Aquaculture Reports 20 (2021) 100746

Contents lists available at ScienceDirect

Aquaculture Reports

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

Development and validation of a SNP-based genotyping tool for pedigree establishment in Australian greenlip abalone *Haliotis laevigata* Donovan, 1808

Phoebe M. Arbon [a,*], Catarina N.S. Silva [a], David B. Jones [a], Damian Jaccoud [c], Mark Gervis [d], Dean R. Jerry [a,b], Jan M. Strugnell [a,e]

[a] Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, 4811, Queensland, Australia
[b] Tropical Futures Institute, James Cook University, Singapore
[c] Diversity Arrays Technology Pty. Ltd., G.P.O. Box 3200, Canberra, ACT, 2601, Australia
[d] Southern Ocean Mariculture Pty Ltd, Port Fairy, 3284, Victoria, Australia
[e] Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, 3086, Victoria, Australia

ARTICLE INFO

Keywords:
Parentage assignment
Selective breeding
DArTag

ABSTRACT

To enable cumulative increases in aquaculture productivity, structured and efficient selective breeding programs are required. These are contingent on the management of genetic resources within the breeding population though the attainment of accurate pedigree information which can be provided by DNA-based parentage analysis. This study developed a SNP panel for greenlip abalone, *Haliotis laevigata,* from DArTSeq data produced from a genetic audit of greenlip abalone (n = 336) from five farms. The initial dataset consisted of 15,320 SNPs. Strict filtering on SNP quality control metrics was conducted to select the most informative 1,004 SNPs for the design of a DArTag™ genotyping panel for greenlip abalone. Sixteen broodstock and 1035 hatchery produced progeny were genotyped using the 1,004 SNP DArTag™ panel. The resulting genotypes were filtered, to produce 705 high performing SNPs. *In silico* parentage analysis using subsets of these SNPs (highest ranked by polymorphic information content) revealed that the inclusion of < 50 SNPs were satisfactory to resolve parentage for 10,000 simulated progeny to up to 200 candidate parents. The resolving power of the panel was also assessed under high levels of inbreeding (0–50 %) and relatedness (0–50 %) between candidate parents, and conditions of missing parental genotypes (10–50 %) *in silico*. Under all levels of inbreeding, relatedness and missing parental genotypes simulated, the panel was able to accurately assign parental pairs to the offspring. To validate these *in silico* results, parentage analysis of the hatchery produced progeny of the 16 candidate broodstock was conducted using both CERVUS and APIS. Complete parentage was assigned to all experimental progeny, with 100 % consensus between the two methods used. This study indicates this panel will serve as an efficient and cost-effective tool for accurate pedigree establishment for greenlip abalone.

1. Introduction

The sustainability of growth in the aquaculture sector is contingent on the development and implementation of selective breeding programs across a diverse range of production species (Gjedrem et al., 2012; Houston et al., 2020; Janssen et al., 2017; Zenger et al., 2019). Consequently, careful long-term management of genetic resources within breeding populations is required to avoid detrimental effects of inbreeding, while allowing selection and retention of superior families (Nayfa et al., 2020; Nguyen, 2016; Oliehoek and Bijma, 2009). Accurate pedigree information is essential to generating appropriate and beneficial selection decisions (Domingos et al., 2014; Nayfa et al., 2020). Selection of broodstock is often based on estimated breeding values (EBVs), derived from both phenotypic and genetic information on the individual. The precision of breeding values and trait heritability estimates improve with the accuracy of pedigree information (Oliehoek and Bijma, 2009; Premachandra et al., 2019). Thus, the capacity to obtain accurate pedigree information is essential to the general viability and

* Corresponding author at: Aquaculture Department, James Cook University, Building 32, 1 James Cook Dr, Douglas, QLD, 4811, Australia.
E-mail addresses: phoebe.arbon@jcu.edu.au (P.M. Arbon), catarina.silva@jcu.edu.au (C.N.S. Silva), david.jones051986@gmail.com (D.B. Jones), damian@diversityarrays.com (D. Jaccoud), mg@som-abalone.com.au (M. Gervis), dean.jerry@jcu.edu.au (D.R. Jerry), jan.strugnell@jcu.edu.au (J.M. Strugnell).

https://doi.org/10.1016/j.aqrep.2021.100746
Received 30 December 2020; Received in revised form 30 May 2021; Accepted 30 May 2021
Available online 16 June 2021

success of selective breeding programs.

Traditionally, to enable tracing of genetic relationships and the integration of pedigree information into selection decisions within breeding programs, physical tagging of stock has been implemented as a low technology solution (Prince, 1991). However, in many instances physical tagging for pedigree maintenance is highly impractical (i.e. for crustaceans), or impossible due to the reproductive vagaries and small early life-stages of many aquaculture species (Frost et al., 2006; Gjedrem, 1998; Yue and Xia, 2014). For many years, DNA microsatellites (regions of DNA with high repeat motif of base pairs) have been the marker of choice to retrospectively assign parentage for many communally reared aquaculture species (García-Fernández et al., 2018; Jones et al., 2011; Nguyen et al., 2014; Thongda et al., 2018). More recently, single nucleotide polymorphism (SNP) markers are becoming popularized for pedigree establishment in aquaculture species (Beacham et al., 2018; Dussault and Boulding, 2018; Kijas et al., 2019; Thongda et al., 2018). Larger SNP marker sets are required to achieve high rates of parental resolution, compared to microsatellites (Liu et al., 2017). However, this is compensated by the high frequency and availability of SNPs within the genome (Bester et al., 2008) and their high-throughput genotyping capacity (Guppy et al., 2020). Additionally, the larger number of SNP markers used in parentage analysis, compared to microsatellites, may offset the effects of reduced heterozygosity on assignment success, allowing for greater accuracy and efficiency of parentage analysis in cases with high rates of inbreeding (Premachandra et al., 2019).

Abalone is a high-value seafood commodity, particularly in Southeast Asia and China. Australia is a significant global producer of abalone, with the majority of exported product derived from wild-capture fisheries (ABARES, 2020). However, with wild-catch volumes expected to remain constrained by conservative catch quotas, aquaculture is expanding production to deliver the majority of projected industry growth (ABARES, 2019, 2020; Cook, 2019). The Australian abalone aquaculture industry is primarily focused on the production of high-value blacklip abalone (*Haliotis rubra* Leach, 1814), greenlip abalone (*Haliotis laevigata* Donovan, 1808) and a cross between these two species, namely, 'hybrid' abalone, (ABARES, 2020; Cook, 2019). Aquaculture production of abalone is challenged by slow growth rates, supply chain mortality and high production costs (Reaburn and Edwards, 2003); accordingly, there is much interest within the industry to develop and implement breeding programs to achieve sustainable productivity gain and improve farm return via genetic improvement (AAGA, 2020). Recently, several genetic marker panels have been developed for various abalone species illustrating their applications in determining natural population structure and for pedigree construction in breeding programs (Harney et al., 2018; Kijas et al., 2019; Miller et al., 2019; Sandoval-Castillo et al., 2018). Effective design and optimal performance of SNPs included on a parentage panel relies on validating the SNPs within the genetic diversity (i.e. farm broodstock) they are intended to be genotyped to reduce ascertainment biases. A cost-effective and efficient approach to achieve this is through genotype by sequencing (GBS) approaches such as DArTSeq and DArTag™. DArTSeq, a double digest, complexity reduction GBS approach, allows the discovery of SNPs throughout the genome using less sequencing. Once there are sufficient SNPs discovered, filtering and prioritisation can be conducted to design a smaller, more cost-effective DArTag™ panel (a targeted SNP genotyping platform), which is an efficient tool for genotyping large numbers of individuals (Guppy et al., 2018, 2020).

The present study aimed to develop and validate a low-density DArTag™ marker panel, consisting of recently isolated genome-wide SNP markers from Australian greenlip abalone, for high throughput and cost-effective genotyping to enable parental assessment. The parentage assignment power of the panel was assessed *in silico* with varying numbers of candidate parents, increasing rates of inbreeding, increasing levels of relatedness between candidate parents and missing parental genotypes. These experimental conditions were formulated to mimic challenging biological conditions within an aquaculture-based context. Additionally, the resolution power of the SNPs was validated under a real-world commercial hatchery situation through genotyping abalone progeny and evaluating the assignment level and confidence to greenlip abalone broodstock.

2. Materials and methods

2.1. SNP discovery and filtration

The genotypic data utilised within this study was produced from a genetic audit of commercial farm stocks of greenlip and blacklip abalone during 2016 and 2017 (Strugnell and Silva, 2017). A total of 336 greenlip abalone were included in the genetic audit. These individuals were collected from five farms located in three states in Australia (Victoria, n = 76, South Australia, n = 52 and Tasmania, n = 36, n = 60 and n = 112). The abalone selected from each farm for the audit were selected to best represent the genetic diversity present on each farm. DArTSeq libraries were created by DArT using the restriction enzymes PstI and SphI. The libraries were sequenced on a Hiseq 2500 (Illumina) at relatively high density at about 2–2.5 million sequence reads per sample to obtain single nucleotide polymorphisms (SNPs) across all individuals. To develop a DArTag™ panel containing a subset of the SNPs discovered within the DARTseq data, investigations into the genotypic data integrity were conducted before SNPs were iteratively filtered to ensure high priority SNPs were kept. First, samples with a call rate of < 80 % were excluded from the DARTseq dataset, followed by the silencing of any genotype call made with < 5 read counts. SNPs were then prioritised for the subarray if their reproducibility (based on 20 % technical replicate samples) was > 90 %, their missing data was < 30 %, they were biallelic, their minor allele frequency (MAF) was > 0.01, they were not in close proximity (based on genome sequence alignment), and polymorphism information content (PIC) was > 0.1. In addition, multi-copy markers were identified (and excluded) by using NCBI BLAST against the Blacklip abalone (*Haliotis rubra*) reference genome (Gan et al., 2019). Markers were further filtered for technical suitability by removing low complexity sequences and selecting those where the SNP position was between 23 and 45bp of a 70bp fragment. A DArTag™ panel was then developed by designing probes flanking the 1,004 prioritised SNPs with the highest MAF.

2.2. Candidate broodstock samples

Candidate greenlip abalone broodstock (N = 16), comprising seven dams and nine sires originating from wild Victorian (n = 14) and South Australian stocks (n = 2; sires 62 and 74), were sampled for genotyping. The experimental abalone were sourced from Southern Ocean Mariculture, a commercial abalone aquaculture facility, Port Fairy, Victoria, Australia, and were selected from a pool of candidate broodstock based on suitable maturation condition. The group selected constituted the broodstock of an actual mating scheme, as described in Section 2.3, representative of those used to generate a single cohort on farm. Epipodial tentacle clips of all candidate parents were sampled and preserved in ~98 % ethanol for subsequent genotyping (Section 2.4).

2.3. Hatchery produced greenlip abalone cohort

The experimental progeny were derived from single pair crosses of the 16 greenlip abalone broodstock (seven dams and nine sires), described in Section 2.2. Briefly, broodstock were spawned separately in 40 L glass aquaria to obtain unfertilised gametes. Gametes were mixed in 20 L buckets to produce 42 single pair crosses of the 63 crosses possible. These crosses were recorded and constituted the farm pedigree records. After fertilization, embryos were stocked into larval rearing tanks and raised using standard commercial husbandry protocols. Once the larvae formed cephalic tentacles, they were communally stocked into a

commercial rearing tank for settling. The larvae were settled onto *Ulvella lens* coated plates and then fed on diatoms and other algae growing on these plates.

A total of 1,035 greenlip abalone post-larvae (2.1 ± 0.5 mm) were sampled from the commercial rearing tank at 62 days of culture. The post-larvae were removed from settlement plates using a spatula and placed individually into tubes with ~1 mL of RNALater. The samples were kept at 2 °C overnight before being stored at −20 °C. Whole post larval individuals were homogenized using a BioSpec Mini-BeadBeater 96 with ~0.5 mL 1.0 mm dia. Zirconia/silica beads (Daintree Scientific Australia; BioSpec), 250 μL of 1x TE (Tris-EDTA) buffer and 5 μL of 10 % sarcosyl solution. To lyse, 100 μL of the sample homogenate was added to 200 μL of 1 x TE buffer and 10 μL of proteinase K (~20 ng/ ul) and incubated at 55 °C for 2 h. DNA extraction was conducted in a 96 well format using the chemagic™ DNA extraction kit and Zephyr® automated workstation (Perkin Elmer™), according to the manufacturer's instructions. Briefly, 30 μL of magnetic beads and 400 μL of binding buffer was added to each sample well and mixed thoroughly. The beads were then settled out for 5 min using a magnetic block and the supernatant was removed. Following this, four wash steps were performed using 450 μL of wash buffer. The final product was eluted into 20 μL of elution buffer for subsequent genotyping (Section 2.4).

2.4. Genotyping

Broodstock tentacle samples and extracted post larval DNA samples were sent to Diversity Arrays Technology, Australia, for DNA extraction and sequencing using the DArTag™ service, containing 1,004 SNPs for Australian greenlip abalone, as described above. Prior to sequencing using the SNP panel, quality control gels were run by DArT on the gDNA for a subset of each sample group. Additionally, DNA quality (260/280) and quantity (μl^{-1} ng) was assessed using a NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific ™) which had an average DNA concentration of 18.09 (± SD 8.15) ng/μl and an average 260/280 ratio of 1.78 (± SD 0.35). Co-analysed broodstock and post larval DArTag™ genotypic data, across the 1,004 SNPs, was filtered using DartQC 2.0 (https://github.com/esteinig/dartqc) to ensure the retention of the most genetically informative markers. SNPs with an average read depth > 5, call rate > 95 %, MAF > 0.02 and sequence similarity < 95 % were selected. Summary statistics, including combined exclusionary probability for the parental pairs (Ex_{PP}), polymorphic information content (PIC), observed heterozygosity (H_o) and expected heterozygosity (H_e) were generated directly using CERVUS 3.0 (Kalinowski et al., 2007). Data across a total of 705 SNPs (70.2 %) were retained for *in silico* SNP power analysis and pedigree reconstruction across the broodstock and progeny genotypes. Within the post-larval samples, 32 post-larvae returned genotypes at less than 50 % of the 705 loci. These samples were removed prior to subsequent analysis.

2.5. In silico SNP power analysis

To assess the theoretical parental assignment power of the SNPs under challenging and variable scenarios pertinent to aquaculture production, a suite of parental assignment simulations were conducted using CERVUS 3.0 (Kalinowski et al., 2007). *In silico* offspring and candidate broodstock genotypes were generated using the experimental post larval and broodstock genotypes across the 705 filtered SNPs, as described above. For each progeny, parental pairs were assigned according to a log-likelihood ratio (LOD score).

Initially, to assess the effect of the number of candidate parents on assignment success, three scenarios were investigated. The first scenario simulated the experimental numbers of broodstock, including a total of 16 candidate parents (seven female and nine male) with all candidate parents being sampled (Scenario 1; equivalent to the commercial hatchery spawn scenario described in Section 2.3). The second and third scenarios investigated the effect of larger numbers of candidate parents, with 100 candidate parents (Scenario 2) and 200 candidate parents (Scenario 3) simulated to comprehensively assess the extent of parentage assignment power of the SNPs. The number of SNPs required to achieve 100 % parental assignment success with 95 % confidence was determined for each scenario by simulating parental assignment using incremental subsets of SNP markers (from 10 to 50 SNPs in increments of 10 SNPs). These three subsets of SNPs were derived using a top-down selection of the filtered 705 SNPs, ranked by PIC (highest to lowest). When 100 % assignment success was reached at an incremental number of SNPs, one SNP was removed, and the simulation was run again to confirm the exact number of SNPs needed for each scenario. Summary statistics, including combined exclusionary probability for the parental pairs (Ex_{PP}), polymorphic information content (PIC), heterozygosity observed (H_o) and heterozygosity expected (H_e) were generated directly using CERVUS 3.0 (Kalinowski et al., 2007) for each subset. Secondly, the parentage assignment performance of the 705 filtered SNPs were assessed under differing levels of inbreeding between candidate broodstock. The rate of inbreeding among broodstock was increased for each simulation from 0 to 50 % in increments of 10 % for 100 simulated candidate parents. To assess the performance of the SNPs with the effect of increasing relatedness between candidate parents, simulations were run using 100 simulated candidate parents with the level of relatedness between parents increasing from 0 to 50 % (equivalent to full-siblings) in increments of 10 % relatedness. Finally, to assess the effect of missing parental genotypes on assignment success, simulations were run using 100 simulated candidate parents with the proportion of missing sires increasing from 0 to 50 %, in increments of 10 %. As conducted by Griot et al., 2020, only sire genotypes were removed as assignment error is more likely when only one true parental genotype is missing, compared to when both are missing (Jamieson and Taylor, 1997). All simulations were run in triplicate with 10,000 simulated offspring.

2.6. Parentage assignment of the hatchery-produced cohort

To validate the parental assignment power of the SNP markers, established by the *in-silico* analysis, the hatchery-produced cohort comprising broodstock and progeny was assessed. Parental pairs of the real hatchery-produced greenlip abalone progeny were assigned using both the maximum likelihood-based method in CERVUS 3.0 (Kalinowski et al., 2007) and the method of observed average Mendelian transmission probabilities, with a 5 % error rate, in APIS 1.0.1 (Griot et al., 2020), across the 705 filtered SNPs.

2.7. Cohort resolution

As well as assessing the power of the SNPs to ascertain parentage where parents were known (i.e. classical parentage assignment methods), we also assessed the discriminatory power of markers to cluster progeny into family groups in the hatchery-produced progeny based purely on allelic compositions and genetic relationships using two commonly applied clustering techniques. The first approach involved using the R package *adegenet* to discriminate relationships among progeny through discriminant analyses of principal components (DAPC; Jombart, 2008). Here, DAPC was used on individual genotypes in a multivariate analysis to determine the best number of genetic clusters (K) to retain by running the function *find.clusters*(). DAPC was conducted based on inclusion of dams and sires separately (i.e. individual broodstock identifiers) to allow the visualisation of independent family groups. Secondly, and as a comparison to analyse family structure, a network analysis was performed using the NetView P pipeline v.0.6 (Steinig et al., 2015).

Although the panel was designed for parentage application in greenlip abalone, the panel's inter-species discriminatory resolution was also assessed. The genotypes across the final 1,004 SNPs identified for DArTag™ were extracted from the genetic audit DArTSeq data for greenlip, blacklip and hybrid abalone. DAPC was used on these

Table 1
Genetic parameters for the SNP subsets used in the parental assignment simulations. Parameters include; the number of SNPs in each subset panel (i.e. the number of SNPs adequate to achieve complete parental assignments at 95 % confidence; N), polymorphic information content (PIC), observed heterozygosity (H_o), expected heterozygosity (H_e), and combined exclusionary probability for the parental pairs (Ex_{PP}).

Simulation factor	Simulation scenario	*N*	*PIC*	H_o	H_e	Ex_{PP}
Number of candidate parents	Scenario 1 - (experimental) 16 candidate parents	26	0.375	0.52	0.50	0.999
	Scenario 2–100 candidate parents	37	0.375	0.51	0.50	1
	Scenario 3–200 candidate parents	41	0.375	0.51	0.50	1
Inbreeding	Inbreeding from 0 to 50 % between 100 candidate parents	705	0.271	0.31	0.34	1
Relatedness	Relatedness from 0 to 50 % between 100 candidate parents	705	0.271	0.31	0.34	1
Missing parental genotypes	Candidate sire genotypes missing 10–50% from 100 candidate parents	705	0.271	0.31	0.34	1

genotypes to visualise the resolution of the species groups.

3. Results

3.1. SNP discovery and filtration

After processing the raw sequencing DArTSeq data across the 336 genetic audit abalone, the datasets consisted of 15,320 SNPs. After initial filtering for the best quality SNPs, 3,527 greenlip SNPs were identified (Strugnell and Silva, 2017). Secondary filtration of these SNPs identified 1,004 highly informative, genome wide SNPs which formed the basis for the DArTag™ panel. All experimental samples were genotyped using the 1,004 DArTag™ (Diversity Arrays Technology) probes.

3.2. In silico SNP power analysis

Filtering of the broodstock and post larval genotypes retained 705 (70.2 %) of the 1,004 SNPs, with an average PIC of 0.271 and a combined exclusionary probability for parental pairs of 1.00, indicating the capacity for accurate parentage assignment (Table 1). *In silico* assessment of the constituent SNPs, under three levels of parentage complexity involving increasing number of candidate parents (Scenario 1 (experimental) – 16 candidate parents; Scenario 2–100 candidate parents; Scenario 3–200 candidate parents), demonstrated that < 50 of the SNPs with highest PIC were required to fully resolve the parentage of 10,000 simulated offspring for all three scenarios. Despite the vast differences in the number of candidate parents simulated across the three parental complexity scenarios, the number of SNPs required across subset panels did not increase dramatically (Table 1), suggesting strong discriminatory power using the SNPs with highest ranked PIC. Specifically, subsets of 26, 37 and 41 SNPs (Table 1) were required to assign progeny correctly for Scenario 1, Scenario 2 and Scenario 3, respectively (Fig. 1). For these final three subsets of SNPs, required to achieve complete parentage assignment across the three simulated parental complexity scenarios, Ex_{PP} was >99 % (Table 1).

Further simulations demonstrated that regardless of the level of inbreeding that was simulated (0–50 %) among 100 candidate broodstock, the SNPs retained high power to resolve parentage, achieving 100 % assignment under all inbreeding scenarios. Similarly, at all levels of relatedness between candidate parents simulated (0–50 %), complete parental assignment success was achieved, at 95 % confidence. Finally, for the simulations addressing missing parental genotypes, while all offspring were assigned a parental dam, only offspring belonging to a sampled sire were assigned to a parental pair, indicating that no offspring were incorrectly assigned to a parental pair when the true sire was missing.

3.3. Hatchery-produced greenlip abalone cohort and species-level resolution

In validation of the *in silico* assessments of the SNPs, all experimental post-larvae were assigned to the most likely parental pair (100 % assignment, 95 % confidence), with complete consensus between both the CERVUS (Kalinowski et al., 2007) and APIS (Griot et al., 2020) generated pedigrees. Parentage analysis revealed the presence of 47 families in the cohort of progeny examined. An additional six families which were not identified by the farm pedigree were found to be present using the genetic pedigree. Additionally, one family which was recorded as being in the farm pedigree (Sire 66 x Dam 60) was not detected in the post-larval samples (Table 2).

In general, DAPC showed a clear clustering by broodstock identifier (i.e. family), supporting the findings of the simulations. For the post-

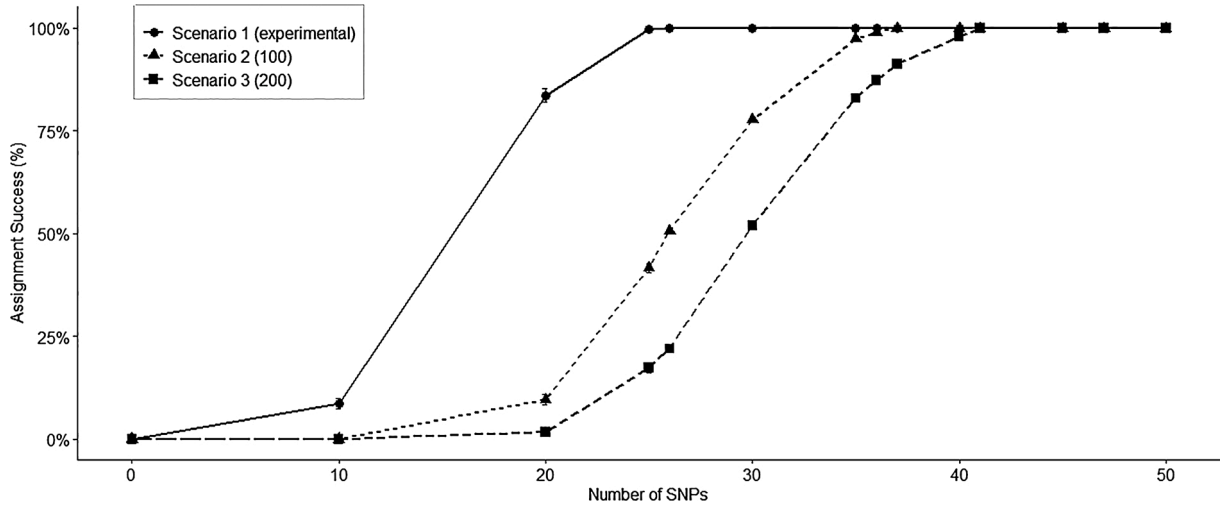

Fig. 1. The effect of increasing the number of candidate parents (Scenario 1 - 16 parents; Scenario 2 - 100 parents; Scenario 3 - 200 parents) on the number of SNPs required to fully resolve parentage of 10,000 simulated greenlip abalone offspring, with 95 % confidence, using CERVUS 3.0.

Table 2
Broodstock crosses from farm records (black crosses) and additional broodstock crosses identified through genetic parentage analysis of post-larvae (red crosses). Farm record crosses that were not detected using genetic parentage analysis in the post-larvae samples are shown by an empty red circle. The total number of crosses for each sire and dam are shown in black for the farm record values, and red for the genetic parentage analysis values. Here, numbers (e.g. 62) represent individual broodstock identifiers.

		Sire ID									
		62	63	64	65	66	69	71	72	74	Total
Dam ID	51				X	X	X			X	4 (4)
	53	X	X	X	X	X	X	X	X		6 (8)
	55		X	X	X	X	X	X		X	6 (7)
	56	X	X	X	X	X	X	X		X	7 (8)
	57		X	X	X	X	X			X	5 (6)
	60	X	X	X	X	XO	X	X		X	7 (7)
	61	X	X	X	X		X	X	X		7 (7)
	Total	3 (4)	6 (6)	6 (6)	6 (7)	6 (5)	5 (7)	4 (5)	1 (2)	5 (5)	42 (47)

larvae dataset all experimental broodstock clustered separately, except Sire 69, Sire 71 and Sire 72 (Fig. 2a) and Dam 51 and Dam 55 (Fig. 2c). The family groups for each sire or dam that did not distinctly cluster were of the lowest relative contributors to the cohort (data not shown). Sire 62 and Sire 74 (originating from South Australia, Australia) clustered distinctly from the other abalone sires (originating from Victoria, Australia) (Fig. 2a). Netview results supported the DAPC results and showed finer scale clustering with subgroups of individuals from the same broodstock clustering separately (Fig. 2b and Fig. 2d). This clustering within each broodstock group reflects the experimental design with broodstock contributing to different families (i.e. half-sib relationships).

The species-level resolution for the 1,004 SNP panel was low, with the DAPC revealing no distinct clustering by abalone species (Supplementary Fig. 1). This was expected given the design of the panel was tailored on polymorphic SNPs with good design parameters in greenlip abalone.

4. Discussion

The present study developed and validated a SNP-based genotyping tool to provide accurate and robust genetic data for the establishment of pedigrees in Australian greenlip abalone aquaculture. The parental assignment power of our SNP panel was assessed *in silico,* using

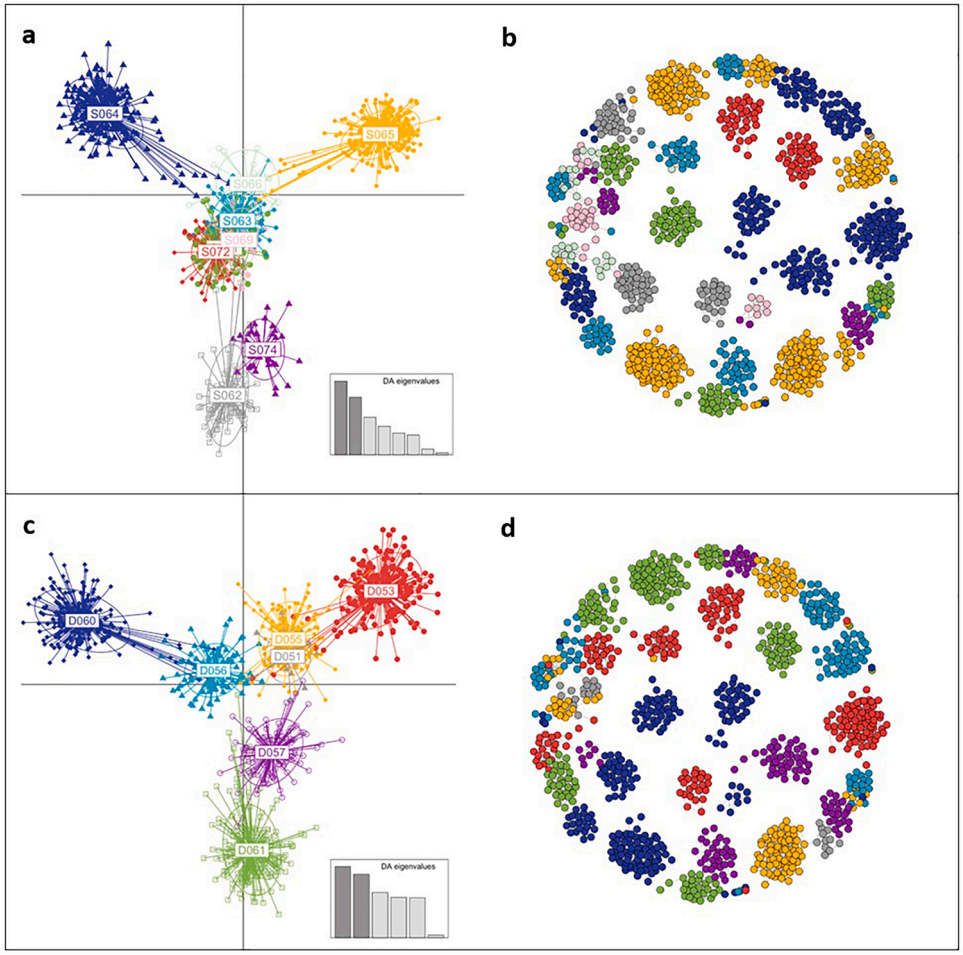

Fig. 2. Cohort genetic resolution for greenlip abalone post-larvae samples. Discriminant analyses of principal components (DAPC; Jombart, 2008) plot for individual post larval genotypes assigned to parental (a) sires and (c) dams of independent family groups. Supporting network analysis plot (NetView P pipeline v.0.6) showing family structure of the post-larvae samples assigned to parental (b) sires and (d) dams of independent family groups.

parentage assignment simulations in CERVUS 3.0., and validated for a commercially produced greenlip abalone cohort. To capture sufficient genetic information to render the panel broadly applicable across the Australian greenlip abalone aquaculture industry, the initial sequencing effort for variant discovery included a total of 336 greenlip commercial farmed abalone from five Australian farms across three states.

The parentage assignment power of the SNPs was assessed *in silico,* under complex and commercially relevant conditions. Increasing parentage complexity through increasing the number of candidate parents on the SNP panel's assignment success was simulated under three candidate parent scenarios. The simulations of the first scenario, representing the experimental scenario of 16 candidate parents (63 potential parental crosses), demonstrated that only the top 26 SNPs, ranked by descending PIC, were required to fully resolve parentage with 95 % confidence. The second and third simulated scenarios revealed that with a marginal increase in the number of SNPs, parentage was able to be completely resolved under highly complex parentage scenarios exceeding those indicative of multiple mixed cohorts on farm (Gervis, Southern Ocean Mariculture, pers comm, 2019; Hamilton et al., 2009). Similar studies have found the performance of SNPs to be comparably robust and accurate, requiring on average 55 SNPs to resolve >95 % parentage assignment of aquaculture species, including Pacific (*Crassostrea gigas*), Eastern (*Crassostrea virginica*) and European flat (*Ostrea edulis*) oysters and European abalone (*Haliotis tuberculata*) (Harney et al., 2018; Jin et al., 2014; Lapègue et al., 2014; Liu et al., 2017; Thongda et al., 2018). However, when using the likelihood-based method for parentage assignment, such as in CERVUS, type 1 error rates (false positive assignments) can be inflated when a limited number of markers are used (< 100, equivalent to exclusionary power < 0.9999) (Griot et al., 2020). Thus, while the results from the first set of simulations in our study serve in demonstrating the power of the SNPs, where only 26, 37 and 41 SNPs were required to assign parentage for the three parentage scenarios, consideration of these results, and the results of other studies applying a similar approach, should recognise this caveat. A recent study conducted by Kijas et al. (2019), focusing on Australian greenlip, blacklip and hybrid 'Tiger' abalone, developed and validated a SNP-based genotyping tool which was subsequently applied for parentage assignment. They reported 100 % assignment success to the 86 candidate broodstock with an average of 100 SNPs using CERVUS. The present study reports the requirement of a lower number of SNPs under similarly complex scenarios *in silico*. The panel of Kijas et al. (2019) was derived from whole genome sequencing of 24 (eight blacklip, eight greenlip, and eight hybrid) commercially produced abalone. Comparatively, the panel in the present study was developed from a larger sample group (n = 336), comprising greenlip abalone sampled from five separate abalone farms and three states. As suggested by Kijas et al. (2019), sampling of animals sourced from a wider genetic background would likely identify a higher number of SNP variants. Here, the larger sample size used to initially develop the panel in the present study may have captured more informative and discriminative markers with a higher average overall minor allele frequency in the design of the panel, allowing for increased parentage assignment resolution from a smaller number of SNPs.

The second and third simulation sets of the current study investigated the effects of inbreeding and relatedness between candidate parents on the assignment resolution of the panel. Candidate parents (n = 100) were simulated with relatedness levels up to an equivalent of full-sibling relationships. Here, the SNP panel was able to resolve 100 % parentage with 95 % confidence. Similarly, full parentage resolution was achieved when rates of inbreeding exceeding those which would be realistically encountered on farm, without considerable and obvious detriment to production. The performance of the panel under these conditions is especially relevant to aquaculture selective breeding programs, where relatedness between candidate parents and inbreeding are pervasive (Gjerde et al., 1996; Nguyen, 2016; You and Hedgecock, 2018). As such, these results further substantiate the capacity of the SNP panel to perform with high accuracy and confidence under relevant and challenging commercial conditions.

Application of the panel as a SNP-based genotyping tool to a commercially produced greenlip abalone cohort established its capacity for fine-scale, high resolution pedigree construction, and validated the *in silico* findings. Accurate assignment of parental pairs was achieved for all sampled progeny, demonstrating the value of the panel for retrospective pedigree construction of the communally reared species. Further, through DAPC analysis of individual progeny genotypes, the resolution of the panel was shown to distinctly cluster individual families and separate those belonging to South Australian or Victorian greenlip abalone sires, capturing the genetic distinction between these populations (Mayfield et al., 2014). The panel did not yield high resolution to discriminate between species, likely due to the fact that it was designed using greenlip-specific SNPs during the initial selection of the DArTag™ markers.

Six families which were not identified in the farm-kept pedigree records (unexpected) were identified using our SNP-based genotyping tool. Assignment of offspring to a parental pair when the true sire was missing was not observed in the parentage assignment simulations. Further, complete parentage assignment was achieved *in silico* under challenging conditions, including when candidate parents were highly related, supporting the validity of the parental assignments made using the SNP panel to the unexpected families in the hatchery cohort, contrary to the farm-kept records. Five of the six unexpected families each had a relative contribution less than 0.8 % (data not shown). Potentially, these families were a product of unintended or contaminant gamete mixing during the production of single pair crosses from the candidate parents, or from unaccounted fertilization of residual viable gametes when embryos were stocked communally into larval rearing tanks. The sixth unexpected family represented a considerably larger relative contribution to the cohort. Potentially, the identification of this larger family was a result of a hatchery record-error, as an expected family from the farm-kept pedigree was not identified from the cohort by the SNP-based genotyping and may have been mistakenly recorded instead of the true cross. Missing data and errors within farm kept pedigree records are common within many aquaculture selective breeding programs (Nayfa et al., 2020; Nguyen, 2016; Oliehoek and Bijma, 2009; Visscher et al., 2002; Weller, 2006). Pedigree errors can have a significant impact on the accuracy and capacity for breeding value estimation, response to selection and can result in incorrect estimations of family diversity on farm. Studies investigating the effects of pedigree errors circa 10 % have reported losses in genetic gain of 2–4 % (Israel and Weller, 2000; Visscher et al., 2002). In such cases, the value of DNA parentage analysis to accurately construct pedigrees, such as the analysis presented in this study, is highlighted.

4.1. Conclusion

This study validated a SNP panel for greenlip abalone parentage analysis though both *in silico* and *in situ* application. The panel's constituent SNPs achieved complete parentage resolution for a commercially produced Australian greenlip abalone cohort and performed robustly under challenging and complex simulated scenarios, with large numbers of candidate parents, high rates of inbreeding, high levels of relatedness between candidate parents and cases of missing parental genotypes. Considering the findings of the present study, we suggest that the SNP-based genotyping tool established in this study will return accurate pedigree construction for Australian greenlip abalone, supporting the pursuit of genetic improvement for this species.

Data availability statement

The supporting data of this study are available upon reasonable request to the corresponding author.

CRediT authorship contribution statement

Phoebe M. Arbon: Methodology, Data curation, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Catarina N.S. Silva:** Methodology, Investigation, Formal analysis, Visualization, Writing - review & editing. **David B. Jones:** Formal analysis, Methodology, Writing - review & editing. **Damian Jaccoud:** Methodology, Data curation, Investigation, Writing - review & editing. **Mark Gervis:** Resources, Writing - review & editing. **Dean R. Jerry:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Jan M. Strugnell:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Methodology, Investigation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Southern Ocean Mariculture for providing and rearing the experimental animals. This work was supported by funding from the FRDC on behalf of the Australian Government. This funding includes projects '2016-142 Genetic diversity audit of farm held stocks of Greenlip and Blacklip abalone' and also '2017-220 Testing established methods of early prediction of genetic merit in abalone broodstock'.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aqrep.2021.100746.

References

Aaga, 2020. AAGA Strategic Plan 2020-25. Australian Abalone Growers Association, Beaconsfield, Tasmania.

ABARES, 2020. Australian Fisheries and Aquaculture Outlook 2020.

ABARES, 2019. Annual Fisheries Outlook - Department of Agriculture [WWW Document]. URL http://www.agriculture.gov.au/abares/research-topics/fisheries/fisheries-economics/fisheries-forecasts#abalone (Accessed 10.19.19).

Beacham, T.D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J.R., Willis, D., Lynch, C., Moore, J.-S., Bernatchez, L., Withler, R.E., 2018. Comparison of coded-wire tagging with parentage-based tagging and genetic stock identification in a large-scale coho salmon fisheries application in British Columbia, Canada. Evol. Appl. 12, 230–254. https://doi.org/10.1111/eva.12711.

Bester, A.E., Roodt-Wilding, R., Whitaker, H.A., 2008. Discovery and evaluation of single nucleotide polymorphisms (SNPs) for *Haliotis midae:* a targeted EST approach. Anim. Genet. 39, 321–324. https://doi.org/10.1111/j.1365-2052.2008.01728.x.

Cook, P.A., 2019. Worldwide abalone production statistics. J. Shellfish Res. 38, 401. https://doi.org/10.2983/035.038.0222.

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. https://doi.org/10.1016/j.aquaculture.2014.01.004.

Dussault, F.M., Boulding, E.G., 2018. Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: a methodology illustrated using Atlantic salmon. Aquac. Res. 49, 1368–1372. https://doi.org/10.1111/are.13566.

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.

Gan, H.M., Tan, M.H., Austin, C.M., Sherman, C.D.H., Wong, Y.T., Strugnell, J., Gervis, M., McPherson, L., Miller, A.D., 2019. Best foot forward: nanopore long reads, hybrid meta-assembly, and haplotig purging optimizes the first genome assembly for the southern hemisphere blacklip abalone (*Haliotis rubra*). Front. Genet. 10, 889. https://doi.org/10.3389/fgene.2019.00889.

García-Fernández, C., Sánchez, J.A., Blanco, G., 2018. SNP-haplotypes: an accurate approach for parentage and relatedness inference in gilthead sea bream (*Sparus aurata*). Aquaculture 495, 582–591. https://doi.org/10.1016/J.AQUACULTURE.2018.06.019.

Gjedrem, T., 1998. Developments in fish breeding and genetics. Acta Agric. Scand. Sect. A - Anim. Sci.

Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture. https://doi.org/10.1016/j.aquaculture.2012.04.008.

Gjerde, B., Gjoen, M., Villanueva, B., 1996. Optimum designs for fish breeding programmes with constrained inbreeding Mass selection for a normally distributed trait. Livest. Prod. Sci. 47, 59–72.

Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., Vandeputte, M., 2020. APIS: an auto-adaptive parentage inference software that tolerates missing parents. Mol. Ecol. Resour. 20, 579–590. https://doi.org/10.1111/1755-0998.13103.

Guppy, J.L., Jones, D.B., Jerry, D.R., Wade, N.M., Raadsma, H.W., Huerlimann, R., Zenger, K.R., 2018. The state of "omics" research for farmed penaeids: advances in research and impediments to industry utilization. Front. Genet. https://doi.org/10.3389/fgene.2018.00282.

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. https://doi.org/10.1186/s12864-020-06960-w.

Hamilton, M.G., Kube, P.D., Elliott, N.G., Mcpherson, L.J., Krsinich, A., 2009. Development of a breeding strategy for hybrid abalone. Association for the Advancement of Animal Breeding and Genetics. Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW, pp. 350–353.

Harney, E., Lachambre, S., Roussel, S., Huchette, S., Enez, F., Morvezen, R., Haffray, P., Boudry, P., 2018. Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*. Aquaculture 491, 105–113. https://doi.org/10.1016/j.aquaculture.2018.03.006.

Houston, R.D., Bean, T.P., Macqueen, D.J., Gundappa, M.K., Jin, Y.H., Jenkins, T.L., Selly, S.L.C., Martin, S.A.M., Stevens, J.R., Santos, E.M., Davie, A., Robledo, D., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat. Rev. Genet. https://doi.org/10.1038/s41576-020-0227-y.

Israel, C., Weller, J.I., 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J. Dairy Sci. 83, 181–187. https://doi.org/10.3168/jds.S0022-0302(00)74869-7.

Jamieson, A., Taylor, S.C.S., 1997. Comparisons of three probability formulae for parentage exclusion. Anim. Genet. 28 (6), 397–400. https://doi.org/10.1111/j.1365-2052.1997.00186.x.

Janssen, K., Chavanne, H., Berentsen, P., Komen, H., 2017. Impact of selective breeding on European aquaculture. Aquaculture 472, 8–16. https://doi.org/10.1016/j.aquaculture.2016.03.012.

Jin, Y.-L., Kong, L.-F., 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. https://doi.org/10.1007/s13258-013-0150-0.

Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jones, D.D.B., Zenger, K.R., Jerry, D.R., 2011. In silico whole-genome EST analysis reveals 2322 novel microsatellites for the silver-lipped pearl oyster, *Pinctada maxima*. Mar. Genomics 4, 287–290.

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

Kijas, J., Hamilton, M., Botwright, N., King, H., McPherson, L., Krsinich, A., McWilliam, S., 2019. Genome sequencing of blacklip and greenlip abalone for development and validation of a SNP based genotyping tool. Front. Genet. 9, 687. https://doi.org/10.3389/fgene.2018.00687.

Lapègue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., Ballenghien, M., Genestout, L., Barbotte, L., Mahla, R., Haffray, P., Klopp, C., 2014. Development of SNP-genotyping arrays in two shellfish species. Mol. Ecol. Resour. 14, 820–830. https://doi.org/10.1111/1755-0998.12230.

Liu, T., Li, Q., Kong, L., Yu, H., 2017. Comparison of microsatellites and SNPs for pedigree analysis in the Pacific oyster *Crassostrea gigas*. Aquac. Int. 25, 1507–1519. https://doi.org/10.1007/s10499-017-0127-0.

Mayfield, S., Miller, K.J., Mundy, C.N., 2014. Towards Understanding Greenlip Abalone Population Structure: Final Report for the Fisheries Research and Development Corporation. FRDC Project no. 2010-013. http://www.frdc.com.au/project?id=360.

Miller, A.D., Hoffmann, A.A., Tan, M.H., Young, M., Ahrens, C., Cocomazzo, M., Rattray, A., Ierodiaconou, D.A., Treml, E., Sherman, C.D.H., 2019. Local and regional scale habitat heterogeneity contribute to genetic adaptation in a commercially important marine mollusc (*Haliotis rubra*) from southeastern Australia. Mol. Ecol. 28, 3053–3072. https://doi.org/10.1111/mec.15128.

Nayfa, M.G., Jones, D.B., Lind, C.E., Benzie, J.A.H., Jerry, D.R., Zenger, K.R., 2020. Pipette and paper: combining molecular and genealogical methods to assess a Nile tilapia (*Oreochromis niloticus*) breeding program. Aquaculture 523, 735171. https://doi.org/10.1016/j.aquaculture.2020.735171.

Nguyen, H., 2016. Genetic improvement for important farmed aquaculture species with a reference to carp, tilapia and prawns in Asia: achievements, lessons and challenges. Fish Fish. 17, 483–506. https://doi.org/10.1111/faf.12122.

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. https://doi.org/10.1016/j.aquaculture.2013.11.021.

Oliehoek, P.A., Bijma, P., 2009. Effects of pedigree errors on the efficiency of conservation decisions. Genet. Sel. Evol. 41, 9. https://doi.org/10.1186/1297-9686-41-9.

Premachandra, H.K.A., Nguyen, N.H., Knibb, W., 2019. Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. Aquaculture 499, 24–31. https://doi.org/10.1016/J.AQUACULTURE.2018.09.022.

Prince, J.D., 1991. A new technique for tagging abalone. Mar. Freshw. Res. 42, 101–106. https://doi.org/10.1071/MF9910101.

Reaburn, R., Edwards, S., 2003. Modelling the growth of cultured greenlip abalone (*Haliotis laevigata* Leach): practical modes of measurement. Aquac. Res. 34, 253–259. https://doi.org/10.1046/j.1365-2109.2003.00814.x.

Sandoval-Castillo, J., Robinson, N.A., Hart, A.M., Strain, L.W.S., Beheregaray, L.B., 2018. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. Mol. Ecol. 27, 1603–1620. https://doi.org/10.1111/mec.14526.

Steinig, E.J., Neuditschko, M., Khatkar, M.S., Raadsma, H.W., Zenger, K.R., 2015. NetView P: a network visualization tool to unravel complex population structure using genome-wide SNPs. Mol. Ecol. Resour. 16, 216–227. https://doi.org/10.1111/1755-0998.12442.

Strugnell, J.M., Silva, C., 2017. Genetic Diversity Audit of Farm Held Stocks of Greenlip and Blacklip Abalone. Fisheries Research and Development Corporation FRDC Project No.2016/142, Townsville. October.

Thongda, W., Zhao, H., Zhang, D., Jescovitch, L.N., Liu, M., Guo, X., Schrandt, 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. https://doi.org/10.1007/s10126-018-9803-y.

Visscher, P.M., Woolliams, J.A., Smith, D., Williams, J.L., 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J. Dairy Sci. 85, 2368–2375. https://doi.org/10.3168/jds.S0022-0302(02)74317-8.

Weller, J.I., 2006. Lessons for aquaculture breeding from livestock breeding. Isr. J. Aquac. 58 (4), 257–265.

You, W., Hedgecock, D., 2018. Boom-and-bust production cycles in animal seafood aquaculture. Rev. Aquac. https://doi.org/10.1111/raq.12278.

Yue, G.H., Xia, J.H., 2014. Practical considerations of molecular parentage analysis in fish. J. World Aquac. Soc. 45, 89–103. https://doi.org/10.1111/jwas.12107.

Zenger, K.R., Khatkar, M.S., Jones, D.B., Khalilisamani, N., Jerry, D.R., Raadsma, H.W., 2019. Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. Front. Genet. 9, 693. https://doi.org/10.3389/fgene.2018.00693.