



**HAL**  
open science

# Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*

Ewan Harney, Sebastien Lachambre, Sabine Roussel, Sylvain Huchette, Florian Enez, Romain Morvezen, Pierrick Haffray, Pierre Boudry

## ► To cite this version:

Ewan Harney, Sebastien Lachambre, Sabine Roussel, Sylvain Huchette, Florian Enez, et al.. Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*. *Aquaculture*, 2018, 491, pp.105-113. 10.1016/j.aquaculture.2018.03.006 . hal-02626095

**HAL Id: hal-02626095**

**<https://hal.science/hal-02626095>**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Aquaculture**

April 2018, Volume 491, Pages 105-113

<http://dx.doi.org/10.1016/j.aquaculture.2018.03.006><http://archimer.ifremer.fr/doc/00429/54061/>

© 2018 Elsevier B.V. All rights reserved.

**Archimer**<http://archimer.ifremer.fr>

---

## Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*

Harney Ewan<sup>1,2,\*</sup>, Lachambre Sébastien<sup>1,3</sup>, Roussel Sabine<sup>1</sup>, Huchette Sylvain<sup>3</sup>, Enez Florian<sup>4</sup>, Morvezen Romain<sup>4</sup>, Haffray Pierrick<sup>4</sup>, Boudry Pierre<sup>2</sup>

<sup>1</sup> Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR 6539 CNRS/UBO/IRD/Ifremer, Institut Universitaire Européen de la Mer, University of Brest (UBO), Technopôle Brest Iroise, 29280 Plouzané, France

<sup>2</sup> Ifremer, UMR 6539 LEMAR (UBO/CNRS/IRD/Ifremer), Centre Bretagne, 29280 Plouzané, France

<sup>3</sup> France Haliotis, 29880 Plouguerneau, France

<sup>4</sup> Syndicat des Sélectionneurs Avicoles et Aquacoles Français (SYSAAF), 35000 Rennes, France

\* Corresponding author : Ewan Harney, email address : [ewan.harney@gmail.com](mailto:ewan.harney@gmail.com)

---

### Abstract :

Selective breeding strategies require pedigree information over generations, but many species produced in aquaculture are too small to be physically tagged at early stages. Consequently, maintaining a sufficient number of separate families is often needed but costly and logistically difficult. Alternatively, parentage assignment can be obtained using DNA markers. We developed a panel of single nucleotide polymorphism (SNP) markers for the European abalone *Haliotis tuberculata* using an existing transcriptomic resource. An initial set of 2,176,887 SNPs was filtered to select 500 for high throughput genotyping. Of these, 298 SNPs were amplified in at least 90% of our *H. tuberculata* samples, consisting of a mixed family cohort (945 offspring) generated by crossing 40 abalones, and 5 full-sib training families (70 offspring). Based on amplification success among parents, minimum allele frequency and checks carried out against the training families, a subset of 123 markers was used to carry out parentage assignment in our mixed family cohorts. Maximum likelihood and exclusion-based methods of parentage assignment yielded consistent results, allowing parentage to be assigned in 98.9% of the studied progeny. Optimization of markers suggests that the 60 most informative SNPs may be sufficient for 95% assignment success in these progeny. The panel was also used to estimate effective population size, and revealed a low  $N_e$  due to high variance of reproductive success between parents. Our panel could be used to estimate genetic parameters of traits in mixed family cohorts, an essential stage to initiate selective breeding in *H. tuberculata*. It could also be useful tool in the context of monitoring stock enhancement and population genetics studies.

---

## Highlights

► This is the first SNP panel developed for parentage assignment in an abalone species. ► SNPs discovery was carried out using an existing transcriptomic data set, and yielded a panel of 298 SNPs with >90% call success across >1000 genotyped individuals. ► A subset of 123 SNPs was used to successfully assign parentage in 98.9% of 945 offspring from 40 parents representing 189 mixed families. ► Parentage assignment revealed that broodstock reproductive success was highly variable, and suggests that careful management of genetic variability would be needed if implementing individual selection in hatchery produced progenies.

**Keywords** : Abalone, Parentage assignment, Pedigree, Selective breeding, Transcriptome, SNP

## Abbreviations

- SNP, Single Nucleotide Polymorphism;
- QTL, Quantitative Locus Traits;
- $N_e$ , Effective population size;
- PIC, Polymorphic Information Content

## Introduction

Abalone production is an emerging industry in Europe, and is mainly based on the European abalone *Haliotis tuberculata* (Huchette and Clavier, 2004), although *Haliotis discus hannai* and *Haliotis rufescens* have been introduced to Ireland (Hannon et al., 2013) and Iceland (Jonasson et al., 1999) respectively for aquaculture production. *H. tuberculata* takes between 3 and 4 years to reach a minimal commercial size of 30g, and no intentional selective process has yet been initiated (Lachambre et al., 2017). This slow growth leads to high production costs and increases the risk of accidents during the production cycle. To reduce growing time, many producers and breeders have adopted selective breeding programs for the different abalone species farmed worldwide (Elliott, 2000). These programs require considerable investment in the long run, including dedicated research projects, special rearing facilities, precise traceability and phenotyping costs (Gjedrem and Baranski, 2010). Furthermore, as with many aquatic species, abalones are too small to be physically tagged in the early stage of production, which can complicate the implementation of family-based selective breeding programs (Vandeputte and Haffray, 2014). Either all the families are mixed at the larval stage, with individuals selected only on the basis of their phenotype at the end of the rearing without pedigree information, or all the families are reared separately until physical tagging.

Individual selection can generate improved phenotypes, but cannot control for inbreeding unless multiple cohorts are selected and intercrossed. Rearing families separately can control inbreeding, but is costly and heritability estimates may be influenced by common environmental conditions between tanks (Gjedrem and Baranski, 2010).

Alternatively, DNA markers can be used to identify the pedigree of animals reared in mixed families under commercial conditions; parentage assignment with DNA markers is thus an

invaluable tool for assessing the extent of inbreeding and can help to optimise the breeding strategy during the domestication process (Vandeputte and Haffray, 2014). Until recently, microsatellites were the most commonly used genetic marker for parentage analysis in non-model organisms (Jones et al., 2010; Weinman et al., 2015), including species of interest for aquaculture (Liu et al., 2017; Trøng et al., 2013). Among different abalone species, several studies have used microsatellites to infer pedigree, resulting in up to 90-95% assignment success (Hara and Sekino, 2007; Lucas et al., 2006; van den Bergb and Roodt-Wilding, 2010). Yet recent advances in gene sequencing and genotyping technologies have allowed the use of alternative markers such as single nucleotide polymorphisms (SNPs) in parentage assignment. Such markers display a number of advantages over traditional microsatellite markers, including better assignment (Sellars et al., 2014; Trøng et al., 2013), more accurate genotyping (Anderson and Garza, 2006), fewer null alleles (van den Bergb and Roodt-Wilding, 2010) and the potential for quantitative trait loci (QTL) detection (Avia et al., 2017; Gutierrez et al., 2014). Furthermore, SNPs can be derived from increasingly common expressed sequence data such as the short reads and transcriptomes generated during RNA-Seq experiments (De Wit et al., 2015), making the development of novel markers straightforward compared to microsatellites. Non-synonymous substitutions in SNPs from coding regions may help to identify candidate genes for marker assisted selection (Merwe et al., 2013), and more broadly can be used in genetic analyses to understand the structure of the population and manage the stock (Rhode et al., 2017).

The number of markers required for successful parentage assignment can be quite variable, depending on the heterozygosity of the markers in progenies. Liu et al. (2017) found that 50 SNPs allowed 100% assignment in the Pacific oyster *Crassostrea gigas*; yet in another study

of a different population of *C. gigas*, Lapègue et al. (2014) found that 150 SNPs were necessary for unambiguous parentage assignment. While assignment success for SNP panels is clearly influenced by numerous population genetic factors and by the quality of the markers, optimization approaches based on using the most polymorphic markers (Perez-Enriquez and Max-Aguilar, 2016) or testing subsets of SNPs from a larger panel (Holman et al., 2017) can help to determine smaller optimal panels for particular populations. The assignment success of a panel may also be influenced by the method used to achieve it. Numerous methodologies for assigning parentage exist (Jones et al., 2010), but among the most widely used are exclusion-based methods and maximum likelihood-based methods. Exclusion-based methods are sensitive to genotyping errors whereas likelihood-based methods generally give higher assignment rates (especially with low power marker sets), but can give inconsistent results (Vandeputte and Haffray, 2014). Direct comparisons of the two methods (e.g. Perez-Enriquez and Max-Aguilar, 2016; Trøng et al., 2013) showed similar levels of assignment were achieved with the same number of markers. Furthermore, Trøng et al. (2013) found that likelihood-based methods could help resolve certain cases in which exclusion methods assigned multiple parents. Comparing the results from multiple methods can increase assignment confidence (Morvezen et al., 2013), and help to determine the optimal number of markers necessary to assign parentage in a given population.

In the context of selective breeding programs, an important output of an SNP panel is its capacity to estimate effective size ( $N_e$ ) of populations under selection (van den Bergb and Roodt-Wilding, 2010), which is an important indicator of inbreeding. The  $N_e$  of cultured populations is often lower than the number of parents, a fact that is especially true for aquatic species which are characterized by high fecundity and high variability in reproductive success

(Boudry et al., 2002; Hedgecock, 1994). In closed populations such as selected farm populations, low  $N_e$  (i.e. high genetic drift) can lead to a deleterious loss of diversity (Falconer and Mackay, 1996). If no temporal information is available,  $N_e$  can be calculated from the variability of reproduction success of the different parents contributing to a cohort (Chevassus, 1989).  $N_e$  can also be estimated from genetic parameters such as heterozygotes excess (Zhdanova and Pudovkin, 2008) or linkage disequilibrium (Waples and Do, 2008). Methods for estimating  $N_e$  with molecular markers are still being developed and tested (Waples, 2016), and different methods applied to the same data set often result in different estimates of  $N_e$  (Morvezen et al., 2016; Wang, 2016; Waples and Do, 2010); however, comparison of these values can provide a measure of confidence, particularly if they are based on independent information (Waples, 2016). In abalone hatcheries,  $N_e$  has been estimated at 18 to 75 for *H. rubra* farms in Australia (Evans et al., 2004), and, at 15-67 for *H. midae* following two generations of selection on farms in South Africa (Rhode et al., 2014). In *H. midae*, these  $N_e$  estimates were associated with elevated rates of inbreeding (5%), and assessing  $N_e$  is thus an important step in understanding loss of genetic diversity when implementing new breeding practices.

The first aim of this study was to design a panel of SNPs using data from a transcriptomic database for high throughput parentage assignment. The second aim was to test the panel's utility for parentage assignment in a real cohort reared during three years at an abalone farm. Finally the inferred pedigree was used to estimate  $N_e$  in this cohort.

## Materials and Methods

### 2.1. Rearing of mixed families cohorts and bi-parental families

A total of 945 abalone were reared during 32 months at France Haliotis (48°36'46N, 4°33'30W; Plouguerneau, France), a commercial French hatchery. The studied progeny resulted from a full factorial mating of 24 males and 16 females randomly sampled from France Haliotis farmed stock. Two epipodia (approximately 50 mg of tissue) were sampled from each parent after the spawning and stored in 70% ethanol. Farm stock from France Haliotis was not under intentional selective pressure and animals were produced from systematic crosses between wild and farmed individuals (either males or females were wild broodstock) in order to avoid inbreeding.

Progeny were reared under common conditions with the standard production system of a sea-based abalone farm. More details on hatchery procedure can be found in Daume et al. (2000). A full factorial mating design was created by fertilizing eggs from 16 dams with the sperm of 24 sires. Briefly, the male and female gametes were collected separately by individual spawning or sperm ejaculation in beakers. Eggs of the different dams were equalized in number and pooled. They were then divided between 24 buckets of 20 L, with each bucket fertilized by the sperm from a single male. After fertilization, all embryos were pooled and abalone larvae shared a common larvae tank. They were induced to settle in one nursery tank, where they were fed with seaweed cultured on plates, mainly *Ulvelia lens* and diatoms. Abalones were reared for 14.5 months in nursery tanks, after which 3000 juveniles were randomly chosen to be sent to sea cages. Limited mortality was observed during the cage rearing. Abalones were reared at standard farm density and fed monthly with macro-algae (mainly *Palmaria palmata* and *Laminaria digitata*) during 17 months in sea cages, during



which time limited mortality was observed. After this rearing procedure (31 months in total), animals were brought back to the laboratory facility for tissue samples. For each of 945 individuals, a piece of foot (approximately 50 mg) was cut and stored in 70% ethanol for DNA extraction.

In addition, a training set of five full sib families were produced separately. A piece of tissue from the parents was sampled by cutting a small fragment of the basal part of the foot at the time of spawning. Settlement plates were put in 20L buckets after two months in the nursery to prevent any mixing. The juveniles were reared during 4 months in the bucket with constant input of water and fed with dry powder of *Palmaria palmata* two times a week until a size of 8 mm. At that time the foot (approximately 50 mg) was sampled from 70 juveniles (between 11 and 15 offspring per training family) and stored in 70% ethanol for DNA extraction. DNA samples from both mixed families and full sib families were sent to LGC Genomics (Hoddesdon, UK) for DNA extraction and genotyping.

## 2.2. SNP discovery

The transcriptomic data which served as the basis for SNP discovery are described in detail in Harney *et al.*, (2016), but a brief overview is provided here. The transcriptome comprised 2 distinct datasets: an experiment on early development in *H. tuberculata* consisting of 24 RNA-Seq libraries of several thousand pooled larvae reared under different environmental conditions; and an experiment on the immune response of adult *H. tuberculata* to the pathogen *Vibrio harveyi*, consisting of 24 RNA-Seq libraries derived from 12 individual hemolymph samples (a total of 48 libraries across 36 samples). Larval samples originated

from the hatchery France Haliotis, while adult samples were from one of two wild populations: Molène (France, 48° 23' 40" N, 4° 57' 33" W) and Saint-Malo (France, 48° 38' 57" N, 2° 1' 32" W). Illumina sequencing generated more than 1.5 billion paired-end reads. Filtration of low quality reads left 1.3 billion high quality reads (790,015,978 for larvae and 544,556,316 for adults) which were used in the generation of a shotgun transcriptome assembly using TRINITY (Haas et al., 2013). Following filtration, this resulted in a transcriptome of 41,099 non-redundant contigs (i.e. not including potential splice variants or gene duplicates) with an average length of 1,015 base pairs. The raw sequence data and reference transcriptome are available from the NCBI under the accessions SRA303338 and GEAU01000000 respectively.

Variant detection followed the WGS workflow proposed by the creators of SAMTOOLS (<http://www.htslib.org/workflow/>, accessed 07/02/2018). Initially the *H. tuberculata* transcriptome was indexed using the BURROWS WHEELER ALIGNER (Li and Durbin, 2009). SAMTOOLS (Li et al., 2009) was then used to map the 1.3 billion high quality reads to the reference. PICARD was used to mark and remove duplicates (<http://broadinstitute.github.io/picard/>, accessed 07/02/2018) and realignment was carried out with the GENOME ANALYSIS TOOLKIT (McKenna et al., 2010). Finally SAMTOOLS was used to combine reads from the same samples that were in different libraries, and to produce a variant call file.

### 2.3. SNP Filtration

The variant call file was initially subject to a heavy filtration using the filter argument in SAMTOOLS. All non-biallelic alleles and SNPs less than 50 base pairs from the start of the contig were removed; furthermore, the 'thin' argument was set to 50, to prevent inclusion of SNPs that were less than 50 base pairs from one another. Any SNPs that were not present in at least 34 of the 36 samples from Harney et al. (2016) were also removed. Finally, SNPs with a minor allele frequency (MAF) less than 0.2 were also filtered. Although pooled samples do not provide accurate estimates of allele frequencies, this filtration step still allows SNPs with particularly high or low allele frequencies to be identified and removed.

Following this filtration, phred scaled quality scores and read depth were assessed to ensure only high quality SNP sequences remained. Because larval RNA samples had not been sequenced with strand specific primers, transcriptome assembly produced contigs with unknown sense. Sense was therefore inferred by running the open reading frame (ORF) detection program TRANSDECODER (Haas et al., 2013) against all the contigs remaining in the data set. In cases where contigs contained multiple ORFs of the same sense, this was taken to be the correct direction. To resolve cases where ORFs were found in both the sense and antisense direction, BLASTP (Altschul et al., 1990) was used to search for hits against the UniProt database. In the majority of cases, blast hits either matched ORFs in only one direction, or blast hit sense matched ORF sense, resolving the sense of the contig. In cases where ORF and blast hit senses did not correspond, contigs and the SNPs associated with them were removed from the analysis.

Following sense assessment, SNPs in which heterozygosity was low (no heterozygotes observed among the 12 adults genotyped) were removed, and linkage disequilibrium scores

(LD) were assessed in cases where contigs carried multiple SNPs. As with AF estimation, LD for pooled samples is not strictly accurate, but still provides an indication of SNPs that are not in equilibrium. For SNP pairs with high LD scores ( $r^2 > 0.1$ ), a single one was selected to take forward on the basis of one of the quality values (base quality bias score). Assay design required 50 base pairs either side of the SNP; therefore when SNPs were within 100 base pairs of each other on the same contig, a single SNP was taken forward on the basis of quality, and SNPs less than 50 base pairs from the end of the contig were excluded. Filtration to the final target of 500 SNPs was achieved by using quality scores provided by SAMTOOLS (for a list of the quality scores and thresholds used, see supplementary table 1). These 500 SNPs and the 50 bp flanking sequences either side of the SNP were sent to LGC Genomics for assay design.

#### **2.4. Genotyping and selection of an SNP subset**

DNA extractions were performed by LGC genomics according to their standard protocol, with tissue lysis achieved using Proteinase K, and subsequent binding, washing and elution steps achieved using proprietary buffers. Following DNA extraction, LGC Genomics performed high throughput genotyping using KASP (Kompetitive Allele Specific PCR) assays, allowing bi-allelic scoring of SNPs (Semagn et al., 2014). Following genotyping, clusterplots were assessed, and SNPs for which homozygote and heterozygote groups did not cluster clearly were removed. Furthermore, only SNPs with  $> 90\%$  genotyping success among all samples (mixed cohort and training families) were retained for the panel of SNP markers.

In order to carry out the parentage assignment, a subset of this panel was selected. Only SNPs with  $MAF > 0.1$  (among all samples) and with a call success rate of  $> 90\%$  (among the

parents of the mixed cohort families) were selected, as these would likely be the most informative. To assess the quality of the markers, we compared parent and offspring genotypes and SNP heredity among the five full-sib training families. At this point, any SNPs with < 90% genotyping success among training family parents were removed. SNPs for which offspring genotypes did not match parental genotypes, or for which offspring genotypes were deemed improbable, were removed (although up to two genotyping 'errors' among offspring were allowed per SNP/family, i.e. an error rate of less than 20%). Additionally, clusterplots for these SNPs were inspected using SNPVIEWER2 (LGC), and those with ambiguous clustering of heterozygotes (e.g. the presence of two groups of heterozygotes) were removed.

## **2.5. Parentage assignment**

Parentage assignment for the mixed families using the SNP subset was performed with CERVUS 3.0 (Kalinowski et al., 2007) using a maximum likelihood based methodology and with VITASSIGN 8.5 (Vandeputte et al., 2006) using an exclusion based methodology. The outputs of the two programs were compared. A simulation was carried out to obtain a theoretical assignment power of the SNP panel with CERVUS with the following parameters: 10,000 offspring generated, 24 sires, 16 dams, 100% parents sampled, 99% of loci typed, and 1% error rate. For each individual, assignment was awarded to the most probable pair of parents according to a log-likelihood ratio (LOD score). The observation of the mismatch distribution of offspring-dam-sire trios informed the decision to allow VITASSIGN up to 5 mismatches ( $\approx$  2% of discordance within a trio). For both CERVUS and VITASSIGN, all crosses between the 24 sires and 16 dams were assumed to be possible.

## 2.6. Optimization of the SNP panel

The optimization of the SNP panel was carried out using the most conservative approach of parentage assignment, the exclusion method. In order to know the minimum number of markers required to obtain an assignment rate of 95%, SNPs were ordered by information content, represented by their MAF. Assignment rate was then calculated for the top five, 10, 15, 20 (etc., up to 120) best SNPs ranked by VITASSIGN, allowing zero, one, two, three, four, or five mismatches for each step.

## 2.7. $N_e$ estimation methodologies

Effective population size ( $N_e$ ) was estimated using both a basic formula and using linkage disequilibrium methods. The basic formula followed Chevassus (1989), and is expressed as:

$$N_e = \frac{4(N - 2)}{\left(K_s + \frac{V_s}{K_s}\right) + \left(K_d + \frac{V_d}{K_d}\right) - 2}$$

Where  $N$  is the total number of offspring,  $K_s$  and  $K_d$  are the mean numbers of offspring per sire and dam, and  $V_s$  and  $V_d$  are the variances in the number of offspring per sire and dam.  $N_e$  was then estimated with NEESTIMATOR V2.01 (Do et al., 2014) using linkage disequilibrium methods (Waples and Do, 2008). For this analysis, all SNPs with a minimum allele frequency (MAF)  $<0.05$  were excluded (as advised by Waples & Do, 2008).

## Results

### 3.1. SNP discovery and filtration *in silico*

The variant call file (vcf) output by SAMTOOLS identified a total of 2,176,887 SNPs in the *H. tuberculata* transcriptome. The initial heavy filtration removed nearly 99% of these, leaving 2,368 SNPs across 1,923 contigs, all of which had phred scaled quality scores above 200 and read depths between 150 and 310. TRANSDECODER identified 556 contigs with at least one ORF (a total of 986 ORFs), which in combination with BLASTP results allowed the sense of the contig to be resolved in most cases. Just 10 contigs contained ORF and blast hit senses that did not correspond, and so these contigs (and their associated SNPs) were removed from the analysis. This resulted in a list of 706 SNPs across 546 contigs. Following the removal of SNPs for which no heterozygous adult individuals were present, 639 SNPs on 478 contigs remained. A check for linkage disequilibrium and the removal of SNPs that were too close to each other, or to the end of the contig, reduced this to 582 SNPs across 478 contigs, from which 500 SNPs across 441 contigs were chosen based on Samtools quality scores (see supplementary table S1). Two SNP sequences that failed LGC Genomics' assay design were replaced with sequences from the list of 582 SNPs. Filtration steps are summarized in Fig. 1. Full information relating to these SNPs can be found in Harney et al. (2017).

### 3.2. Selection of informative SNPs

From the 500 SNPs that were genotyped, 298 SNPs showed at least 90% genotyping success among all samples (parents and offspring from both mixed family cohorts and full sib training families), and the average call success for these SNPs was 98.4%. This novel panel of SNP

markers is provided in supplementary table S2 and full genotyping results are available in Harney et al. (2017). Of these, 152 SNPs had an MAF  $> 0.1$  among all samples, and a call success rate of  $\geq 90\%$  among the 40 parents of the mixed-cohort families. Parent and offspring genotypes were assessed among the five full sib training families for these markers. For 21 of these SNPs, parent and offspring genotypes did not match in certain families (most likely because of genotyping errors among the parents), while for 8 SNPs, heterozygotes did not cluster as a single group in SNPVIEWER2 clusterplots. As a precaution, these 29 SNPs were also removed, leaving a subset of 123 SNPs for parentage assignment. The filtrations and quality controls used to produce this final group of SNPs are summarized in Fig. 1, and the subset of 123 SNPs used for parentage assignment is indicated in supplementary table S2.

### 3.3. Parentage assignment

The simulation carried out in CERVUS showed a potential estimation rate of 100% with the subset of 123 SNPs. Observed assignment rate of offspring from mixed family cohorts to a single pair of parents with CERVUS was 99.99%, with only two offspring with no probable parent pairs (negative LOD score). The exclusion method carried out by VITASSIGN gave 98.9% unambiguous assignment with 99% congruent parent pairs for these individuals compared to CERVUS. Four individuals were not assigned by VITASSIGN: the same two unassigned individuals found with CERVUS, and two individuals that had six mismatches with their most probable parent pair (and were therefore above our selected threshold of five mismatches). These individuals may have suffered from multiple genotyping errors. Seven other individuals were ambiguously assigned to two sires with an equal number of mismatches for either sire, illustrating a limitation of the exclusion method. Although the



likelihood method could help with identification of the correct sire, a conservative approach was chosen, and no sire was assigned to those individual in the final pedigree. Summary statistics for the 123 SNPs (observed and expected heterozygosities, PIC values and non-exclusion probabilities) from CERVUS are reported in supplementary table S3.

### **3.4. Optimization of SNP panel**

VITASSIGN calculations showed that it was not possible to achieve 90% assignment with mismatch values of 0 and 1 (Fig. 2). However, with three, four or five mismatches allowed, 95% assignment was obtained with the 60 most informative SNPs. Furthermore, for four or five mismatches, increasing the number of markers (up to 100) improved assignment success to 98%. When considering four or five mismatches, increasing the number of markers beyond 100 did not result in large changes in assignment success. On the other hand, for lower mismatches (zero, one or two), genotype errors resulted in declining assignment success beyond 65 markers, and with a dramatic fall in assignment success associated with the 10 least informative markers (Fig. 2).

### **3.5. Variability of reproductive success and $N_e$ estimates**

The reproductive success of parents was highly variable (Fig. 3.A). The variability of contribution was higher for dams than for sires: the three dams with the highest reproductive success contributed to 63% of the progeny, while the three dams with the least reproductive success had less than five offspring in the progeny (<0.5%). Of the 384 families possible with our crossing design, 189 (49%) were found in the progeny, with a high variability of number

of individuals per family (Fig. 3.B). Variability in reproductive success resulted in  $N_e$  estimates that were smaller than the initial number of parents, but generally consistent across different methods. The calculation based on Chevassus's formula (1989) estimated  $N_e$  to be 18, while the estimate based on linkage disequilibrium using NEESTIMATOR ( $N_e = 20.2$ , CI = 19.4 – 21.1) was slightly higher, but noticeably consistent.

## Discussion

In this study we successfully used a previously published transcriptomic dataset to generate a panel of informative SNPs for parentage assignment of the European abalone *H. tuberculata*. Comparing maximum-likelihood and exclusion-based methods to assign parentage yielded similar results and showed that a panel consisting of as few as 60 SNPs was enough to achieve more than 95% assignment success. Parentage assignment revealed high variance in reproductive success, which when taken with the  $N_e$  estimate of 20 suggests that this cohort could suffer from inbreeding following further selective breeding. However, the SNP panel presented here could be an invaluable tool for developing this program in the future.

We provide an example of the power of using existing transcriptomic resources for discovering SNPs for parentage assignment, and present a panel of 298 SNPs which were reliably genotyped in *H. tuberculata*. The success rate for SNP discovery was 59.6%, which is comparable to other recent studies in aquatic species which have developed novel SNP panels: Holman et al. (2017) successfully genotyped for 61.3% of tested markers in Atlantic salmon, while Gonçalves da Silva et al. (2015) were able to genotype 61.2% of SNPs in the orange roughy (*Hoplostethus atlanticus*). However, in both of these studies, SNP discovery

was based on novel genomic sequencing carried out as part of the study. Our use of an existing transcriptome saved a considerable amount of time and money. Our success rate was somewhat lower than the success rates of 70.3% and 78.4% observed by Lapègue *et al.*, (2014) in two oyster species, where transcriptomic data was also used. However it is interesting to note that Lapègue *et al.*, (2014) used both an *in vitro* approach (i.e. using targeted sequencing for SNP discovery) approach and a purely *in silico* approach (similar to our own, using existing transcriptomic data) for SNP discovery, and found broadly similar levels of genotyping success with both methods. Indeed, for the European flat oyster *Ostrea edulis*, *in silico* SNP discovery was more reliable than *in vitro* SNP discovery (Lapègue *et al.*, 2014). These results suggests that *in silico* SNP discovery using existing transcriptomic resources is a valid and cost-effective strategy for identifying polymorphic markers for parentage assignment.

To our knowledge this is the first SNP panel able to assign parentage in abalone. It demonstrates the utility of using low density SNP panels to achieve efficient assignment as observed in other aquatic species (Holman *et al.*, 2017; Lapègue *et al.*, 2014; Nolasco-Alzaga *et al.*, 2018; Sellars *et al.*, 2014). The assignment rate of 98.9% is higher than in previous studies that used microsatellites, which achieved between 82% (Slabbert *et al.*, 2009) and 95% (Hara and Sekino, 2007) assignment highlighting the utility of SNPs for parentage assignment in this genus. Of the 11 individuals not assigned in our study, two were clearly contaminants and two had six mismatches, and were probably not assigned because of cumulative genotyping errors. Allowing up to 2% of alleles to mismatch within a trio (here, 5 mismatches) is commonly done (e.g. Nguyen *et al.*, 2014; Rohrer *et al.*, 2007; Trøng *et al.*,

2013), and is necessary to avoid the strong effect of genotyping errors (Vandeputte et al. 2006).

In our study, the comparison of two main methods of assignment allowed the identification of seven individuals with high LOD scores in CERVUS but ambiguous assignment in VITASSIGN. The observation of the second pair of possible parents shows that this ambiguity derives from the sires. As these sires had a higher rate of missing data, this ambiguity does not seem to result from a weak power of exclusion for our SNP panel. It is unclear why certain males had more missing data, although if sires were related to one another, this may have complicated their assignment (Weinman et al., 2015).

Our optimization approach showed that a set containing the 60 best markers resulted in more than 95% unambiguous assignment for this particular cohort, allowing higher rates of assignment than any previous parentage assignment based on microsatellites. The use of 60 SNPs is congruent with the analysis of Holman et al (2017), who found that a subset of 60 SNPs from their panel of 94 total SNPs was sufficient for 100% assignment in 85% of simulations. These authors demonstrated considerable variation in the success of different panels of SNPs, highlighting the utility of optimization approaches for determining the most informative SNPs for a given population. Indeed, because allele frequencies and null alleles vary between populations, a cost-effective strategy for panel design may involve genotyping an initial group of individuals with all possible markers, carrying out an optimization step, and then selecting a subset of the most informative markers to genotype in the whole cohort. Furthermore, while ascertainment bias deriving from our choice of SNPs should not prove problematic for parentage assignment (Kaiser et al., 2017), the use of these markers to infer genetic parameters in other populations of *H. tuberculata* should be carried out with careful

consideration of how such bias may affect the analyses (Lachance and Tishkoff, 2013).

Appropriate steps can be taken to correct ascertainment biases, for example by incorporating them explicitly into subsequent population genetic models (Nielsen and Signorovitch, 2003).

Our study provides the first estimate of the variance of reproductive success in a full factorial cross for the European abalone. In our mating design, eggs were equalized by eye, pooled, then split and fertilized with the sperm of an individual male. This procedure was conducted to limit variance in egg number, which can have a strong effect on reproductive success (Bilbao et al., 2012), and avoid sperm competition, which has been suggested to exist in the abalone *H. asinina* (Selvamani et al., 2001). Nevertheless, parentage assignment showed that overall variability of reproductive success was high, particularly in dams. The three dams with the highest reproductive success contributed to 63% of the progeny, and the three sires with the highest reproductive success contributed to 31% of the progeny. Conversely, the three dams and three sires with the least reproductive success each contributed less than 9 offspring (<1%) towards the total progeny. Spawning protocols with strict, replicated counts of eggs may help to reduce this variation, although variation in reproductive success of broadcast spawning marine invertebrates may arise for many reasons, including differences in gamete proteins (Corporeau et al., 2012) or lipids (Boulais et al., 2015), genetic variation in fertilization success (Palumbi, 1999; Schlegel et al., 2012) or genetic inviability in crosses (Plough et al., 2016). This variation may also arise due to differential survival at later life stages; for example *C. gigas* families mixed during larval development show high variance in family representation in the post larval stage (Matson, 2010). High variability of reproductive success has also previously been reported in other abalone studies based on microsatellites (Hara and Sekino, 2007; Lucas et al., 2006). Variability was either balanced between the two

sexes (Slabbert and Roodt-Wilding, 2006; van den Bergb and Roodt-Wilding, 2010), or higher between sires (Hara and Sekino, 2007) when sperm competition was possible (Boudry et al., 2002). This high variability of reproductive success led to the absence of 51% of possible parental combinations in our sampled progeny. In contrast, Lucas et al. (2006) observed 89% of parental combinations in a smaller experiment (6 dams and 14 sires) with *H. asinina*. Their spawning protocol involved carrying out separate fertilizations for each sire/dam combination, which not only eliminated sperm competition, but may have reduced the effects of spawning asynchrony and variable egg quality on initial variability in reproductive success. Thus separate sire/dam fertilizations seem to be important for minimizing variance in reproductive success. Additionally, the effects of intra-gametic interactions and inter-familial variability in survival during the first three months of rearing have not previously been studied for abalone and may help to explain the variability in reproductive success that we observed (Boudry et al. 2002).

High variability in reproductive success is a common feature of marine broadcast spawning invertebrates, and high fecundity and high mortality can result in a sweepstakes pattern of reproductive success (Hedgecock, 1994; Hedgecock and Pudovkin, 2011), whereby only a small fraction of parents contribute towards the following generation. Even if many parents produce some offspring, a few highly successful parents producing many offspring may cause a reduction in  $N_e$  with respect to the census population size (Hedgecock and Pudovkin, 2011; Hedrick, 2005). Our calculation of  $N_e$  based on that of Chevassus (1989), which directly considered variation in reproductive success, suggested a rather low  $N_e$  of approximately 20; a result that was closely supported by our estimate of  $N_e$  based on LD methods (Do et al., 2014). In the context of abalone aquaculture, low  $N_e$  values have already been observed in

several studies (Evans et al., 2004; Rhode et al., 2014; Slabbert et al., 2009). A temporal  $N_e$  estimate of 50, evaluated over two successive generations, was associated with an inbreeding increase of 5% per generation in a commercial South African population of *H. midae* (Rhode et al., 2014). This level is higher than the 1% that is generally recommended for a sustainable breeding strategy (Bentsen and Olesen, 2002). Our results therefore confirm that the cohort study here is probably not large enough to support a selective breeding program in the long term. The SNP panel we present will, however, help in the development of this program. The panel could be used to assess population genetic parameters in the broodstock, and investigate whether the current reproductive strategy does result in any inbreeding. Over multiple generations, this SNP panel could be used for temporal estimates of  $N_e$ , allowing better understanding of genetic structure in the population (Rhode et al., 2017). Moreover, SNPs could be used for marker-assisted selection, ensuring a genetically diverse base population (Hayes et al., 2006; Loughnan et al., 2016) for future selective breeding programs that will maximize initial genetic variation. Markers can also be used to determine optimal breeding strategies that will reduce the relatedness of subsequent crosses (Hely et al., 2013; Herbinger et al., 1995).

## Conclusion

Here we present a novel panel of 298 SNP markers for the European abalone *H. tuberculata*. The panel was developed using an existing transcriptomic database and demonstrated comparable success rates to SNP discovery experiments based on intentionally collected genomic data. A subset of 123 markers had a theoretical potential assignment of 100%, and provided 98.9% assignment in a real farm cohort. The 60 most-informative markers were

sufficient to achieve more than 95% assignment. Parentage assignment revealed highly variable reproductive success, and  $N_e$  was estimated at 20, suggesting that selective breeding on this particular cohort may result in a higher-than-desirable rate of inbreeding.

## Acknowledgements

The authors would like to thank Xavier Lesage, Frederic Laurans, Michael Gleeson, Iain McKensy, and Maryvonne Leroux, from the production team of France Haliotis, for the 4 year rearing of the mixed cohort families. Abims (Analysis bioinformatics for marine science, Roscoff Marine Station) provided server space for bioinformatic analyses. In addition, the authors would like to thank Anastasia Bestin and Anne-Sophie Tiriau from the SYSAAF for help with phenotyping, and Thomas Bisch and Amy Courant for help with the demanding work of tissue sampling. E.H. was supported by a grant from the Regional Council of Brittany, from the European Funds (ERDF), and by the “Laboratoire d’Excellence” LabexMER (ANR-10-LABX-19) through a grant from the French government under the program “Investissements d’Avenir”. This work was conducted in the context of a PhD partially funded by France Haliotis and ANRT (CIFRE N° 2014 0643) and benefiting from the support of the European Union via program 42 of the FEAMP: “Innovation in Aquaculture”. We thank two anonymous reviewers for constructive criticisms.

## References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment



- search tool. *J. Mol. Biol.* 215, 403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anderson, E.C., Garza, J.C., 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172, 2567–2582. <https://doi.org/10.1534/genetics.105.048074>
- Avia, K., Coelho, S.M., Montecinos, G.J., Cormier, A., Lerck, F., Mauger, S., Faugeron, S., Valero, M., Cock, J.M., Boudry, P., 2017. High-density genetic map and identification of QTLs for responses to temperature and salinity stresses in the model brown alga *Ectocarpus*. *Sci. Rep.* 7, 43241. <https://doi.org/10.1038/srep43241>
- Bentsen, H.B., Olesen, I., 2002. Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture* 204, 349–359. [https://doi.org/10.1016/S0044-8486\(01\)00846-8](https://doi.org/10.1016/S0044-8486(01)00846-8)
- Bilbao, A., Uriarte, I., del Pino Viera, M., Sosa, B., Fernández-Palacios, H., Hernández-Cruz, C.M., 2012. Effect of macroalgae protein levels on some reproductive aspects and physiological parameters for the abalone, *Haliotis tuberculata coccinea* (Reeve 1846). *J. World Aquac. Soc.* 43, 764–777. <https://doi.org/10.1111/j.1749-7345.2012.00617.x>
- Boudry, P., Collet, B., Cornette, F., Hervouet, V., Bonhomme, F., 2002. High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* 204, 283–296. [https://doi.org/10.1016/S0044-8486\(01\)00841-9](https://doi.org/10.1016/S0044-8486(01)00841-9)
- Boulais, M., Corporeau, C., Huvet, A., Bernard, I., Quere, C., Quillien, V., Fabioux, C., Suquet, M., 2015. Assessment of oocyte and trochophore quality in Pacific oyster, *Crassostrea gigas*. *Aquaculture* 437, 201–207. <https://doi.org/10.1016/j.aquaculture.2014.11.025>
- Chevassus, B., 1989. Aspects génétiques de la constitution de populations d'élevage destinées au repeuplement. *Bull. Français la Pêche la Piscic.* 146–168. <https://doi.org/10.1051/kmae:1989010>
- Corporeau, C., Vanderplancke, G., Boulais, M., Suquet, M., Quéré, C., Boudry, P., Huvet, A., Madec, S., 2012. Proteomic identification of quality factors for oocytes in the Pacific oyster *Crassostrea gigas*. *J. Proteomics* 75, 5554–5563. <https://doi.org/10.1016/j.jprot.2012.07.040>
- Daume, S., Krsinich, A., Farrell, S., Gervis, M., 2000. Settlement, early growth and survival of *Haliotis rubra* in response to different algal species. *J. Appl. Phycol.* 12, 479–488. <https://doi.org/10.1023/a:1008110828581>
- De Wit, P., Pespenti, M.H., Palumbi, S.R., 2015. SNP genotyping and population genomics from expressed sequences — current advances and future possibilities. *Mol. Ecol.* 24, 2310–2323. <https://doi.org/10.1111/mec.13165>
- Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Mol. Ecol. Resour.* 14, 209–214.

<https://doi.org/10.1111/1755-0998.12157>

- Elliott, N.G., 2000. Genetic improvement programmes in abalone: What is the future? *Aquac. Res.* 31, 51–59. <https://doi.org/10.1046/j.1365-2109.2000.00386.x>
- Evans, B.S., Bartlett, J., Sweijd, N., Cook, P., Elliott, N.G., 2004. Loss of genetic variation at microsatellite loci in hatchery produced abalone in Australia (*Haliotis rubra*) and South Africa (*Haliotis midae*). *Aquaculture* 233, 109–127. <https://doi.org/10.1016/j.aquaculture.2003.09.037>
- Falconer, D.S., Mackay, T.F.C., 1996. *Introduction to Quantitative Genetics*. 4th edition, Fourth Edi. ed. Longman, London, UK.
- Gjedrem, T., Baranski, M., 2010. *Selective Breeding in Aquaculture: an Introduction*. Vol. 10. Springer Science & Business Media, New York, USA.
- Gonçalves da Silva, A., Barendse, W., Kijas, J.W., Barris, W.C., McWilliam, S., Bunch, R.J., McCullough, R., Harrison, B., Hoelzel, A.R., England, P.R., 2015. SNP discovery in nonmodel organisms: Strand bias and base-substitution errors reduce conversion rates. *Mol. Ecol. Resour.* 15, 723–736. <https://doi.org/10.1111/1755-0998.12343>
- Gutierrez, A.P., Lubieniecki, K.P., Fukui, S., Withler, R.E., Swift, B., Davidson, W.S., 2014. Detection of quantitative trait loci (QTL) related to grilising and late sexual maturation in Atlantic salmon (*Salmo salar*). *Mar. Biotechnol.* 16, 103–110. <https://doi.org/10.1007/s10126-013-9530-3>
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M.G., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hannon, C., Officer, R.A., Le Dorven, J., 2013. Review of the technical challenges facing aquaculture of the European abalone *Haliotis tuberculata* in Ireland. *Aquac. Int.* 21, 243–254. <https://doi.org/10.1007/s10499-012-9584-7>
- Hara, M., Sekino, M., 2007. Parentage testing for hatchery-produced abalone *Haliotis discus hannai* based on microsatellite markers: Preliminary evaluation of early growth of selected strains in mixed family farming. *Fish. Sci.* 73, 831–836. <https://doi.org/10.1111/j.1444-2906.2007.01403.x>
- Harney, E.D., Dubief, B., Boudry, P., Basuyaux, O., Schilhabel, M.B.M.B., Huchette, S., Paillard, C., Nunes, F.L.D.F.L.D., 2016. De novo assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome. *Mar. Genomics* 28, 11–16. <https://doi.org/10.1016/j.margen.2016.03.002>
- [dataset] Harney, E.D., Lachambre, S., Roussel, S., Huchette, S., Enez, F., Morvezen, R., Haffray, P., Boudry, P., 2017. SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*. SEANOE.

<https://doi.org/http://doi.org/10.17882/51704>

- Hayes, B., He, J., Moen, T., Bennewitz, J., 2006. Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture* 255, 573–578. <https://doi.org/10.1016/j.aquaculture.2005.11.038>
- Hedgcock, D., 1994. Does variance in reproductive success limit effective population sizes of marine organisms, in: Beaumont, A.R. (Ed.), *Genetics and Evolution of Aquatic Organisms*. Chapman & Hall, London, UK, pp. 122–134.
- Hedgcock, D., Pudovkin, A.I., 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: A review and commentary. *Bull. Mar. Sci.* 87, 971–1002. <https://doi.org/10.5343/bms.2010.1051>
- Hedrick, P., 2005. Large variance in reproductive success and the  $N_e/N$  ratio. *Evolution* (N. Y). 59, 1596–1599. <https://doi.org/10.1038/438753a>
- Hely, F.S., Amer, P.R., Walker, S.P., Symonds, J.E., 2013. Optimised parent selection and minimum inbreeding mating in small aquaculture breeding schemes: a simulation study. *Animal* 7, 1–10. <https://doi.org/10.1017/S1751731112001371>
- Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M., Cook, D., 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245–256. [https://doi.org/10.1016/0044-8486\(95\)01109-9](https://doi.org/10.1016/0044-8486(95)01109-9)
- Holman, L.E., Garcia de la Serrana, D., Onoufriou, A., Hillestad, B., Johnston, I.A., 2017. A work flow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture* 476, 59–64. <https://doi.org/10.1016/j.aquaculture.2017.04.001>
- Huchette, S., Clavier, J., 2004. Status of the ormer (*Haliotis tuberculata* L.) industry in Europe. *J. Shellfish Res.* 23, 951–955.
- Jonasson, J., Stefansson, S.E., Gudnason, A., Steinarsson, A., 1999. Genetic variation for survival and shell length of cultured red abalone (*Haliotis rufescens*) in Iceland. *J. Shellfish Res.* 18, 621–625.
- Jones, A.G., Small, C.M., Paczolt, K.A., Ratterman, N.L., 2010. A practical guide to methods of parentage analysis. *Mol. Ecol. Resour.* 10, 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
- Kaiser, S.A., Taylor, S.A., Chen, N., Sillett, T.S., Bondra, E.R., Webster, M.S., 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Mol. Ecol. Resour.* 17, 183–193. <https://doi.org/10.1111/1755-0998.12589>
- 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. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>

- Lachambre, S., Huchette, S., Day, R., Boudry, P., Rio-Cabello, A., Fustec, T., Roussel, S., 2017. Relationships between growth, survival, physiology and behaviour — A multi-criteria approach to *Haliotis tuberculata* phenotypic traits. *Aquaculture* 467, 190–197. <https://doi.org/10.1016/j.aquaculture.2016.04.028>
- Lachance, J., Tishkoff, S.A., 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays* 35, 780–786. <https://doi.org/10.1002/bies.201300014>
- Lapègue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadiou, 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>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- 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>
- Loughnan, S.R., Smith-Keune, C., Jerry, D.R., Beheregaray, L.B., Robinson, N.A., 2016. Genetic diversity and relatedness estimates for captive barramundi (*Lates calcarifer*, Bloch) broodstock informs efforts to form a base population for selective breeding. *Aquac. Res.* 47, 3570–3584. <https://doi.org/10.1111/are.12807>
- Lucas, T., Macbeth, M., Degnan, S.M., Knibb, W., Degnan, B.M., 2006. Heritability estimates for growth in the tropical abalone *Haliotis asinina* using microsatellites to assign parentage. *Aquaculture* 259, 146–152. <https://doi.org/10.1016/j.aquaculture.2006.05.039>
- Matson, S.E., 2010. Development, Evaluation and Application of a Mixed-family Selective Breeding Method for the Pacific Oyster (*Crassostrea gigas*). Oregon State University.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriell, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110.20>
- Merwe, A.B. Der, Blaauw, S., Plessis, J. Du, Roodt-wilding, R., 2013. Transcriptome-wide single nucleotide polymorphisms (SNPs) for abalone (*Haliotis midae*): Validation and application using GoldenGate medium-throughput genotyping assays. *Int. J. Mol. Sci.* 14, 19341–19360. <https://doi.org/10.3390/ijms140919341>
- Morvezen, R., Boudry, P., Laroche, J., Charrier, G., 2016. Stock enhancement or sea ranching? Insights from monitoring the genetic diversity, relatedness and effective

- population size in a seeded great scallop population (*Pecten maximus*). *Heredity* 117, 1–7. <https://doi.org/10.1038/hdy.2016.42>
- Morvezen, R., Cornette, F., Charrier, G., Guinand, B., Lapègue, S., Boudry, P., Laroche, J., 2013. Multiplex PCR sets of novel microsatellite loci for the great scallop *Pecten maximus* and their application in parentage assignment. *Aquat. Living Resour.* 26, 207–213. <https://doi.org/10.1051/alr/2013052>
- Nguyen, T.T.T., Hayes, B.J., Ingram, B.A., 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>
- Nielsen, R., Signorovitch, J., 2003. Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium. *Theor. Popul. Biol.* 63, 245–255. [https://doi.org/10.1016/S0040-5809\(03\)00005-4](https://doi.org/10.1016/S0040-5809(03)00005-4)
- Nolasco-Alzaga, H.R., Perez-Enriquez, R., Enez, F., Bestin, A., Palacios-Mechetnov, E., Haffray, P., 2018. Quantitative genetic parameters of growth and fatty acid content in the hemolymph of the Whiteleg shrimp *Litopenaeus vannamei*. *Aquaculture* 482, 17–23. <https://doi.org/10.1016/j.aquaculture.2017.09.015>
- Palumbi, S.R., 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12632–12637. <https://doi.org/10.1073/pnas.96.22.12632>
- Perez-Enriquez, R., Max-Aguilar, A., 2016. Pedigree traceability in whiteleg shrimp (*Litopenaeus vannamei*) using genetic markers: A comparison between microsatellites and SNPs. *Ciencias Mar.* 42, 227–235. <https://doi.org/10.7773/cm.v42i4.2662>
- Plough, L. V., Shin, G., Hedgecock, D., 2016. Genetic inviability is a major driver of type III survivorship in experimental families of a highly fecund marine bivalve. *Mol. Ecol.* 25, 895–910. <https://doi.org/10.1111/mec.13524>
- Rhode, C., der Merwe, A.E.B., Roodt-Wilding, R., 2017. An assessment of spatio-temporal genetic variation in the South African abalone (*Haliotis midae*), using SNPs: implications for conservation management. *Conserv. Genet.* 18, 17–31. <https://doi.org/10.1007/s10592-016-0879-5>
- Rhode, C., Maduna, S.N., Roodt-Wilding, R., Bester-Van Der Merwe, A.E., 2014. Comparison of population genetic estimates amongst wild, F1 and F2 cultured abalone (*Haliotis midae*). *Anim. Genet.* 45, 456–459. <https://doi.org/10.1111/age.12142>
- Rohrer, G.A., Freking, B.A., Nonneman, D., 2007. Single nucleotide polymorphisms for pig identification and parentage exclusion. *Anim. Genet.* 38, 253–258. <https://doi.org/10.1111/j.1365-2052.2007.01593.x>
- Schlegel, P., Havenhand, J.N., Gillings, M.R., Williamson, J.E., 2012. Individual variability in reproductive success determines winners and losers under ocean acidification: a case study with sea urchins. *PLoS One* 7, e53118. <https://doi.org/10.1371/journal.pone.0053118>

- Sellars, M.J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, G.J., Barendse, W., Henshall, J., 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquac. Res.* 45, 417–426. <https://doi.org/10.1111/j.1365-2109.2012.03243.x>
- Selvamani, M.J.P., Degnan, S.M., Degnan, B.M., 2001. Microsatellite genotyping of individual abalone larvae: Parentage assignment in aquaculture. *Mar. Biotechnol.* 3, 478–485. <https://doi.org/10.1007/s1012601-0062-X>
- Semagn, K., Babu, R., Hearne, S., Olsen, M., 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. <https://doi.org/10.1007/s11032-013-9917-x>
- Slabbert, R., Bester, A., D'Amato, M., 2009. Analyses of genetic diversity and parentage within a South African hatchery of the abalone *Haliotis midae* Linnaeus using microsatellite markers. *J. Shellfish Res.* 28, 369–375. <https://doi.org/http://dx.doi.org/10.2983/035.028.0220>
- Slabbert, R., Roodt-Wilding, R., 2006. Non-destructive sampling of juvenile abalone using epipodial tentacles and mucus: method and application. *African J. Mar. Sci.* 28, 719–721. <https://doi.org/10.2989/18142320609504221>
- Trọng, T.Q., van Bers, N., Crooijmans, R., Dibbits, B., Komen, H., 2013. A comparison of microsatellites and SNPs in parental assignment in the GIFT strain of Nile tilapia (*Oreochromis niloticus*): The power of exclusion. *Aquaculture* 388–391, 14–23. <https://doi.org/10.1016/j.aquaculture.2013.01.004>
- van den Bergb, N.C., Roodt-Wilding, R., 2010. Parentage assignment in *Haliotis midae* L.: A precursor to future genetic enhancement programmes for South African abalone. *Aquac. Res.* 41, 1387–1395. <https://doi.org/10.1111/j.1365-2109.2009.02428.x>
- Vandeputte, M., Haffray, P., 2014. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5, 432. <https://doi.org/10.3389/fgene.2014.00432>
- Vandeputte, M., Mauger, S., Dupont-Nivet, M., 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Mol. Ecol. Notes* 6, 265–267. <https://doi.org/10.1111/j.1471-8286.2005.01167.x>
- Wang, J., 2016. A comparison of single-sample estimators of effective population sizes from genetic marker data. *Mol. Ecol.* 25, 4692–4711. <https://doi.org/10.1111/mec.13725>
- Waples, R.S., 2016. Making sense of genetic estimates of effective population size. *Mol. Ecol.* 25, 4689–4691. <https://doi.org/10.1111/mec.13814>
- Waples, R.S., Do, C., 2010. Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evol. Appl.* 3, 244–262. <https://doi.org/10.1111/j.1752-4571.2009.00104.x>

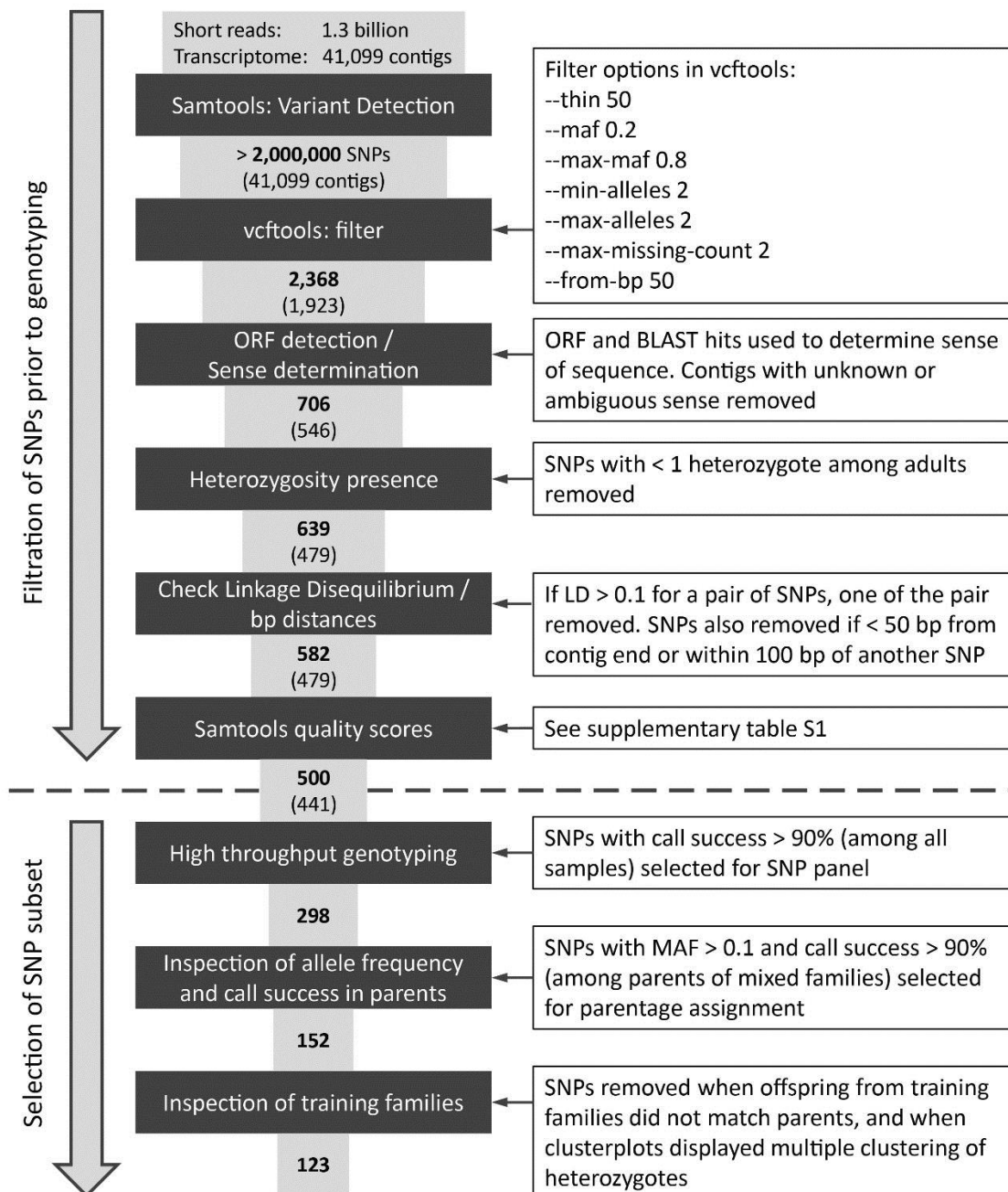
Waples, R.S., Do, C., 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Mol. Ecol. Resour.* 8, 753–756.  
<https://doi.org/10.1111/j.1755-0998.2007.02061.x>

Weinman, L.R., Solomon, J.W., Rubenstein, D.R., 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Mol. Ecol. Resour.* 15, 502–511.  
<https://doi.org/10.1111/1755-0998.12330>

Zhdanova, O.L., Pudovkin, A.I., 2008. Nb\_HetEx: A program to estimate the effective number of breeders. *J. Hered.* 99, 694–695. <https://doi.org/10.1093/jhered/esn061>

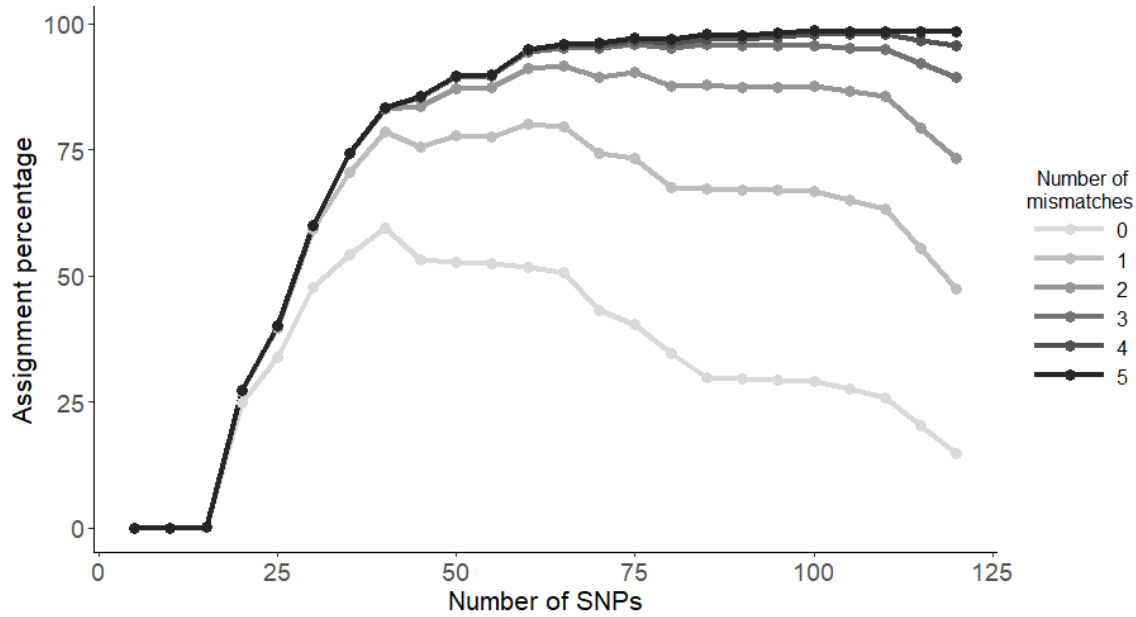
ACCEPTED MANUSCRIPT

## List of Figures

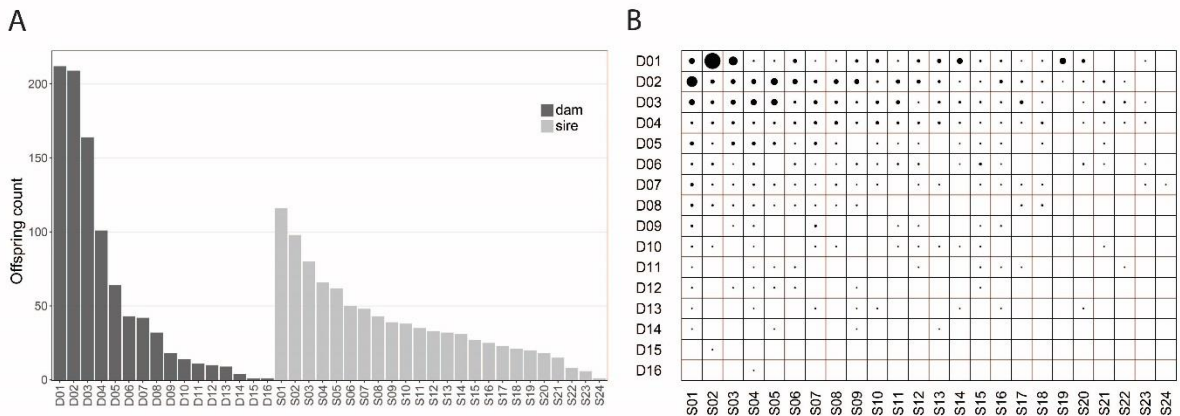


**Fig. 1.** Filtration steps applied to list of variants in order to produce 500 SNPs for high throughput genotyping, and subsequent selection of 123 informative SNPs following genotyping





**Fig. 2.** Change in single assignment rate with VITASSIGN for different levels of allowed mismatch and an increasing number of SNPs, added in groups of 5 at a time based on their MAF.



**Fig. 3. A)** Number of offspring in the progeny for every dam and sire in the mating scheme, **B)** Representation of the mating scheme with dams in lines and sires in columns, when individuals of every possible cross are presented on the progeny, a dot with an area proportional to the number of offspring is presented.

ACCEPTED MANUSCRIPT

SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*

### Highlights

5. This is the first SNP panel developed for parentage assignment in an abalone species.
6. SNPs discovery was carried out using an existing transcriptomic data set, and yielded a panel of 298 SNPs with > 90% call success across >1000 genotyped individuals.
7. A subset of 123 SNPs was used to successfully assign parentage in 98.9% of 945 offspring from 40 parents representing 189 mixed families.
8. Parentage assignment revealed that broodstock reproductive success was highly variable, and suggests that careful management of genetic variability would be needed if implementing individual selection in hatchery produced progenies.

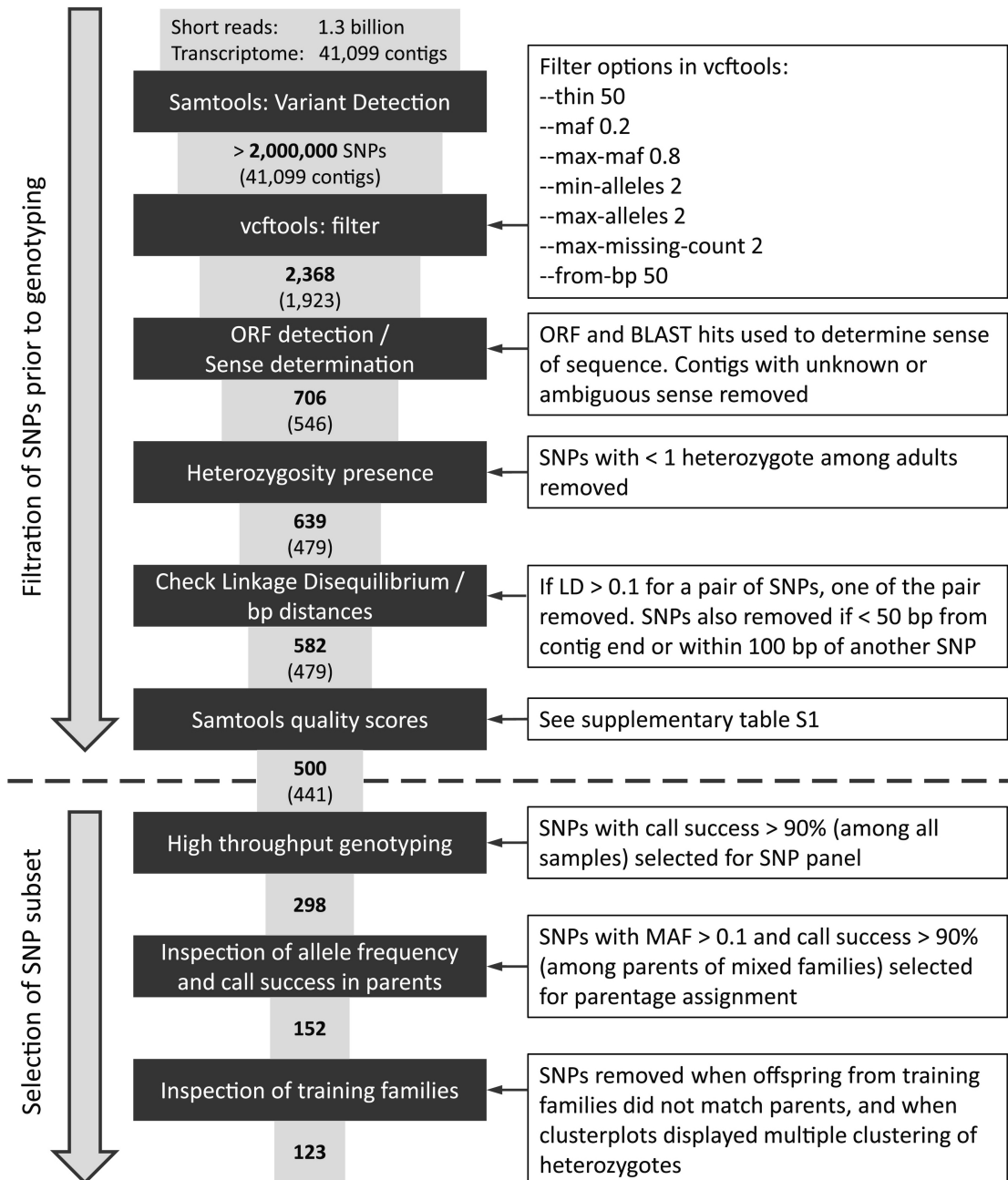


Figure 1

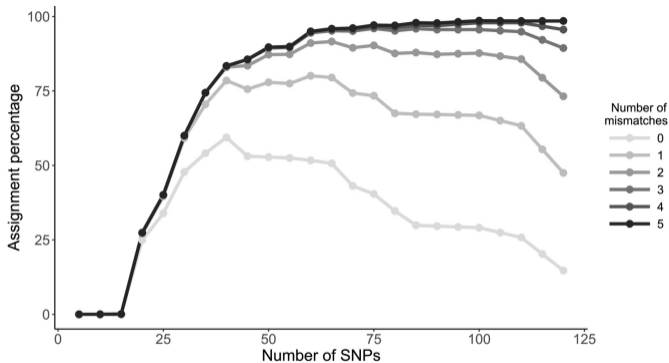
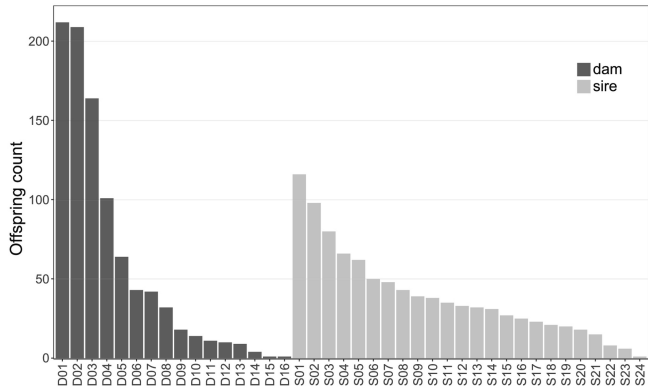


Figure 2

A



B

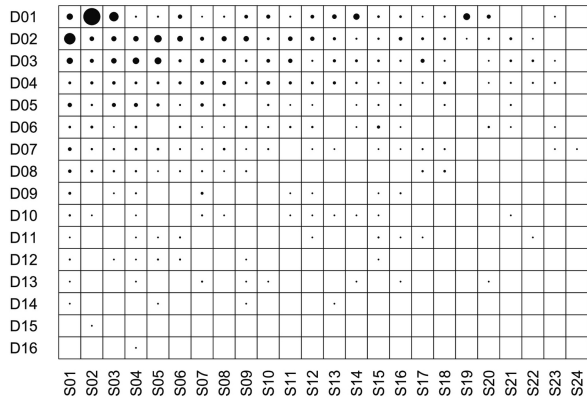


Figure 3