Technical Reports of the Institute of Cetacean Research (2024) pp. 8–14

Technical Report (not peer reviewed)

Development and implementation of SNP markers for whale research at the Institute of Cetacean Research's genetic laboratory based on ddRAD technique

Katrin Kiemel[1,*], Mioko Taguchi[1], Ralph Tiedemann[2] and Luis A. Pastene[1]

[1] *Institute of Cetacean Research, 4–5 Toyomi-cho, Chuo-ku, Tokyo 104–0055, Japan*

[2] *Unit of Evolutionary Biology/Systematic Zoology, Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht Straße 24-25, 14476 Potsdam, Germany*

Contact e-mail: kiemel.katrin@gmail.com

ABSTRACT

The inference of population structure and species determination has relied on the availability of molecular information. This information was traditionally gained through markers such as microsatellite DNA, single mitochondrial gene sequencing or Restriction Fragment Length Polymorphisms (RFLPs). However, with the advent of high throughput genomics, which began more than two decades ago with the development of next generation sequencing (NGS), there has been a significant shift in methodology. The increasing cost efficiency of NGS has led to a rapid increase in publications utilizing genome-wide single nucleotide polymorphisms (SNPs) to determine species, study population structure and allow to examine the related adaptations shaping the observed population structure. The use of SNPs allows for a more comprehensive analysis of genetic variation across entire genomes, often providing a more detailed and accurate understanding of fine scale population structure and species relationships than traditionally used markers. Furthermore, SNP assays (i.e., a defined number of informative SNPs) allow the screening of large numbers of samples with the added benefit of between-laboratory comparability. Currently, the Institute of Cetacean Research (ICR) employs traditional markers. However, the advantages of SNPs strongly support the need for a transition to this method. To facilitate the transition, a SNP pipeline has been established that allows the ICR to (i) identify SNPs from double digest Restriction-site Associated DNA (ddRAD) sequencing data and quality filter SNPs depending on different datasets, (ii) conduct the SNP analysis and (iii) select a SNP assay which facilitates further processing of unprocessed samples not subject to ddRAD sequencing or DNA samples of low quality.

INTRODUCTION

To accurately identify species and assess their population structure, molecular information is essential. Before the advent of high throughput technologies, scientists relied on methods such as Restriction Fragment Length Polymorphisms (RFLPs), microsatellite DNA and single gene sequencing, often focusing on the mitochondrial DNA (mtDNA), to study species and their population structures. However, using these markers, often resulted in unresolved fine scale population structures. The emergence of next generation sequencing (NGS) and its various methods to process DNA prior to sequencing, including double digest Restriction-enzyme Associated DNA (ddRAD) and Genotyping by Random Amplicon Sequencing-Direct (GRAS-Di), has introduced the high throughput era, enabling the transition from population genetics to population genomics (Hemmer-Hansen *et al.*, 2014; Hohenlohe *et al.*, 2021).

The analysis of Single Nucleotide Polymorphisms (SNPs) or entire genomes offers significant advantages by overcoming the resolution limitations of traditional population genetics often encountered in highly mobile species/groups such as cetaceans (Lah *et al.*, 2016). This approach leverages a vastly increased number of loci, often ranging from hundreds to several thousands, providing comprehensive genomic coverage that includes both coding and non-coding regions, which in turn allowing the identification of signs of selection and the investigation of adaptations that shape the observed population structures (e.g., Autenrieth *et al.*, 2024; Celemín *et al.*, 2023).

SNP assays, also referred to as panels (i.e., a fixed number of informative SNPs), have become particularly

valuable for their ability to enable a high comparability of population genetics conducted in different laboratories without the need for prior calibration, as is required for microsatellites (Ellis *et al.*, 2011). Technologies such as microfluidics (e.g., the Fluidigm System, Standard BioTools as shown in Figure 1), facilitate the high throughput processing of SNP assays through integrated fluidic circuits (IFCs), which can perform multiple PCRs simultaneously. This method allows semi-automated screening of large numbers of samples and SNPs, thereby significantly enhancing efficiency and consistency in genetic analysis (Fabbri *et al.*, 2012; Kraus *et al.*, 2014; Holman *et al.*, 2015).

Blue whales (*Balaenoptera musculus*), a highly mobile marine species, have been significantly depleted in the 19th and 20th century due to commercial whaling and are currently listed as an endangered species (IUCN, 2024). They have attracted considerable scientific interest, with population structure being studied using traditional markers such as the mtDNA control region, microsatellite DNA (e.g., LeDuc *et al.*, 2007; 2016; Torres-Lorez *et al.*, 2014), but also SNPs (Attard *et al.*, 2024). These studies have revealed pronounced genetic differentiation among ocean regions (*i.e.*, North Pacific, South Pacific, Southern Ocean and Indian Ocean). The combination of results based on genetics (LeDuc *et al.*, 2007; 2016; Torres-Lorez *et al.*, 2014), morphometrics (i.e., Branch *et al.*, 2007; Pastene *et al.*, 2020) and acoustics (McDonald *et al.*, 2006), suggest the existence of five subspecies (i.e., *B. m. musculus*, *B. m. intermedia*, *B. m. indica*, *B. m. brevicauda*, *B. m. unnamed subspecies*, aka Chilean blue whale).

To facilitate the transition from population genetics to population genomics for the investigation of cetaceans at the ICR, this study established a SNP panel pipeline, using blue whales as an example, which will allow the processing of data generated from ddRAD sequencing. The aim was to develop a pipeline which can (i) identify and quality filter SNPs, (ii) perform SNP analysis, and (iii) design a SNP panel allowing for species/population assessment and kinship analysis for subsequent use on the Fluidigm system available at the ICR.

SAMPLE SELECTION, DNA EXTRACTION AND ddRAD SEQUENCING

Samples (n=314) of the subspecies *Balaenoptera m. musculus*, *B. m. intermedia* (Antarctic blue whale), *B. m. brevicauda* (pygmy blue whale) *and B. m. unnamed subspecies* (aka) Chilean Blue whale were used for the development of SNP genotyping techniques at the ICR. These samples were biopsied under various surveys including the International Whaling Commission Pacific Ocean Whale and Ecosystem Research (IWC POWER), Japanese Whale Research Program under Special Permit in the western North Pacific, Phases II (JARPNII), New Scientific Whale Research Program in the North Pacific (NEWREP-NP), Japanese dedicated sighting surveys, IWC International Decade for Cetacean Research-Southern Ocean Whale and Ecosystem Research (IWC IDCR-SOWER), Japanese Whale Research Program under Special Permit in the Antarctic, Phases I and II (JARPA and JARPAII), New Scientific Whale Research Program in the Antarctic (NEWREP-A) and Japanese Abundance and Stock-structure Surveys in the Antarctic (JASS-A).

Samples were selected to cover global distribution from the following regions with sample sizes as indicated: Indian Ocean (IO) n=21, Southern Ocean (SO) n=224, eastern South Pacific (ESP) n=17, eastern North Pacific (ENP) n=10, western North Pacific (WNP) n=42. Total genomic DNA was extracted from approximately 0.05 g of

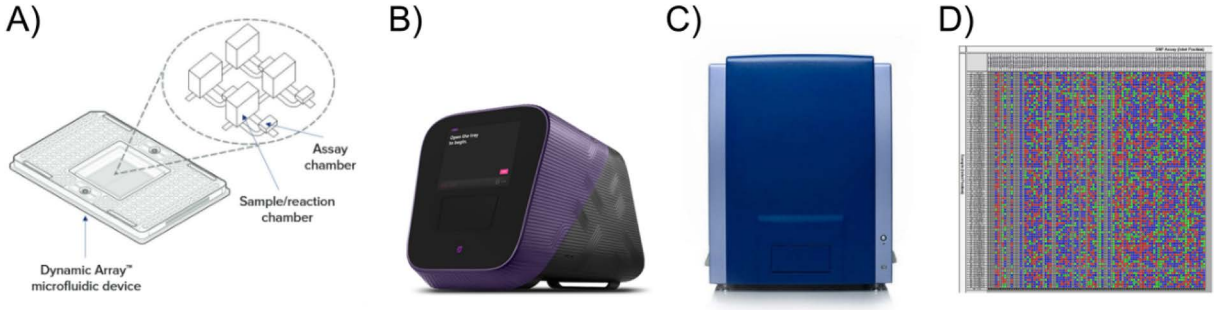

Figure 1. Fluidigm system consisting of (A) the IFC available in different sizes 24×96, 48×48 and 96×96, (B) JUNO to conduct PCRs and (C) EP1 to visualize results. Example of a result output is shown in (D). Each column is a sample while each row is a SNP. Red dots represent a homozygous call for allele (X/X), green dots represent a homozygous call for allele (Y/Y), blue dots represent a heterozygous call (X/Y), and grey dots indicate unsuccessful amplification. Figures (A), (B) and (C) are from Standard BioTools (2024).

tissue sample (i.e., skin, muscle or blubber), using either the phenol-chloroform method (Sambrook *et al.*, 1989) or the Gentra Puregene kit (QUIAGEN), following the manufacturer's protocol for animal tissue. The extracted DNA was stored in TE buffer at 4°C until further processing.

To retrieve ddRAD sequences, 25 µL of extracted DNA (5–10 ng/µL) was sent to the Giken Biotechnology Lab for sequencing. For genomic library preparation, 100 ng of DNA was treated with the restriction enzymes MspI and EcoRI for 3 hours at 37°C. The DNA was then cleaned using DNA Clean Beads (MGI Tech CO Ltd) and ligated for 16 hours using T4-DNA-Ligase (TaKaRa), followed by another cleaning with DNA Clean Beads. The library was amplified using PCR, and a size selection performed with a size range of 240–400 bp. Libraries were quantified using the Qubit dsDNA HS Assay kit and enriched with DNA Clean beads. Final libraries were checked with the Agilent 2100 Bioanalyzer and the High sensitivity DNA Kit (Agilent technology). Libraries were circularized using the MGIEasy Circularization kit (MGI Tech Co Ltd) as per the manufacturer's protocol. Paired-end reads of 100 bp or 200 bp were sequenced on a DNBSEQ G400 with a sequencing depth of 1–3 million read pairs, respectively. The company provided demultiplexed reads.

SNP IDENTIFICATION

To identify SNPs, demultiplexed reads were processed using radtags to check for intact barcodes and restriction enzyme cutting sites. Adapters were trimmed at the 3'-end and filtered for the mean quality using the program fastp (Chen, 2023). The reads were then mapped against the indexed reference genome of *B. m. musculus* (NCBI Accession number: GCA_009873245.3) using BWA-MEM (Li, 2013), and bam files were indexed using samtools index. SNPs were called using STACKS version 2.2 (Rochette *et al.*, 2019) and freebayes version 1.3.6 (Garrison & Marth, 2012). Info tags derived from freebayes were combined with the SNPs called by STACKS using bcftools annotate (Li *et al.*, 2011). SNPs were filtered based on a modified pipeline described by O'Leary *et al.* (2018) using vcftools version 0.1.19 (Danecek *et al.*, 2011), bcftools version 1.13 (Li *et al.*, 2011) and plink version 1.9 (Purcell *et al.*, 2007).

First, SNPs located on non-autosomal chromosomes (i.e., sex-determining chromosomes and mtDNA) were removed, as they can bias subsequent population genetic analysis. To ensure that only high quality SNPs being retained, several filter steps were included to remove low confidence SNPs, as described in O'Leary *et al.* (2018). SNPs were filtered for minor allele frequency (MAF≥0.05), quality score (QUAL≥20), minimum genotype read depth (minDP≥5), minimum mean read depth per locus (minmeanDP≥15) and minor allele count (MAC≥3). Next, SNPs were filtered iteratively by genotype call rate (geno) and individual missing data (imiss) to minimize missing data: step 1: geno≥50%, imiss≤90%; step 2: geno≥60%, imiss≤70%; step 3: geno≥70%, imiss≤50%. After adjusting the SNP dataset, only biallelic SNPs were retained and filtered using the info tags to ensure a high confidence SNP set by the removal of additional putative low confidence SNPs. SNPs were filtered for allele balance (AB: 0.2–0.8), quality/depth ratio (QUAL/DPB>0.2), mapping quality (MQM/MQMR 0.25–1.75) and properly paired status. A final filtering step was done based on genotype call rate and individual missing data (4: geno≥85%, imiss≤25%). Only unlinked SNPs and those in Hardy-Weinberg Equilibrium were kept for further analysis.

IDENTIFICATION OF DUPLICATED INDIVIDUALS AND POPULATION GENETIC ANALYSIS

Before conducting the population genetic analysis, the remaining individuals were checked for duplicates. Samples were derived from biopsy sampling, thus enabling the possibility of repeated sampling. To identify duplicates, the R package *sequoia* v. 2.11.2 (Huisman, 2017) was used. Duplicate individuals, as well as parent offspring pairs (PO) sampled at the same location and time, were removed to ensure the criteria of random sampling. Population genetic analysis was conducted based on the remaining Individuals and SNPs. All analyses were conducted using R version 4.2.2 (R Core Team, 2021) and plink. The population structure was examined through PCA/DAPC and ADMIXTURE. PCA was performed using plink, while DAPC utilized the R package *adegenet* v. 2.1.10 (Jombart, 2008). The program ADMIXTURE (Alexander *et al.*, 2009) was used to infer clusters and the genetic identity of each sample. The optimal K value (range: 1–12) was determined and evaluated via cross validation. All figures were visualized in R using ggplot2 v. 3.5.1 (Wickham, 2016).

SNP PANEL DESIGN AND TESTING

The most informative SNPs from the full SNP set were selected using two approaches: (i) For cluster assignment: 96 SNPs were selected using the program TRES (Kavakiotis *et al.*, 2015), based on maximized F_{ST} across clusters and subspecies, (ii) For kinship and duplicate identification: 96 SNPs with a F_{ST} of 0 and a heterozygosity close to 0.5 were selected. The selected SNP panels

were then dry lab tested for cluster assignment by PCA and ADMIXTURE and for kinship/duplicate assessment using the R package *sequoia* v. 2.11.2. Both SNP panels were designed and ordered with the D3 Assay Design tool provided by Standard BioTools and wet lab tested on the JUNO and EP1 system using the Integrated Preamp SNP Type Genotyping Kit suitable for 96 SNPs and 96 samples (Standard BioTools) following the manufacturer's protocol. SNP panels were wet lab tested on a test set of 95 individuals. The set was chosen to cover all clusters and subspecies, including ddRAD-processed samples, unanalysed samples, low quality samples (≤5.0 ng/μL) that were removed during SNP filtering, duplicates (n=5) and PO pairs (n=6). After they were run on the Fluidigm and EP1, SNP panels were analysed using the Fluidigm SNP Genotyping Analysis Software (Standard BioTools) and subsequently processed in R.

RESULTS OF SNP POPULATION GENETICS AND SNP PANEL DESIGN

After filtering, 12,131 unlinked and Hardy-Weinberg Equilibrium (HWE)-compliant SNPs and 297 individuals were retained from the initial 1,656,931 SNPs for cluster assignment. Kinship and duplicate analysis using *sequoia* identified 49 duplicates which were removed. Additionally, 11 PO pairs, sampled on the same day and location, were excluded to ensure the random sampling criteria. PCA revealed three main clusters corresponding to the ocean basins, supported by ADMIXTURE which suggested a cross-validated K of 3 (Figure 2). Based on these results, a population/cluster SNP panel was designed with 48 SNPs to maximize difference between clusters (F_{ST}: 0.31–0.43) and 48 SNPs to differentiate between subspecies (F_{ST}: 0.39–0.46). For identifying duplicates and assessing kinship status, 96 SNPs with heterozygosity in a range of 0.50–0.54 and an F_{ST} of 0 were selected to ensure that SNPs are present in all clusters.

Wet lab SNP panel validation, based on the 95 individual test set, was run successfully at the ICR, Taiji Office, with an average processing time of 4.5 hours per SNP panel. Wet lab validation revealed that from the 96 selected SNPs in the cluster panel, 90 were amplified with a

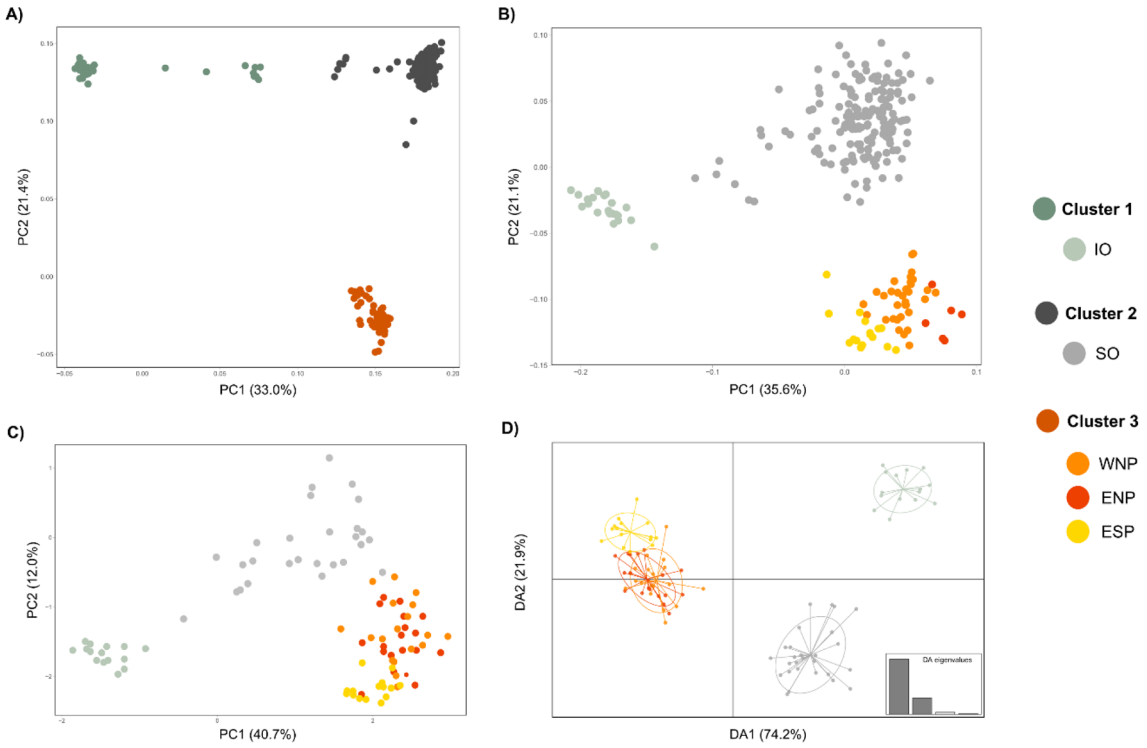

Figure 2. Clustering methods of analysed SNPs. **A)** PCA clustering based on 237 individuals and 12,131 SNPs. Colour code is assigned based on highest ADMIXTURE assignment probability to respective cluster (i.e., Cluster 1—green, Cluster 2—dark grey, Cluster 3—dark orange. **B)** PCA based on 96 most informative SNPs of the population in the SNP panel dry lab tested on 237 individuals. Colours represent subpopulations/regions. **C)** PCA based on 96 SNPs of the population in the SNP panel and the test set of 95 individuals run on the Fluidigm. **D)** DAPC conducted to increase genetic distance between clusters and subpopulations/regions. IO=Indian Ocean; SO=Southern Ocean; WNP=Western North Pacific; ENP=Eastern North Pacific; ESP=Eastern South Pacific.

call rate of ≥85% in 93 individuals. For the kinship panel 88 of the 96 selected SNPs were amplified with a call rate of ≥85% in 93 individuals. The inclusion of low quality samples (≤5.0 ng/µL) and samples which were originally excluded during SNP filtering, resulted in successful amplification in both SNP panels. Cluster assignment of the test set was consistent with previous results based on 237 individuals. Although genetic distance appeared reduced, this could be improved using a DAPC as shown in Figure 2. Kinship analysis and duplicate assignment were conducted and showed a consistency in the identification of all five duplicates which were included. On the other hand, the power to detect PO pairs decreased, as none of the six pairs were detected. Two PO pairs were misclassified as duplicates, likely due to missing data.

DISCUSSION AND FUTURE WORK

The pipeline established and introduced in this work opens a new chapter for the ICR, facilitating the transition to the era of high throughput analysis. This advancement will foster the application of the Fluidigm system in the future for population/cluster assignment, kinship analysis and duplicate detection. The aim of this study was to design a pipeline in a way that it can be readily modified depending on the cetacean species of interest. The established pipeline includes (i) the identification of SNPs from ddRAD data but also is adjustable for GRAS-DI data, including SNP filtering which can be adjusted according to data availability and quality, (ii) population genetics and kinship/duplicate analysis based on SNPs, (iii) SNP

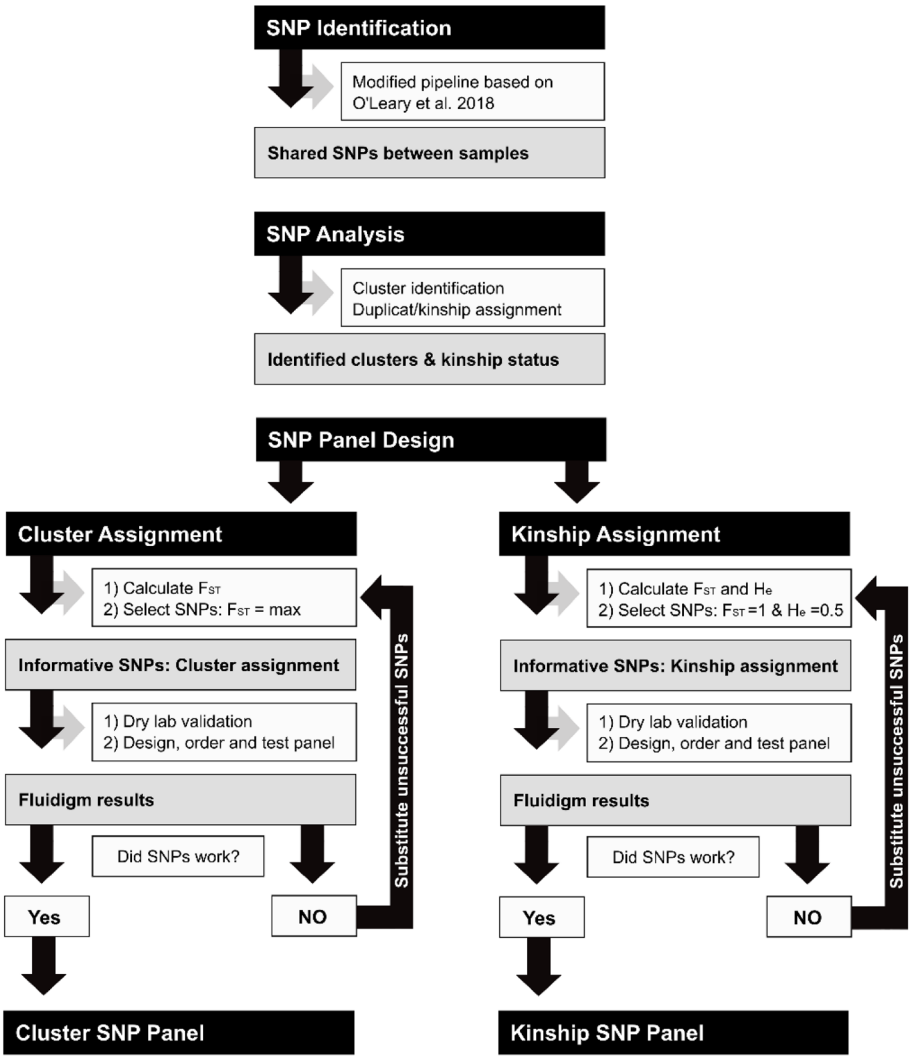

Figure 3. Developed SNP Panel pipeline at the ICR. Pipeline consists of (i) SNP identification based on ddRAD data including a modified filtering scheme described by O'Leary *et al.* (2018), (ii) subsequent SNP analysis to identify genetic clusters, kinship and duplicates and (iii) SNP panel design pipeline that includes the steps to design a SNP panel for cluster assignment, duplicate identification and kinship assessment as well as dry and wet lab testing procedures.

selection and SNP panel design (i.e., F_{ST} outlier, F_{ST} max or heterozygosity) and the establishment of SNP panel wet lab testing at the ICR's Taiji Office (Figure 3).

The results based on the test set of 95 blue whale samples have shown that the designed SNP panels work for low quality samples, which were originally excluded in the SNP filtering process. This makes the SNP panels particularly valuable for stranding samples, which often have a higher level of DNA degradation (Autenrieth *et al.*, 2024), and for non-invasive collected samples such as feces (Thaden *et al.*, 2020; Thavornkanlapachai *et al.*, 2024). Even though most of the SNPs were amplified successfully, some currently unamplified SNPs (six SNPs in cluster SNP panel and eight in the kinship SNP panel) should be substituted in the future to enhance the information provided by each SNP panel. This is particularly important for the kinship SNP panel, as it may enhance the power of PO pair assignments, which was reduced compared to the full 12,131 SNP set, likely due to missing data. A more stringent application of loci exclusion in the analysis may lead to more precise predictions of PO pairs.

While this study has demonstrated that SNPs can be used to assign specimens to clusters and to assess duplicates and potentially kinship status, the potential use of SNP panels are not limited to these applications. Other studies have successfully used SNP panels to assign species (Ciezarek *et al.*, 2022), hybrid status (Thaden *et al.*, 2020; Jarausch *et al.*, 2023), and even sex (Talenti *et al.*, 2018) to screened individuals. With the newly established pipeline and the workshops held recently in Taiji on this topic, such applications can now be explored at the ICR.

ACKNOWLEDGEMENTS

We thank all participating crew members and researchers who contributed to the collection of the samples, as well as the technicians for their support in the laboratory work at the ICR. In addition, we thank the editorial team of TEREP-ICR for editorial checking.

REFERENCES

Alexander, D.H., Novembre, J. and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19(9): 1655–1664.

Attard, C.R.M., Sandoval-Castillo, J., Lang, A.R., Vernazzani, B.G., Torres, L.G., Baldwin, R., Jenner, K.C.S., Gill, P.C., Burton, C.L.K., Barcelo, A., Sironi, M., Jenner, M.-N.M., Morrice, M.G., Beheregaray, L.B. and Moller, L.M. 2024. Global conservation genomics of blue whales calls into question subspecies taxonomy and refines knowledge of population structure. *Anim. Conserv.* DOI: 10.1111/acv.12935.

Autenrieth, M., Havenstein, K., De Cahsan, B., Canitz, J., Benke, H., Roos, A., Pampoulie, C., Sigurosson, G.M., Siebert, U., Olsen, M.T., Biard, V., Heide-Guojon, M.P., Ozturk, A.A., Ozturk, B., Lawson, J.W. and Tiedeman, R. 2024. Genome-wide analysis of the harbour porpoise (*Phocoena phocoena*) indicates isolation-by-distance across the North Atlantic and potential local adaptation in adjacent waters. *Conserv. Genet.* 25(2): 563–584.

Branch, T.A., Abubaker, E.M.N., Mkango, S. and Butterworth, D.S. 2007. Separating Southern Blue Whale Subspecies based on length frequencies of sexually mature females. *Mar. Mam. Sci.* 23(4): 803–833.

Celemín, E., Autenrieth, M., Roos, A., Pawliczka, I., Quintela, M., Lindstrøm, U., Benke, H., Siebert, U., Lockyer, C., Berggren, P., Ozturk, A.A., Ozturk, B., Lesage, V. and Tiedemann, R. 2023. Evolutionary history and seascape genomics of Harbour porpoises (*Phocoena phocoena*) across environmental gradients in the North Atlantic and adjacent waters. *Mol. Ecol. Resour.* 00: 1–18. DOI: 10.1111/1755-0998.13860.

Chen, S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *IMeta* 2(2): e107. DOI: 10.1002/imt2.107.

Ciezarek, A., Ford, A.G.P., Etherington, G.J., Kasozi, N., Malinsky, M., Mehta, T.K., Penso-Dolfin, L., Ngatunga, B.P., Shechonge, A., Tamatamah, R., Haerty, W., Di Palma, F., Genner, M.J. and Turner G.F. 2022. Whole genome resequencing data enables a targeted SNP panel for conservation and aquaculture of Oreochromis cichlid fishes. *Aquaculture* 548: 737637. DOI: 10.1016/j.aquaculture.2021.737637.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15): 2156–2158.

Ellis, J.S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coughlan, J., Cross, T.F., Crozier, W., Dillane, E., Ensing, D., Garcia de Leaniz, C., Garcia-Vazquez, E., Griffiths, A.M., Hindar, K., Hjorleifsdottir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup, D., Nielsen, E.E., Olafsson, K., Primmer, C.R., Prodohl, P., Stradmeyer, L., Vaha, J.-P., Verspoor, E., Wennevik, V. and Stevens, J.R. 2011. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica* 139(3): 353–367.

Fabbri, E., Caniglia, R., Mucci, N., Thomsen, H. P., Krag, K., Pertoldi, C., Loeschcke, V. and Randi, E. 2012. Comparison of single nucleotide polymorphisms and microsatellites in non-invasive genetic monitoring of a wolf population. *Arch. Boil. Sci.* 64(1): 321–335.

Garrison, E. and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint*

arXiv:1207.3907.

Hemmer-Hansen, J., Therkildsen, N.O. and Pujolar, J.M. 2014. Population genomics of marine fishes: Next-generation prospects and challenges. *Biol. Bull.* 227(2): 117–132.

Hohenlohe, P.A., Funk, W.C. and Rajora, O.P. 2021. Population genomics for wildlife conservation and management. *Mol. Ecol.* 30(1): 62–82.

Holman, L.E., La Garcia de serrana, D., Onoufriou, A., Hillestad, B. and Johnston, I.A. 2017. A workflow 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.

Huisman, J. 2017. Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.* 17(5): 1009–1024.

Jarausch, A., Thaden, A., Sin, T., Corradini, A., Pop, M. I., Chiriac, S., Gazzola, A. and Nowak, C. 2023. Assessment of genetic diversity, population structure and wolf-dog hybridisation in the Eastern Romanian Carpathian wolf population. *Sci. Rep.* 13: 22574. DOI: 10.1038/s41598-023-48741-x.

Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11): 1403–1405.

Kraus, R. H., Vonholdt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Forster, D.W., Fickel, J., Roos, C. and Nowak, C. 2015. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Mol. Ecol. Resour.* 15(2): 295–305.

Lah, L., Trense, D., Benke, H., Berggren, P., Gunnlaugsson, Þ., Lockyer, C., Öztürk, A., Öztürk, B., Pawliczka, I., Roos, A., Siebert, U., Skóra, K., Víkingsson, G. and Tiedemann, R. 2016. Spatially Explicit Analysis of Genome-Wide SNPs Detects Subtle Population Structure in a Mobile Marine Mammal, the Harbor Porpoise. *PloS One* 11(10): e0162792. DOI: 10.1371/journal.pone.0162792.

LeDuc, R.G., Dizon, A.E., Goto, M., Pastene, L.A., Kato, H., Nishiwaki, S. LeDuc, C.A. and Brownell, R.L. 2007. Patterns of genetic variation in Southern Hemisphere blue whales and the use of assignment test to detect mixing on the feeding grounds. *J. Cetacean Res. Manage*. 9(1): 73–80.

LeDuc, R.G., Archer, F.I., Lang, A.R., Martien, K.K., Hancock-Hanser, B., Torres-Florez, J.P., Hucke-Gaete, R., Rosenbaum, H.C., van Waerebeek, K., Brownell, R.L. Jr. and Taylor, B.L. 2017. Genetic variation in blue whales in the eastern pacific: Implication for taxonomy and use of common wintering grounds. *Mol. Ecol.* 26(3): 740–751.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21): 2987–2993.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv: 1303.3997*.

McDonald, M.A., Mesnick, S.L. and Hildebrand, J.A. 2006. Biogeographic characterization of blue whale song worldwide: Using song to identify populations. *J. Cetacean Res. Manage*. 8(1): 55–65.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M. and Portnoy, D.S. 2018. These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27: 3193–3206.

Pastene, L.A., Acevedo, J. and Branch, T.A. 2020. Morphometric analysis of Chilean blue whales and implications for their taxonomy. *Mar. Mam. Sci.* 36(1): 116–135.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* 81(3): 559–575.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28(21): 4737–4754.

Talenti, A., Palhière, I., Tortereau, F., Pagnacco, G., Stella, A., Nicolazzi, E.L., Crepaldi, P., Tosser-Klopp, G. and Consortium, A. 2018. Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. *Genet. Sel. Evol.* 50: 55. DOI: 10.1186/s12711-018-0423-9.

Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T.E., Alves, P.C., Lyons, L.A., Mattucci, F., Randi, E., Cragnolini, M., Galián, J., Hegyeli, Z., Kitchener, A.C., Lambinet, C., Lucas, J.M., Mölich, T., Ramos, L., Schockert, V. and Cocchiararo, B. 2020. Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Mol. Ecol. Resour.* 20(3): 662–680.

Thavornkanlapachai, R., Armstrong, K.N., Knuckey, C., Huntley, B., Hanrahan, N. and Ottewell, K. 2024. Species-specific SNP arrays for non-invasive genetic monitoring of a vulnerable bat. *Sci. Rep.* 14(1): 1847.

Torres-Florez, J.P., Hucke-Gaete, R., LeDuc, R., Lang, A., Taylor, B., Pimper, L.E., Bedrinana-Romano, L., Rosenbaum, H.C. and Figueroa, C.C. 2014. Blue whale population structure along the eastern South Pacific Ocean: Evidence of more than one population. *Mol. Ecol.* 23(24): 5998–6010.

Wickham, H. and Wickham, H. 2016. Getting Started with ggplot2. *ggplot2: Elegant Graphics for Data Analysis* 11–31.