl production in two pearl oyster species, *P. Maxima* and *P. margaritifera*: effect on pearl quality and a key to understanding genetic contribution. Aquaculture 302, 175–181.
McGinty, E., Zenger, K.R., Taylor, J.U., Evans, B.S., Jerry, D.R., 2011. Diagnostic genetic markers unravel the interplay between host and donor oyster contribution in cultured pearl formation. Aquaculture 316, 20–24.
McGinty, E., Zenger, K.R., Taylor, J.U., Jerry, D.R., 2012. Transcriptome analysis of biomineralisation-related genes within the pearl sac: host and donor oyster contribution. Mar. Genom. 5, 27–33.
Neuditschko, M., Khatkar, M.S., Raadsma, W., 2012. NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. PLoS One 7 (10), e48375.
Nguyen, T., Hayes, B., Ingram, B., 2014. Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. Aquaculture 420–421, 295–301.
Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., Kilian, A., 2011. Diversity Arrays Technology (DArT) and next-generation sequencingcombined: genome-wide, high throughput, highly informative genotyping fomolecular breeding of Eucalyptus. BMC Proc. 26 (Suppl 7), 54.
Smith, C., Benzie, J.A.H., Wilson, K.J., 2003. Isolation and characterization of eight microsatellite loci from silver-lipped pearl oyster *Pinctada maxima*. Mol. Ecol. Notes 3, 125–127.
Southgate, P.S., Lucas, J., 2008. The Pearl Oyster – a Beginner's Guide. Elsevier B.V.
Steinig, E., Neuditschko, M., Khatkar, M.S., Raadsama, H.W., Zenger, K.R., 2016. NETVIEW P: a network visualization tool to unravel complex population structure using genome-wide SNPs. Mol. Ecol. Resour. 16, 216–227.
Thongda, W., Zhao, H., Jescovitch, L.N., Liu, M., Guo, X., Shrandt, M., Powers, S.P., Peatman, E., 2018. Development of SNP panels as a new tool to assess the genetic diversity, population structure, and parentage analysis of the Eastern Oyster (*Crassostrea virginica*). Mar. Biotechnol. 20, 385–395.
Weir, B.S., 1996. Genetic data analysis II. Sinauer, Sunderland, Mass.
Yue, G.H., Xia, J.H., 2014. Practical considerations of molecular parentage analysis in fish. J. World Aquac. Soc. 45, 89–103.