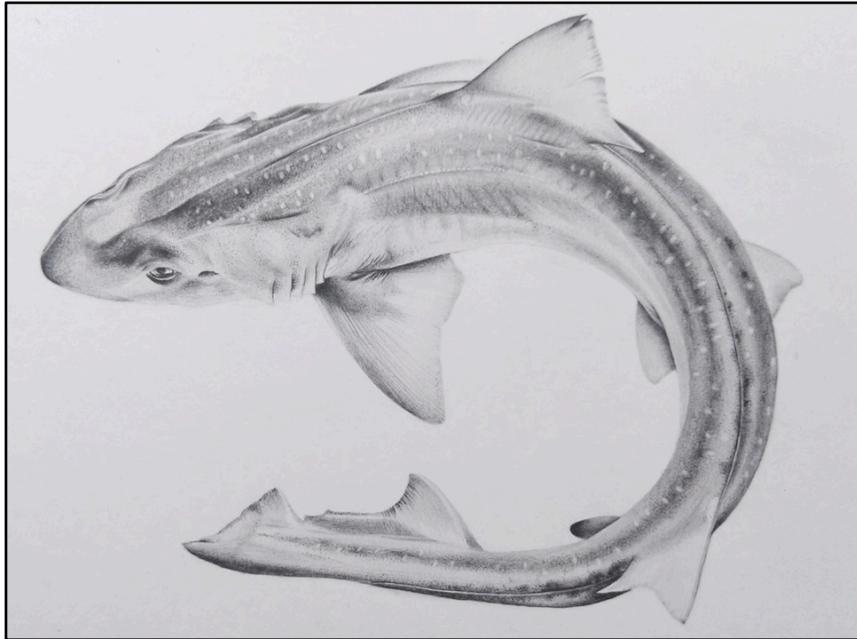


Stock structure and effective population size of the commercially exploited Gummy shark.



Artist: Barbara Langille

Emma Petrolo

Supervisors:

Associate Professor Adam Stow, Dr Kerstin Bilgmann, Dr Jessica Boomer & Jessica O'Hare

Department of Biological Sciences
Faculty of Science and Engineering
Macquarie University

This thesis is prepared for the degree of Master of Research

18th December 2019

This thesis is written in the form of a manuscript for submission to *Marine Ecology Progress Series*, with the following exceptions: cover letter and continuous line numbers are absent, a table of contents has been included, text is 1.5 spaced, tables and figures are integrated into the text and main sections have been numbered. The introduction, methods, results, discussion and appendices are extended.

Declaration:

I wish to acknowledge the following assistance in the research detailed in this report:

Jessica Boomer and Adam Stow for all sampling. Adam Stow and Kerstin Bilgmann for funding to cover DNA extraction and sequencing costs. Diversity Arrays Pty Ltd for DNA extractions and next generation genotyping. Adam Stow, Kerstin Bilgmann and Jessica O'Hare for assistance with experimental design and analyses. Adam Stow, Kerstin Bilgmann, Jessica O'Hare and Natalie Caulfield for comments on drafts of this manuscript.

All other research described in this report is my own original work.

18th December 2019



MACQUARIE
University

Acknowledgments

There are endless amounts of people I could thank for their help during this nine-month journey, however, this group deserves a special mention.

First and foremost, my amazing squad of supervisors. To Jessica Boomer; for whom this project would not exist without, literally. Thank you for your Gummy shark expertise! To Jessica O'Hare; my friend turned supervisor. Your genuine passion for my project continuously kept me motivated - it was an honest pleasure being your first student. To Kerstin Bilgmann; my genetics teacher when I began without a clue. You always approached my project and your teaching with such a gentle touch, I always felt safe in your hands. Extra thanks to beautiful baby Jai for lending me your mum for all those months. To Adam Stow; thank you for accepting me with such open arms. You allowed me to take control when I needed but also never turned me away regardless of my pestering. You are probably the most patient man alive and I am so grateful for that - it was an honour to be a part of your lab.

To my incredible lab friends – Natalie Caulfield, Sonu Yadav, Tess Nelson, Will Ashley, Teagan Parker Kielniacz and Julia Verba. I cannot express how much I appreciate all of the time you each dedicated helping me with analyses, revising my work and most importantly, bringing me snacks from Macquarie Centre.

My best friend Victoria. For taking on this year as if it was your own as you do with all aspects of my life – I will forever appreciate what you do for me.

Ma Famiglia. Thank you for encouraging me through this year, always being supportive of whatever makes me happy and for always pretending to listen when I talk about sharks. Marcus, thank you for actually listening even though you have no idea what I'm saying.

To all the in-betweens too. The big and the small. Thanks to the Starbucks down the road for my daily (and very pretentious) “venti vanilla sweet cream cold brews” and the “ChilledCow” channel on YouTube for mellowing my brain with your sick beats. They all play a vital role in the coming together of this body of work.

Table of Contents

Abstract	1
1. Introduction	2
2. Materials and Methods	7
2.1 Sampling	7
2.2 DNA extraction and sequencing	8
2.3 Generation of SNP dataset	9
2.3.1 Trimming sequences	9
2.3.2 Establishing parameters and filtering.....	9
2.3.3 Identifying loci under selection	10
2.4 Describing genetic differentiation and diversity.....	11
2.5 Determining relatedness by spatial structure	12
2.6 Estimating effective population size (N_e) and the potential for bottleneck	13
3. Results	15
3.1 Sample and SNP dataset selection	15
3.2 Genetic differentiation and diversity	16
3.3 Spatial structure of genetic variation	21
3.4 Estimates of effective population size (N_e).....	24
4. Discussion	28
Conclusions.....	31
References	33
Supplementary Material	39
Appendix 1.1: Sample list.....	39
Appendix 1.2: ipyrad parameters input file	40
Appendix 1.3: Stairway plot blueprint file – East-cluster.....	42
Appendix 1.4: Stairway plot blueprint file – South-cluster	43
Appendix 2.1: Loci identified under positive selection	44
Appendix 2.2: PCA - loci under positive selection.....	48
Appendix 2.3: Mantel test – results	49

Abstract The Australian Gummy shark (*Mustelus antarcticus*) is the main target species of a large fishery that functions across its distribution in Southern and Eastern Australian waters. Commercial harvest of the species is currently considered sustainable based on target biomass estimates that show recovery from past overexploitation. However, previous research regarding stock structure have garnered conflicting results on the genetic structure of the species, necessitating further investigation to better inform fisheries management. Here, we evaluate the genetic structure and effective population size (N_e) of the Gummy shark using genome-wide single nucleotide polymorphisms (SNPs). We identified two distinct genetic clusters, one on the East coast, and one along the South coast. Moderate genetic differentiation was identified between each cluster, while within-cluster comparisons were largely admixed. Spatial analyses revealed some evidence for natal philopatry but no compelling evidence of isolation by distance. In addition, demographic modelling of each cluster showed a comparatively rapid decline of estimated N_e in the most recent past when compared to more historical projections, although current estimates are still considered high. These findings elucidate the current genetic structure of Gummy sharks and estimate the potential impact on N_e that overfishing can generate for the species.

Key words: Gummy shark, genetic structure, genetic diversity, fisheries management

1. Introduction

Human exploitation has increasingly led to the global degradation of marine environments and a loss of biodiversity, with 24% of chondrichthyans (sharks and rays) currently in danger of extinction due to overexploitation (Dulvy et al. 2014; Ortuño Crespo & Dunn 2017; Rubinas 2017). Commercial fisheries and the globalised trade of marine resources are considered the primary cause of this, resulting in a reduction of recruitment and subsequent fishery production in exploited shark populations (Vasconcelos et al. 2013). Reductions in population size can result in a loss of genetic variation, decreasing their evolutionary potential and ultimately increasing their sensitivity to environmental change (Frankham et al. 2002; Frankham 2005; Banks et al. 2013). Fundamental to sustainable management and minimising losses of genetic variation is knowledge of genetic connectivity and stock structure of the target species.

The genetic structure of harvested species is often unknown or inferred from the spatial structuring of changes in catch biomass, resulting in poorly estimated stock assessments (Cadrin et al. 2013; Maunder & Piner 2014). Because target species that lack information on stock structure are typically not managed, the importance of genetic monitoring and the establishment of science-based management plans has been highlighted (Stow et al. 2006; Hilborn & Ovando 2014). As such, utilising genetic analyses and quantifying the genetic variation of commercially targeted species is an important component of their conservation (Frankham 2010). Alongside quantifying the impact of fishing on the genetic health of target species, population genetic approaches provide several advantages in fisheries assessment. Traditional fisheries management techniques have proven capable of rebuilding depleted stocks, however, these approaches can cost up to ~15% of the value of the stock in question as they require practical methods of observation (such as visual surveys) which are often costly and time-consuming to perform (Hilborn & Ovando 2014; Kelly et al. 2014; Port et al. 2016; Simpfendorfer & Dulvy 2017). Furthermore, population genetic approaches are capable of identifying spatial patterns of exchange and the drivers of these exchanges, such as the presence of biogeographic boundaries and their impact on gene flow (Lowe & Allendorf 2010; Selkoe et al. 2016). Genetic methods have the capacity to not only clarify that the delineation of stocks coincides with accurate population boundaries, but are capable of elucidating the demographic and adaptive history of these exploited populations, allowing for fisheries-specific impacts to be quantified (Benestan 2019).

Monitoring fisheries and informing their management using genetic data is helping elucidate the effects of the global catch increase of the 1960's-1970's (Swartz et al. 2010; Dulvy et al. 2014; Rubinas 2017). During this period, the targeting of sharks as a commercial resource increased globally by 227% (from first recordings in 1950) based on biomass catch estimates, leading to a peak in 2003 and a subsequent decline of 15% by 2011 (Collie et al. 2016; Davidson et al. 2016). Relentless harvest of *Galeorhinus galeus* (the School shark) by the Australian Southern Shark Fishery (SSF) from the 1920's-1970's led to a drastic decline in their landings, and a subsequent switch in target species from the School shark to the more biologically fecund Gummy shark (*Mustelus antarcticus*; Walker & Shotton 1999; Pribac 2005; Walker et al. 2005). Gummy sharks have been recorded to live for up to 16 years while growing to a maximum size of 1.75 meters in length (Pribac et al. 2005). Small-bodied, fast-growing shark species such as these are thought to have a higher "rebound" potential from the impacts of long-term commercial fishing pressure as maximum body size is the main predictor of threat status in sharks (Smith et al. 1998; Dulvy et al. 2014). This is likely the result of their quicker maturation rates and high fecundity (Gummy shark litter sizes range from 14-50 pups) when compared to large-bodied sharks (Pribac et al. 2005; Woodhams 2018). This is reflected by the dominance of small-bodied species being listed as "least concern" on the IUCN red list (including houndsharks, the family of which the Gummy shark belongs to; Dulvy et al. 2014).

Although populations of small-bodied sharks are generally large, multiple studies have shown that commercial harvest has the potential to reduce effective population size (N_e) to the point where genetic variation is lost even in populations of relatively large census sizes (Antao et al. 2011; Allendorf et al. 2014). A decrease in N_e (irrespective of census size) can lead to genetic drift, subsequently eroding genetic variation (Hare et al. 2011). This can be particularly problematic for species with high genetic structuring because limited gene flow among populations in concert with low N_e can result in populations with lower genetic variation (Stow et al. 2001; Frankham et al. 2002; Cadrin et al. 2013). Understanding the interplay of N_e , genetic variation, genetic differentiation and the level of demographic connectivity required to sustain healthy populations is fundamental for the long-term sustainable management of target species. However, the historical impacts of population loss on the genetic diversity and N_e of the Gummy shark is unknown (Boomer & Stow 2010).

A rapid decline in Gummy shark numbers occurred between 1970-97, where it comprised 47% of SSF catch (Walker & Shotton 1999). This resulted in a one-third reduction in their abundance (based on biomass catch estimates) along the South coast, recorded between 1973-76 and 1998-2001 (Walker & Shotton 1999). Management of the since-formed “Southern and Eastern scalefish and shark fishery” (SESSF) is hindered by uncertainties in the stock structure of Gummy sharks (Woodhams 2018; Helidoniotis 2019). With no confirmed breeding aggregations and exposure to continuous and unsystematic fishing pressure, the potential vulnerability of the Gummy shark is based on the unknown nature of this on their long-term genetic health. Gummy sharks are endemic to the coastal waters of Southern Australia ranging from Geraldton (Western Australia), around Tasmania and up to Port Stephens (New South Wales) on the East coast (Walker & Shotton 1999). However, fishing pressure from the SESSF is not homogenous across this range. Fishing pressure on the East coast is considered low, with less than 50 tonnes a year taken on average (Fig.1; Woodhams 2018). In comparison, the most recent estimates of shark catch for the south coast (2018) equalled 1,744 tonnes (Fig. 1; Woodhams 2018).

The stock structure of Gummy sharks is used to help delineate the level of fishing pressure across the species range and is currently based on a mixture of low resolution genetic and tagging studies (Woodhams 2018). Whilst the assumed stock on the Southern coast is considered sustainable, the Eastern stock is currently listed as undefined due to the lack of evidence provided from catch estimates (Fig.1; Woodhams 2018). However, the geographic extent of stocks and levels of admixture between population units is often overlooked when stock structure is largely based on tagging studies (Collie et al. 2016; Selkoe et al. 2016; Marandel et al. 2017). Utilising the dispersal potential of target species as a basis for fishing pressure disregards the realised migration and subsequent gene flow occurring between localities (Dawson 2014). While there may be few obvious barriers to dispersal across a species’ range, comparing assessments of stock structure based on catch estimates versus genetic analyses may be different because genetic sampling reveals not only current levels of genetic diversity across a species’ range, but can elucidate the demographic history of a species as influenced by environmental change (Taillebois et al. 2013; Wright et al. 2015). As the seascape heterogeneity of target habitat is predominantly hidden from our view and thus our understanding, using genetic concepts to define discrete fish stocks and the observed rates of interchange between populations is a crucial step to defining suitable management units (Hawkins et al. 2016).

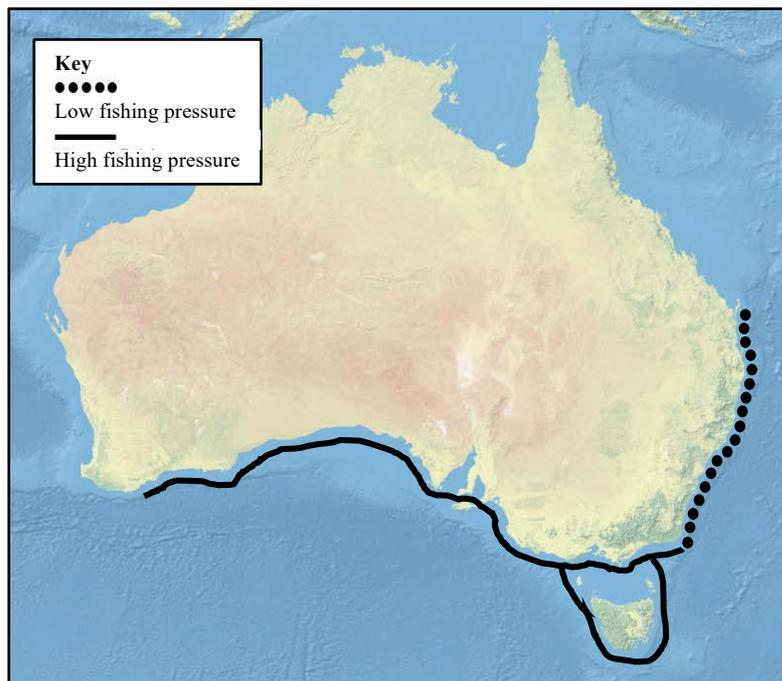


Figure 1. The range extent of Gummy shark fishing from the SESSF. Range corresponds to the degree of fishing pressure experienced by the Gummy shark from the SESSF relative to geographic location.

Although the importance of delineating stock structure and quantifying the genetic health of the commercially important Gummy shark is well understood, ambiguity on the current genetic structure of the species still remains. This uncertainty has resulted from conflicting evidence garnered from both the low reliability of catch estimates (Ovenden et al. 2018), alongside genetic studies that have investigated the population genetics of the species. MacDonald (1988) utilised a single polymorphic allozyme locus as a marker for genetic variation throughout the Southern coast of the species range, finding no evidence for genetic structuring within this region. Gardner and Ward (1998) expanded on this using seven polymorphic allozyme loci to investigate potential genetic differentiation across the East and South coast, with results denoting the presence of two separate stocks. Samples indicated one stock ranging from Newcastle to the Clarence River in NSW, and one ranging from Bunbury in WA to Eden in lower NSW (Gardner & Ward 1998). Contrastingly, there was no evidence of genetic heterogeneity across the species range using eight microsatellite markers specific to the *Mustelus* genus (Boomer & Stow 2010; Boomer 2013).

In recent years, fishery management has benefited from new methods that efficiently generate data sets of many thousands of single nucleotide polymorphisms (SNPs) per individual (Benestan 2019). The increased resolution of this method has the potential to detect changes in allele frequency that have occurred in response to historical factors alongside those that have occurred over few generations (Benestan 2019). As such, SNPs were utilised in this study with the aim of clarifying the results of previous genetic research conducted on the commercially important Gummy shark. The current genetic structure, effective population size (N_e) and the historical impacts of fishing pressure were investigated, in part to ascertain whether the existing genetic variation of the species will be eroded by current harvesting levels. In order to do so, we asked the following questions -

1. Is there genetic partitioning across the distribution of *M. antarcticus*?
2. What is the effective population size of *M. antarcticus*?
3. Is the effective population size sufficient to offset the effects of drift?
4. Has a genetic bottleneck occurred?

2. Materials and Methods

2.1 Sampling

All gummy shark tissue samples were sourced from commercial fishers operating in southern and eastern Australian waters (including Tasmania) from 2007 – 2010 by Boomer (2013). A ~1 cm³ piece of muscle or fin tissue was collected from each sampled shark and stored in 70-90% ETOH. The size and sex of each sampled shark was recorded with all sharks sexed according to their external morphology. Age classes ranged from sub-adults to adults. Sample locations ($n=7$) ranged from northern New South Wales to southern Western Australia (Fig. 2; Tab. 1). Samples from seven locations were included to assure good spatial coverage while maintaining sufficient statistical power at each location for subsequent analyses.

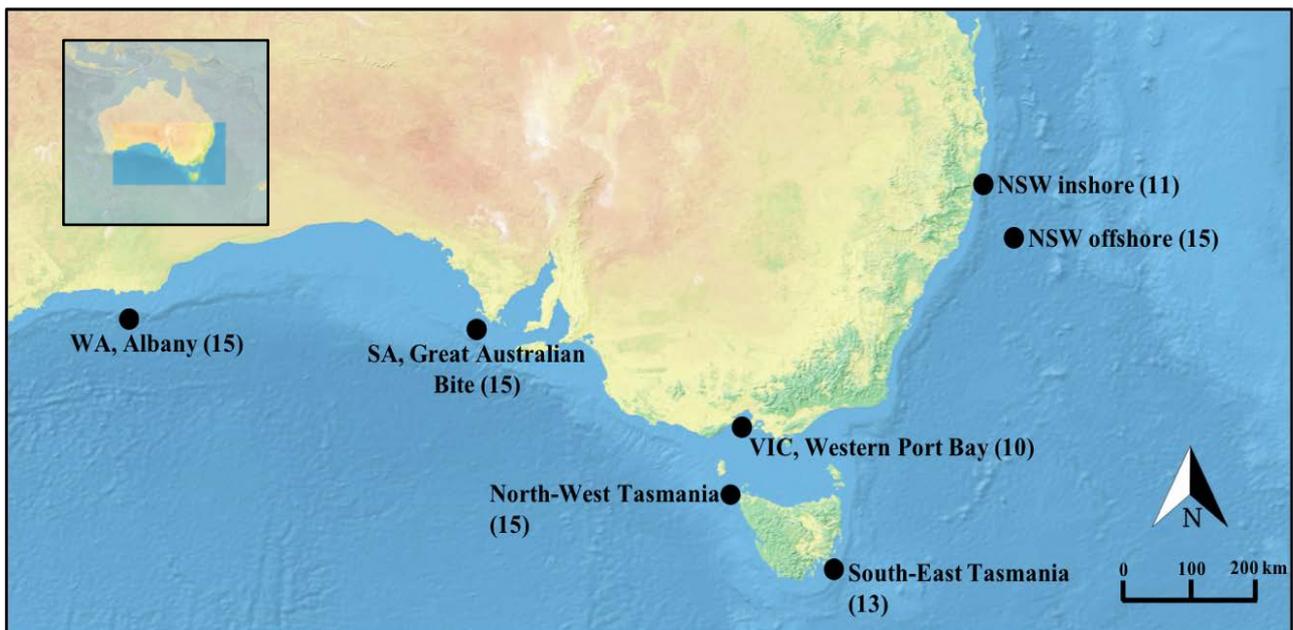


Figure 2. Map of sites where gummy sharks (*Mustelus antarcticus*) were sampled within Southern and Eastern Australian waters ($n=7$). Sample numbers for each site are displayed in brackets.

Table 1. Sample sites corresponding with their site name abbreviation (code) and an average latitude and longitude of their sampling location.

Site Name	Code	Coordinates
NSW inshore	NSW-IN	31°35'540"S, 153°00.555"E
NSW offshore	NSW-OFF	33°18'55.8"S, 152°13'09.1"E
VIC, Western Port Bay	VIC	38°21'22.7"S, 145°14'52.8"E
North-West Tasmania	TAS-NW	40°48'54.2"S, 144°31'11.1"E
South-East Tasmania	TAS-SE	42°58'12.0"S, 147°46'48.0"E
SA, Great Australian Bite	GAB	33°03'58.7"S, 130°52'12.2"E
WA, Albany	WA	34°54'52.6"S, 119°02'19.3"E

Sub-sampling of 94 tissue samples was conducted. Similar numbers of tissue samples were chosen for each sampling location (Appendix 1.1) aiming for an even sex-ratio for each location where possible. Each tissue sample was sub-sampled into a ~10 µg piece appropriate for DNA extraction and placed into an Eppendorf fully skirted, 96 well PCR plate. 70% ETOH was added to each well and each plate row was sealed with a Sarstedt clear flat strip.

2.2 DNA extraction and sequencing

Samples were sent to Diversity Arrays Technology Pty. Ltd. (DArT; Canberra, Australia - <http://www.diversityarrays.com>) for DNA extraction and sequencing. DNA was extracted using the GeneCatch™ Blood and Tissue Genomic Mini Prep Kit (Epoch Life Science, Inc) following the manufacturer guidelines. SNP discovery was performed for each sub-sample using the standard DArTseq protocol (DArT 2018). DArTseq is a genotype-by-sequencing method (Sansaloni et al. 2011; Kilian et al. 2012) able to perform genome-wide marker discovery using the Illumina next-generation sequencing (NGS) platform (Andrews et al. 2016; Zhang et al. 2018). The DArTseq protocol is explained in brief below.

To ensure the quality of all genomic DNA, template DNA was incubated in a 1X solution of Multi-Core™ restriction enzyme buffer (Promega) for 2 hours at 37°C. Approximately 100ng per µL of each DNA sample was then digested with a combination of the two restriction enzymes *PstI* and *SphI*. Each individual sample was ligated to unique barcodes and adapters specific to these enzymes. PCR amplification of each sample followed using primers specific

to the barcode and adaptor sequences used. PCR conditions consisted of 1 min initial denaturation at 94 °C, followed by 30 cycles of 20 s denaturation (94 °C), 30 s annealing (58 °C) and 45 s extension (72 °C), and a final extension of 7 min at 72 °C. To prepare for hybridization to the flow cell, approximately 10 µL of each sample were pooled, diluted and denatured using NaOH. The subsequent library was sequenced on an Illumina HiSeq®2500 platform for 77 cycles, resulting in 77 base pair (bp) long fragments (single read). 20% of the 94 samples were processed a second time following the preceding protocol to create a set of technical replicates that were used to later assess the reproducibility of SNP calls.

Raw genetic sequences were converted to fastq format using the Illumina HiSeq2500 platform. Individual samples were then demultiplexed based on their unique ligated barcode. All remaining reads were checked for contamination using GenBank viral and bacterial example sequences alongside a database curated by DArT to ensure for quality control of reads.

2.3 Generation of SNP dataset

2.3.1 Trimming sequences

Individual fastq files were unzipped and re-labelled giving them unique id's pertaining to their original sample source with all replicates identified and labelled accordingly. Within each sequence, the prefix barcode was removed using the fastx-toolkit (v0.0.14; Gordon & Hannon 2010) leaving the remaining *PstI* restriction enzyme overhang. All barcodes were of random lengths, the maximum being 8bp long. As a result, the shortest sequence was trimmed to 69bp with all remaining sequences trimmed to the same length to ensure comparability among reads.

2.3.2 Establishing parameters and filtering

Trimmed fastq files were processed *de novo* in the pipeline ipyrad (v.0.7.29; Eaton & Overcast 2016). Ipyrad provides default parameters that detail the actions performed during the assembly of consensus sequences (Valencia et al. 2018). The assembly method follows seven steps which assigns reads to individuals, filters for low quality base calls, *de novo* clusters sequences, estimates sequencing error rate and consensus allele sequences before clustering consensus reads a final time to output in variant call format (vcf) file (Fig. 3).

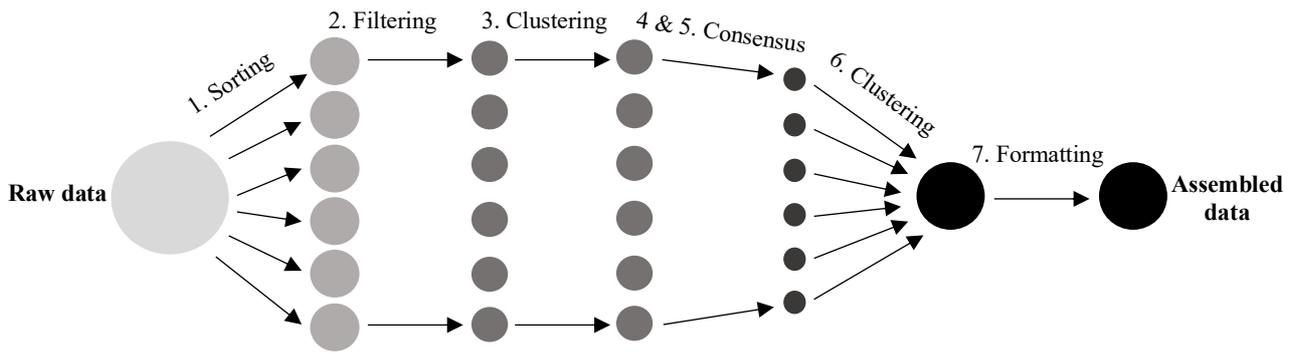


Figure 3. Schematic details the basic assembly workflow implemented by ipyrad, sorting sequences from raw to assembled data.

Where applicable, default parameters were modified to fit the specifications of the dataset with a preference for more stringent filtering options to create a high-quality SNP dataset. The complete input file of chosen parameters is provided in Appendix 1.2. In brief, VCFtools (Danecek et al. 2011) was used to filter and thin all identified SNP sites. Individuals with a read depth $<40\,000$ were removed and were replaced with a replicate sample of better quality where possible. Only polymorphic biallelic loci with $<10\%$ missing data, a minimum read depth >10 and a minor allele frequency (MAF) >0.05 were retained. The 2b-RAD pipeline Hetfilter.pl (Wang et al. 2012) was used to exclude sites with a heterozygosity >0.5 to guard against lumped paralogous loci. Additionally, the 2b-RAD pipeline repMatchStats.pl (Wang et al. 2012) was used to assess all technical replicates ($n=24$ pairs) using the calculated heterozygosity discovery rate. Individuals that met the average heterozygosity were retained, and their replicate counterparts of lower quality were excluded from the final dataset. Finally, the dataset was thinned per 69bp so that no two sites were within this specified distance of one another.

2.3.3 Identifying loci under selection

All remaining SNPs were further filtered to ensure a neutral dataset. Hardy-Weinberg equilibrium (HWE) was calculated per individual locus using default Markov chain parameters in Genepop v.4.2 (Raymond & Rousset 1995, 2004). To account for Type I errors associated with multiple tests, all obtained p-values were converted to q-values using the package *qvalue* v.2.16.0 (Dabney et al. 2004) in R studio (R core development team 2015). All loci that deviated from HWE at a false discovery rate of 5% (q-value <0.05) were

subsequently removed from the dataset. Several F_{ST} outlier tests were then implemented to identify SNPs carrying signals of selection. Arlequin v.3.0 (Excoffier et al. 2005) was run using all default parameters, applying a coalescent method which generates a null distribution of F_{ST} using the island model of population structure (Excoffier et al. 2009). P-values calculated by Arlequin were converted to q-values using the same method as previously mentioned. In addition, Bayescan v.2.1 (Foll & Gaggiotti 2008) was run using default parameters. Bayescan applies the Bayesian simulation method established by Beaumont and Balding (2004) to detect departures from neutrality between groups of loci. The subsequent output was visualised in R using a 5% false discovery rate in the *plot_bayescan* function. OutFLANK v.0.2 (Whitlock & Lotterhos 2015) was additionally run, with both the upper and lower 5% of inferred F_{ST} distribution trimmed prior to estimation of the null model. All loci identified as outliers under a 5% false discovery rate (FDR) by any of the cited methods were subsequently removed from the SNP dataset. This method ensured greater confidence in the neutrality of the remaining SNP dataset, thus the interpretation of their results.

2.4 Describing genetic differentiation and diversity

The residual dataset of selectively-neutral SNPs was used to evaluate genetic differentiation between sample sites in all subsequent analyses. The R package *Adegenet* v.2.1.1 (Jombart & Ahmed 2011) was used to execute a K-means clustering analysis to identify the presence of distinct genetic clusters. These clusters were then further explored in *Adegenet* using the discriminant analysis of principal components (DAPC) function. Genetic clusters were then visualised using a principal components analysis (PCA, retaining 100 PC's) which employs the multivariate method designed by Jombart and Ahmed (2011). In addition, least-squares estimates of ancestry proportions (Frichot et al. 2014) was calculated using the R package *LEA* (Frichot & François 2015) to create a barplot of individual admixture coefficients based on the number of previously identified clusters (K). Genetic differentiation was calculated using θ Weir and Cockerham's (Weir & Cockerham 1984) method of estimating pairwise F_{ST} using the *diveRsity* v.1.9.90 (Keenan et al. 2013) package in R. Pairwise estimates were made based on 999 bootstraps with 95% bootstrapped confidence intervals comparing each sampling location ($n=7$) and each identified genetic cluster.

Summary statistics were calculated to establish the genetic variation of each identified cluster. The R package *strataG* (Archer et al. 2017) was used to calculate the number of private alleles (N_p), mean allelic richness (A_r), the proportion of polymorphic loci (PPL) and both the mean observed and expected heterozygosity (H_o , H_e) across all loci for each identified cluster. Wrights inbreeding coefficient (F) was calculated (per locus) based on the values of H_o and H_e determined by *strataG*, using the formula –

$$F = 1 - \frac{H_o}{H_e} .$$

2.5 Spatial analyses of genetic variation

Once genetic clusters were identified within the dataset, spatial analyses of genetic variation were performed. Only clusters that encompassed >2 sampling locations were included in the following spatial analyses. Analyses were performed on all individuals comprising the qualifying southern cluster ($n=70$).

A spatial autocorrelation analysis was performed in GenAlEx v.6.5 (Smouse & Peakall 2012) to explore r between individuals on different spatial scales. Pairwise genetic (GD) and geographic distance (GGD) matrices were calculated for each analysis based on the input SNP and sample coordinate data with all individuals collected from the same sampling location comprising the same coordinates (Tab. 1). Distance classes were chosen based on the maximum calculated GGD for each analysis, with distance classes chosen at even distances within the maximum threshold. Each spatial autocorrelation was run with 1000 bootstraps per distance class, generating (upper and lower) 95% confidence intervals around each mean (r). The null hypothesis (no spatial structure, $r=0$) was tested using 999 permutations, generating (upper and lower) 95% confidence intervals for each distance class. The presence of isolation-by-distance (IBD) was examined using heterogeneity tests (Banks and Peakall 2012). In addition, this dataset was split between female ($n=41$) and male ($n=29$) individuals in order to explore the potential influence of sex on the spatial structure of genetic correlation (r) for the species.

In order to further investigate for the presence of IBD, the mixed-effects linear based maximum likelihood population effects model (MLPE) was run in R using the packages

corMLPE (Pope 2018) and *nlme* (Pinheiro et al. 2012). The *R* package *Adegenet* v.2.1.1 (Jombart & Ahmed 2011) was used to calculate Edward's genetic distance (Edwards 1971) between pairs of sites, which was then compared to the coordinate data of each site (Tab. 1) within the qualifying cluster ($n=70$). Genetic distance and geographic distance (km) were compared in MLPE on the population-level to explore the potential broad-scale spatial structuring of genetic diversity for the species. This analysis was not used to compare males and females separately due to sample size limitations at some locations.

In addition, a mantel test (Mantel 1967) was run in *R* using the package *ade4* (Dray & Dufour 2007). A distance matrix was generated, mapping the Euclidean distance between the coordinates of each sample location. This matrix was then tested against the estimated genetic distance between corresponding sample locations. Calculations of genetic distance followed the method described by Edwards (1971). Each analysis ran for 100,000 repetitions. Due to the recent conjecture surrounding the use of the mantel method to test hypotheses of IBD (Harmon & Glor 2010; Legendre et al. 2015), the results of the preceding analyses are discussed in Appendix 2.3.

Finally, the *R* package *SNPRelate* (Zheng et al. 2012) was used to calculate the within-site pairwise relatedness of all samples within the qualifying southern cluster ($n=70$).

2.6 Estimating effective population size (N_e) and the potential for bottleneck

The demographic history of *M. antarcticus* was explored using the model-flexible method *Stairway plot* which estimates changes in effective population size (N_e) based on site frequency spectra (SFS; Liu & Fu 2015). Unique "Blueprint files" were created for each cluster (Appendix 1.3 & 1.4) as per the method outlined by *Stairway plot* v.2 (Liu & Fu 2015). The blueprint files contained both default and unique parameters calculated specifically for *M. antarcticus*, with all unique values obtained using the following methods. The pipeline *easySFS* (Overcast 2017) was used to calculate the SFS of each identified genetic cluster. Site frequency spectra was calculated from the initial dataset of raw variants identified in *ipyrad* (SNPs = 60,995), which preceded filtering for polymorphic biallelic loci and MAF (see section 2.3.2). This was done as stringent filtering to exclude raw variants (particularly MAF) strongly affects calculations of SFS and can create a bias in the inference

of population size (Linck & Battey 2019). As the specific mutation rate of *M. antarcticus* is currently unknown, an estimate of mutation rate for the species was made based on the method followed by Galván-Tirado et al. (2013). In brief, a mutation rate of 0.62% per million years was used. This was calculated as the average of the known mutation rates for the Scalloped hammerhead (*Sphyrna lewini*; 0.8%; Duncan et al. 2006), the Sicklefin lemon shark (*Negaprion acutidens*; 0.67%; Schultz et al. 2008), the Blacktip reef shark (*Carcharhinus melanopterus*; 0.43; Keeney & Heist 2006) and the Nurse shark (*Ginglymostoma cirratum*; 0.57%; Karl et al. 2012). Because Stairway plots require mutation rates to be per generation, this average (0.62% per million years) was then divided by the generation time of *M. antarcticus* (16 years; Woodhams 2018) garnering a mutation rate of 0.0000001. The total number of observed nucleic sites (L) was calculated based on the total number of filtered loci (35,271) from the initial dataset of identified raw variants (SNPs = 60,995), multiplied by the number of base pairs (69; $L = 2,433,699$). All other parameters included in the blueprint files of each identified cluster can be found in Appendix 1.3 and 1.4.

3. Results

3.1 Sample and SNP dataset selection

Five individuals were identified as being of poor quality, all of which were fin tissue samples collected from the South-East Tasmania sampling location (Fig. 2; Appendix 1.1). One individual had a read depth <40 000 and the remaining four did not meet the average heterozygosity discovery rate observed for all other samples. Fortunately, each of these five samples was made a technical replicate, all of which were of acceptable quality (having passed all filtering stages). As such, these five samples were replaced with their superior replicate, leaving a dataset of 94 individuals.

Raw variants that were genotyped across all 94 individuals based solely on the initial input parameters selected for ipyrad (Appendix 1.2) contained 60,995 SNPs. Executing all chosen filtering options in ipyrad eventuated in a dataset of 9,394 SNPs (Tab. 2). The greatest loss of SNPs during this process (48,950 SNPs removed) occurred during filtering for biallelic loci and a minimum minor allele frequency (MAF) of 0.05 (Tab. 2). This is to be expected as executing a cut-off for minimum allele frequency generally excludes a large portion of genetic variants that do not meet the prescribed threshold. In this case, the chosen threshold was quite high (MAF=0.05) resulting in a loss of a large portion of SNPs (in addition to those excluded based on polymorphism).

Table 2. Workflow of filtering steps undertaken following data assembly detailing the number of SNPs removed per filtering stage.

Filtering stage	Remaining SNPs (no.)
Raw variants identified in ipyrad	60995
Biallelic loci < 0.1 missing; MAF 0.05	12045
Heterozygosity < 0.5	11995
Thinning 69 (one site per read)	9394
HWE $q > 0.05$	9257
Putatively selectively neutral loci	8881

Filtering of the raw dataset in ipyrad eventuated in a dataset of 9,394 SNPs. Following this, a total of 137 loci were determined to be out of HWE ($q < 0.05$) and were removed, leaving 9,257 SNPs (Tab. 2; Appendix 2.1). Subsequent testing of this filtered dataset (SNPs = 9,257) for F_{ST} outlier loci (FDR < 0.05) in Bayescan identified 214 loci under positive selection, whilst Arlequin determined 60 (Appendix 2.1). All outlier loci previously identified in either Bayescan or Arlequin were also recognised by testing in OutFLANK, revealing 376 F_{ST} outlier loci (Appendix 2.1). As a result, all 376 loci recognised by OutFLANK were removed as being identified as putatively under positive selection by one or more F_{ST} outlier tests (Tab. 2). Removal of these outliers resulted in a final neutral dataset of 8,881 SNPs which was used for all analyses and produced all following results.

3.2 Genetic differentiation and diversity

A K-means clustering analysis inferred the presence of two genetic clusters within the dataset ($K=2$). Further investigation through DAPC analysis supported this result (BIC 654.6453; Fig. 4a) and identified two distinct clusters (Fig. 4b) indicating a difference in the allele frequency of the individuals assigned to each cluster. Amongst the identified clusters of *M. antarcticus*, one cluster was found to encompass most individuals sampled from the East coast of Australia (hereafter East-cluster; $n=24$; Fig. 5), while the second cluster included the majority of the individuals sampled along the Southern coast including Tasmania (hereafter South-cluster; $n=70$; Fig. 5). Two individuals sampled from the NSW inshore site (Fig. 2) were assigned to the inferred South-cluster (Fig. 5a).

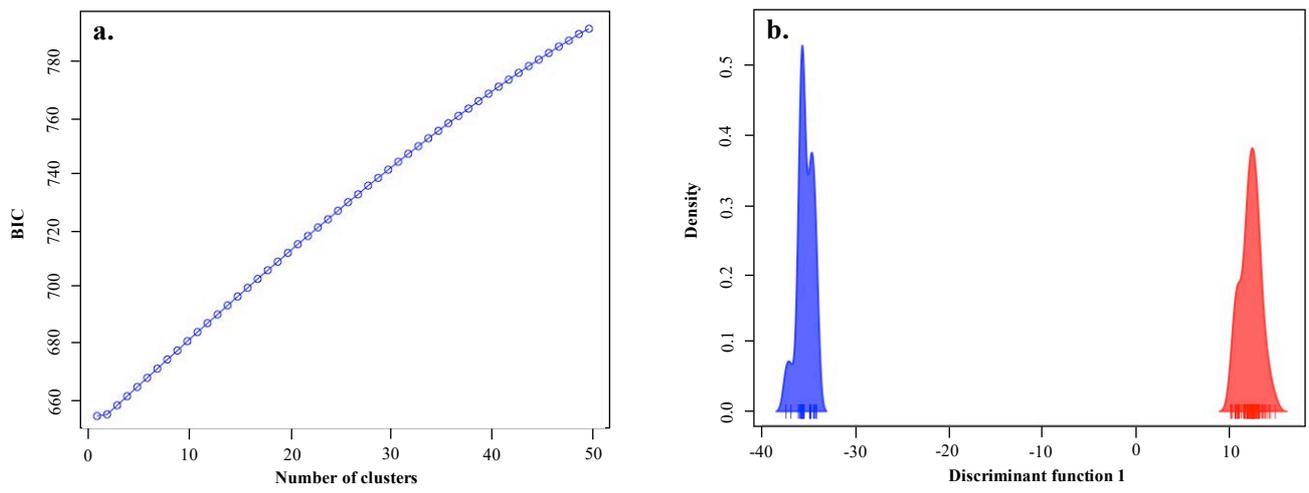


Figure 4. Identification of two genetic clusters by discriminant analysis of principal components (DAPC). a) Potential number of clusters (K) ranging from 0-50 with K indicated by BIC value. b) Distinct clusters displayed based on the density of the first discriminant function.

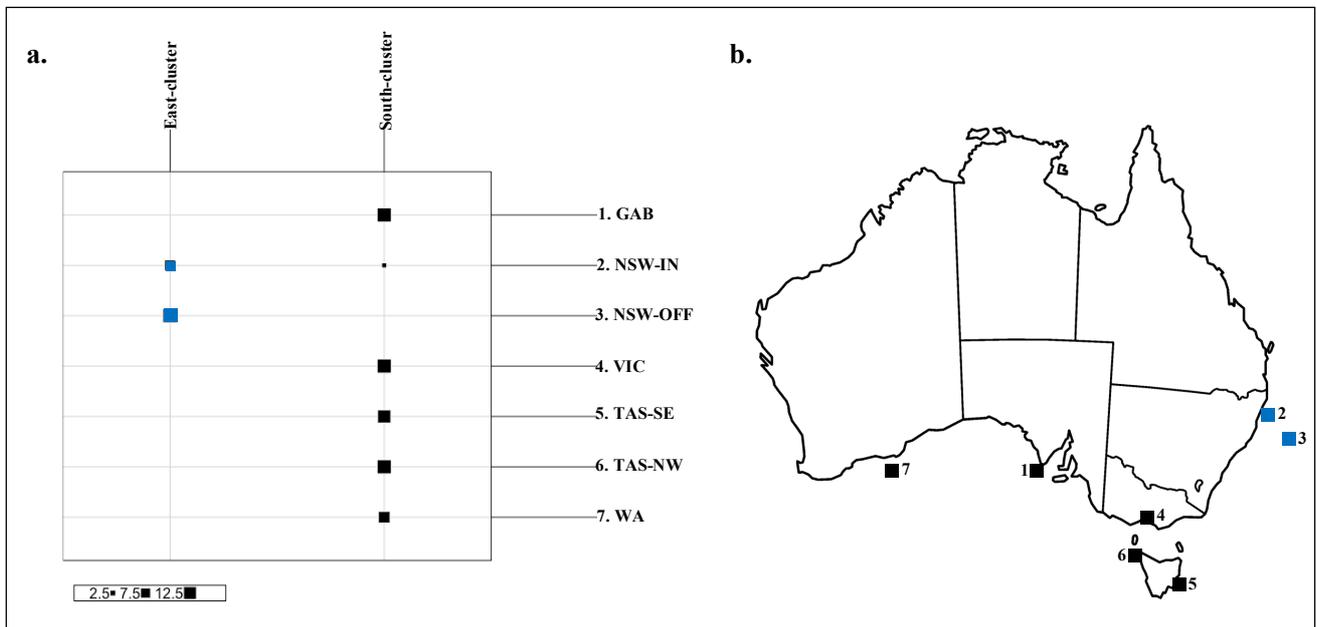


Figure 5. a) Individuals ($n=94$) allocated to each inferred cluster based on sample site. b) Numbers correspond names of sample sites to their geographical location while colours indicate the location of inferred East and South-cluster based on sample site.

The signal of spatial genetic structure was dually demonstrated in the results of the PCA analysis (Fig. 6). Two distinct clusters can be seen, which represent two groups of individuals that are genetically distant from one another. Individuals sampled along the East coast of Australia ($n=24$) formed a single cluster (Fig. 6, Axis 1) as did those sampled along the South coast ($n=70$; Fig. 6, Axis 1). Individuals within the East-cluster (EC) were more densely grouped (Fig. 6, Axis 2) than those of the South-cluster (SC) which were more broadly spread (Fig. 6, Axis 2), indicating less genetic variation within the EC when compared to the SC.

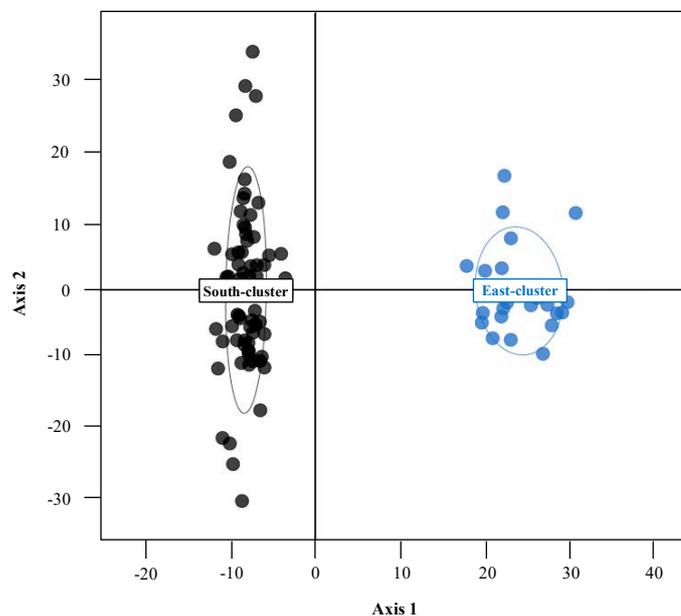


Figure 6. Principal component analysis (PCA) shows the separation of individuals ($n=94$; indicated by circle points) from 0 into two distinct groups. Ellipses indicate individuals with a normal distribution (within) and those that are outliers (outside) within each group. PCA of loci under positive selection ($n = 376$) returned comparable results (see appendix 2.2).

These two genetic clusters were further supported by the admixture analysis. Admixture proportions for each individual were minimal with a strong gradient between blocks (Fig. 7). Within the SC ($n=70$), ten individuals were observed to share no ancestral genes with the EC. Concurrently, within the EC ($n=24$) five individuals presented no admixture with the SC. The SC featured a maximum shared ancestry proportion admixed from the EC of 23%. Similarly, the maximum admixture of ancestral genes from the SC to the EC was 22%.

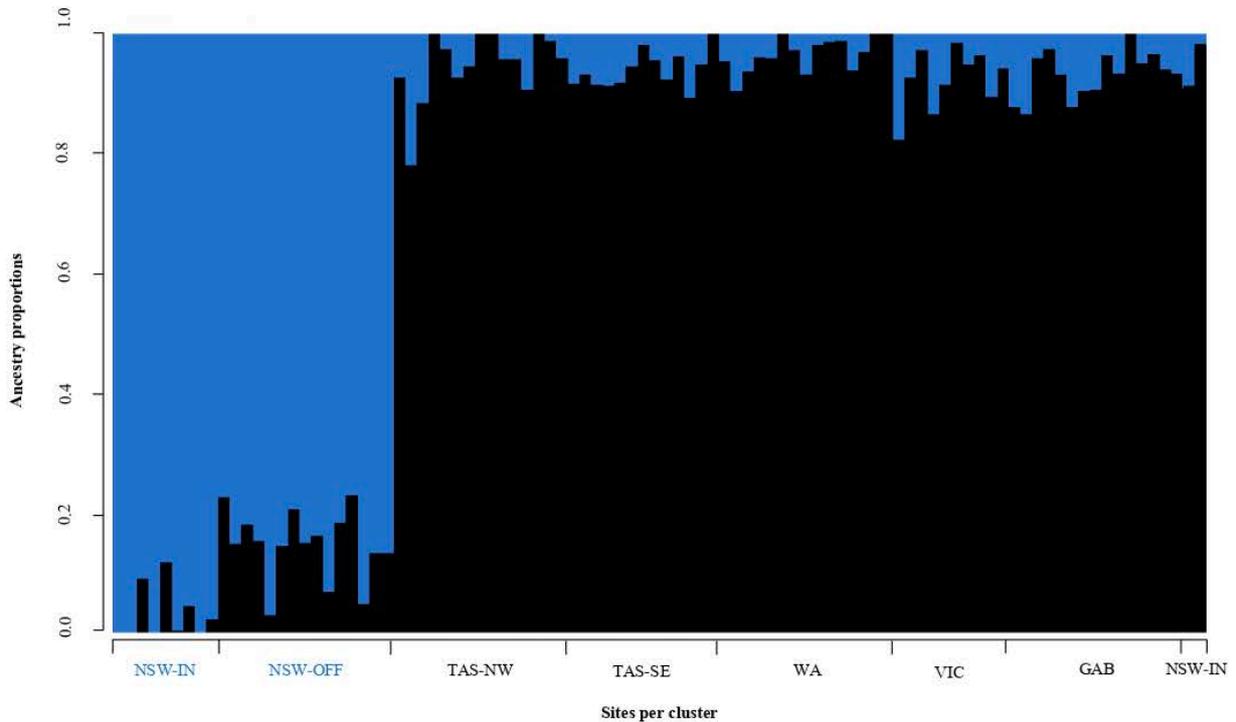


Figure 7. Admixture plot displaying ancestry proportions of gene frequencies per individual ($n=94$), with individuals assigned to clusters ($K=2$) based on shared coancestry. Each individual column denotes one individual. Site name text-colour indicates cluster assignment; East-cluster = blue, South-cluster = black.

The genetic difference between the Eastern and Southern cluster was quantified in pairwise F_{ST} calculations which indicated moderate genetic difference between the East and South-cluster with the F statistic of 0.0298 (95% CI: 0.0169 - 0.0447) being significantly different from zero. There is a clear distinction between the range of F_{ST} values quantified between each sample site. Those comparing between-cluster sites are consistently higher (Fig. 8; $F_{ST}=0.0279 - 0.0323$) than within-cluster site comparisons which are largely admixed (Fig. 8; $F_{ST}= -0.0001 - 0.007$).

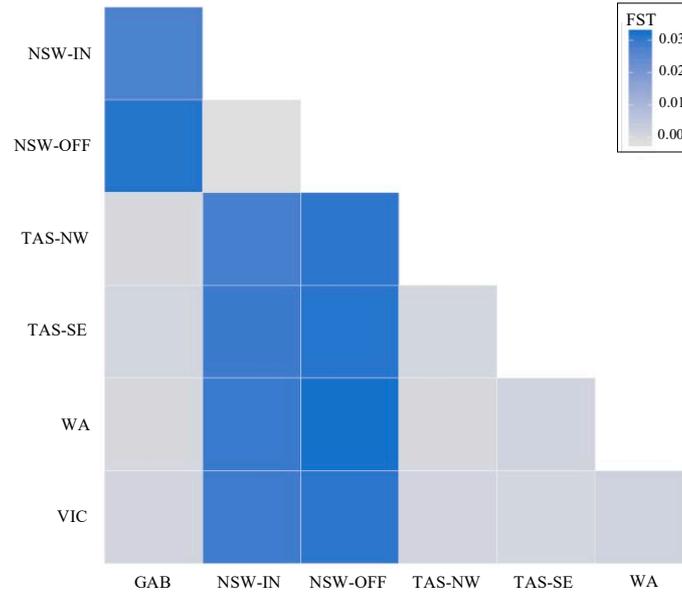


Figure 8. Heat map displaying the variation of pairwise F_{ST} values between each site ranging from -0.0001 - 0.0323. East-cluster = NSW-IN, NSW-OFF. South-cluster = TAS-NW, TAS-SE, WA, VIC.

Summary statistics comparing both the East and South-cluster are presented in Table 3. There was a greater number of private alleles (N_p) identified in the SC ($N_p = 252$) than in the EC ($N_p = 12$). Contrastingly, the mean number of alleles calculated per locus (A_r) was greater in the EC ($A_r = 0.0833$) than in the SC ($A_r = 0.0294$). The proportion of polymorphic loci (PPL) were comparably high for each cluster ranging from 0.97 – 0.99 (EC to SC respectively; Tab. 3). The observed heterozygosity (H_o) determined for each cluster per locus ranged from 0.0416 – 0.5121 in the EC and 0.0142 – 0.5040 in the SC. Concurrently, the expected heterozygosity (H_e) per locus values ranged from 0.0416 – 0.75 in the EC and 0.0142 – 0.619 in the SC. Estimates of mean H_o (EC = 0.2374, SC = 0.2357) and H_e (EC = 0.2461, SC = 0.2468) indicate similar levels of genetic variation between each identified cluster of *M. antarcticus* (Tab. 3). Moreover, calculations of F do not indicate a significant deficit of heterozygosity within each cluster (EC = 0.0351, SC = 0.0451).

Table 3. Summary statistics. Number of private alleles (N_p), mean Allelic richness (A_r), proportion of polymorphic loci (PPL), mean observed heterozygosity (H_o), mean expected heterozygosity (H_e) and Wrights inbreeding coefficient (F).

	East-cluster ($n=24$)	South-cluster ($n=70$)
N_p	12	252
A_r	0.0833	0.0294
PPL	0.97	0.99
H_o	0.2374	0.2357
H_e	0.2461	0.2468
F	0.0351	0.0451

3.3 Spatial structure of genetic variation

Analysis of all individuals in the *South-cluster ($n = 70$) identified positive spatial structure in the first distance category (within-site comparisons) and a significant decline of genotypic similarity (r) across the whole correlogram (Heterogeneity test: $\omega = 68.584$, $P < 0.01$). Genotypic similarity (r) was significantly greater than the null hypothesis at the 0 km distance class ($r = 0.014$, $P < 0.01$), with positive autocorrelation remaining until 348.419 km ($r = 0$ intercept; Fig. 9). Beyond the first distance class there was little evidence of a decline in r with geographic distance.

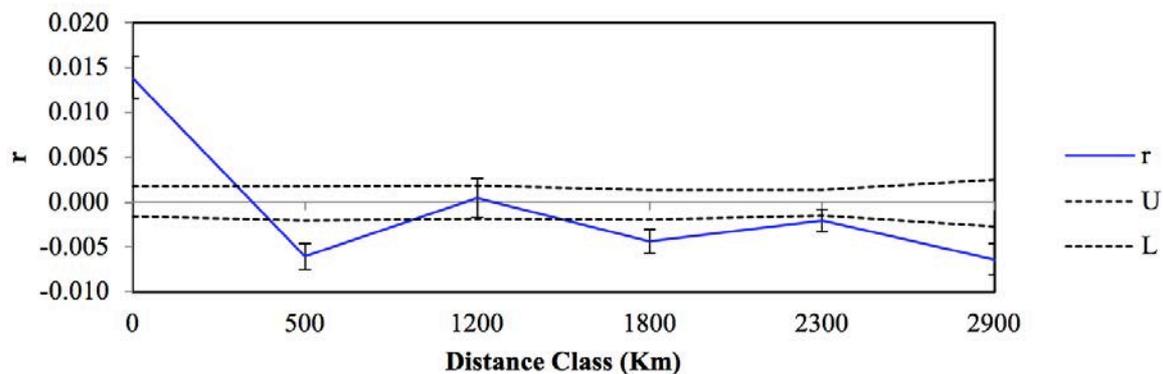


Figure 9. Spatial correlogram comparing mean genotypic similarity (r) within the South-cluster across six different classes. Solid blue line indicates calculations of r per distance class with upper and lower error bars providing 95% confidence intervals for each comparison. Distance class 0 ($n=439$), 500 ($n=345$), 1200 ($n=375$), 1800 ($n=431$), 2300 ($n=600$) and 2900 km ($n=195$).

No signal of isolation by distance (IBD) was identified by the MLPE analysis of the **SC of *M. antarcticus* (AIC = -68.118756, delta = 0, $P = >0.05$). No spatial pattern of genetic variation was estimated as there is no linear relationship between genetic and geographic distances within the SC (Fig. 10). The estimated genetic variation of individuals across the geographic range of samples is relatively homogenous, with genotypic distance decreasing by only 0.006 D_{CSE} across the maximum 2500 km distance range. Comparisons of genetic distance are split between a higher genetic distance range (0.127-0.133 D_{CSE}) and a lower range (0.105-0.116 D_{CSE}) with no comparisons found outside of these ranges. Comparison points within these upper and lower ranges are evenly spread across the sample sites contributing to the geographic distance of the test (Fig. 10).

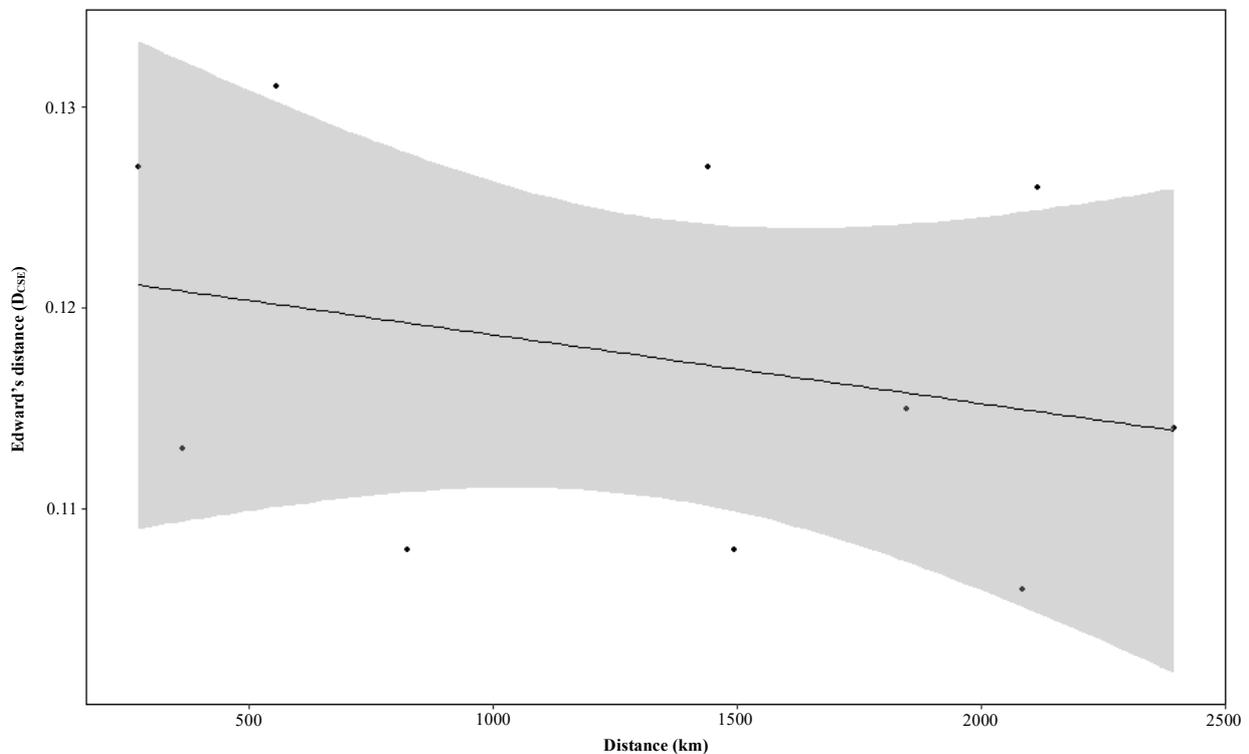


Figure 10. MLPE plot comparing the genotypic distance of all individuals in the South-cluster (excluding the NSW-IN site; $n=68$) to the range of distances (km) encompassed by their associated sample sites. Solid lines indicate points of comparison (GenoDist/Dist) based on the number of individuals tested, sheer lines denote 95% confidence intervals encompassing these comparisons.

A decline in r across the whole correlogram was identified by heterogeneity tests for both females ($n = 41$) and males ($n = 29$) in the SC (Females: $\omega = 57.208$, $P < 0.01$; Males: $\omega = 41.139$, $P < 0.01$). Whilst spatial autocorrelation analyses suggest a mildly different pattern of spatial structure between the sexes, the 95% CI's around estimates of r overlap (0 km: Females; $r = 0.012$, $P < 0.01$, Males; $r = 0.018$, $P < 0.01$; Fig. 11a & b). However, differences in the tested distance classes caused by the uneven sex ratio of samples at some sites (Tab. 4) makes calculations of genotypic similarity not directly comparable.

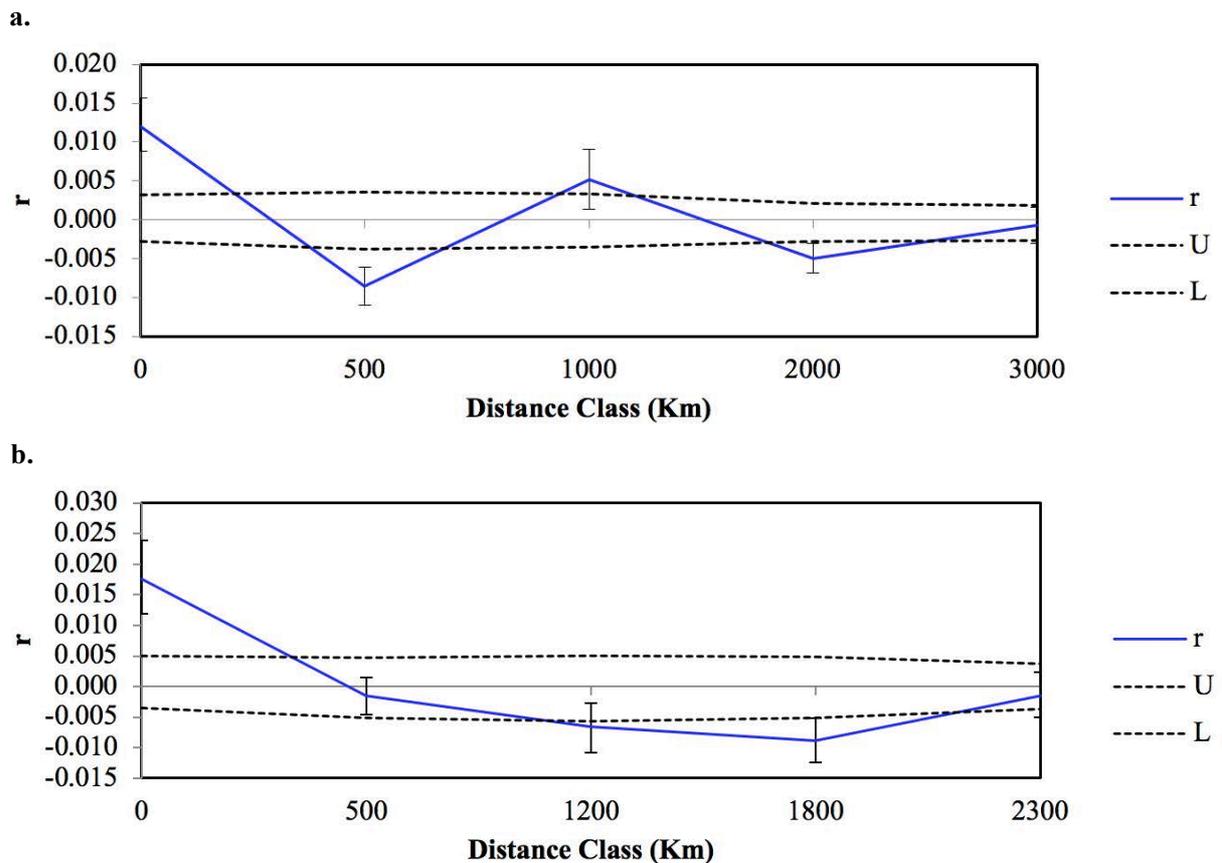


Figure 11. Spatial correlogram comparing mean genotypic similarity (r) between a) female and b) male individuals within the South-cluster. Solid blue line indicates calculations of r per distance class with upper and lower error bars detailing 95% confidence intervals for all comparisons. a) Distance class 0 ($n=439$), 500 ($n=345$), 1200 ($n=375$), 1800 ($n=431$), 2300 ($n=600$) and 2900 km ($n=195$). b) Distance class 0 ($n=136$), 500 ($n=112$), 1000 ($n=120$), 2000 ($n=220$) and 3000 km ($n=232$). b) Distance class 0 ($n=79$), 500 ($n=63$), 1200 ($n=56$), 1800 ($n=63$) and 2300 km ($n=103$).

The average kinship calculated for individuals within sample sites support findings for a lack of IBD across the SC as relatedness estimates were comparable between all sites (Tab. 4).

The mean relatedness of individuals within sites differs by only 0.0004 across the SC (Tab. 4). Sites with a more even sex ratio had a higher average kinship (GAB, TAS-NW, TAS-SE = 0.0005-0.0008) when compared to sites that were majority female (WA, VIC = 0.0004; Tab. 4). Relatedness estimates show a general pattern across all sites inferring that elevated r calculations shown at the first distance class (0 km) in all spatial autocorrelations (Fig 9 & 11) was not driven by a single location.

Table 4. Relatedness estimates per South-cluster sample site, excluding the **NSW-IN site. Sex ratio ($\text{♀}/\text{♂}$), mean relatedness (μ) and standard deviation (σ).

Site name	$\text{♀} / \text{♂}$	μ	σ
GAB	8/7	0.0007	0.0021
TAS-NW	8/7	0.0005	0.0023
TAS-SE	6/7	0.0008	0.0018
WA	9/6	0.0004	0.0012
VIC	8/2	0.0004	0.0010

*All spatial autocorrelation analyses include both individuals ($n = 2$) sampled from East coast site (NSW-IN) that were identified as genetically similar to the SC in analyses for genetic structure (Fig. 4).**These same individuals were excluded from the MLPE and relatedness estimates as the small sample size ($n = 2$) was not comparable to the other sample sites included in the population-level analyses (Fig. 10, Tab. 4).

3.4 Estimates of effective population size (N_e)

Stairway plot estimates of changes in effective population size (N_e) for the EC show a variable past. The first prediction of N_e (74,910 years ago (ya)) presented the lowest size with a median estimate of 1036.2688 (Fig. 12). A gradual increase in N_e from 74,910 – 28,421 ya made way for a comparatively rapid increase in N_e from 1104.063 - 5348.0602 individuals between 28,421 – 17,829 ya (Fig. 12). A long-standing plateau follows this increase from 17,829 - 155 ya where estimates of N_e increase by only 52.3638 individuals (Fig. 12; Tab. 5). The first estimated decrease in N_e follows this, with a gradual decline in size shown between 155 - 2 ya from 5400.424 - 1394.7342 individuals. A range of confidence intervals

encompass these estimates, with the broadest intervals shown surrounding periods of gradual increase and decrease (Fig. 12; Tab. 5). The particularly broad confidence intervals surrounding the gradual decrease observed in the most recent past (from 155 - 2 ya) as seen in Fig. 12 and Tab. 5 will need to be considered in any interpretation of these results.

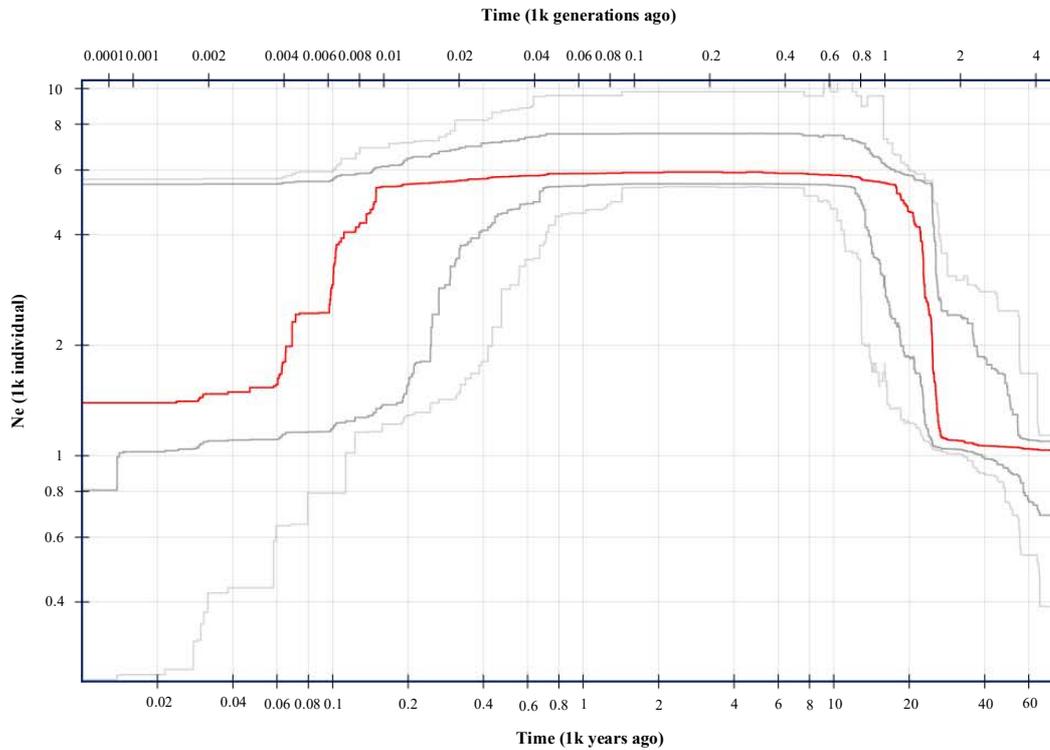


Figure 12. Stairway plot presenting estimated changes in effective population size (N_e) for the East-cluster from 74910 - 2 years ago. Comparisons of changes in N_e over time are made between 1k generations ago and 1k years ago on the top x-axis and bottom x-axis respectively. Dark grey lines denote 75% confidence intervals and light grey lines indicate 95% confidence intervals.

Table 5. Summary of major points of change in effective population size (N_e) for the East-cluster as estimated in Stairway plot. Year denotes the time (1k years ago) estimates were calculated for, while percentages show corresponding confidence intervals around the median approximation of N_e .

Year	N_e _median	75% CI		95% CI	
		lower	upper	lower	upper
74,910	1036.2688	388.5549	1136.4784	689.0854	1095.0648
28,421	1104.063	1012.4071	3085.6668	1044.2148	2413.4547
17,829	5348.0602	1343.3602	6602.0454	2294.4325	5930.8064
155	5400.424	1162.9875	6898.9141	1360.3227	6134.5144
2	1394.7342	242.9939	5662.4417	806.0623	5484.2

While this general trend appears analogous for both clusters, fine-scale observations of the increase and decline in N_e shows the variability in population size experienced by each cluster of *M. antarcticus* in their demographic history. A gradual increase in N_e in the SC from its first calculation 28,447 ya to 10,470 ya shows the median increase from a small 221.4731 individuals to 1508.6058 over the 17,977 year timespan (Fig. 13; Tab. 6). A more rapid increase in N_e follows this, occurring between 10,470 and 5,032 ya in which the median increases by 2,742.3068 individuals over a 5,438 year timespan (Fig. 13; Tab. 6). A long-standing plateau follows this whereby N_e shows an increase of only 781.527 individuals over 4,972 years (Fig. 13; Tab. 6). Rapid declines in N_e follow this plateau from a recent 44 ya to 10 ya where a decrease in N_e projections by 3128.953 individuals occurs over the short 33 year timespan (Fig. 13; Tab. 6). This projection marks the most rapid decrease in N_e experienced by both the EC and SC. A comparatively slower decline follows this until the final estimate of N_e (0.1128 ya) where a decrease of 59.053 individuals occurs over a 10 year time span (Fig. 13; Tab. 6), equating to an N_e of 782.5996 predicted for the present day. As with the EC, broader confidence intervals encompass median estimates of N_e through periods of gradual increase and decrease (Fig. 13; Tab. 6). The most recent decrease in N_e as described from 10 - 0 ya feature particularly broad confidence intervals as seen in Fig. 13 and Tab. 6.

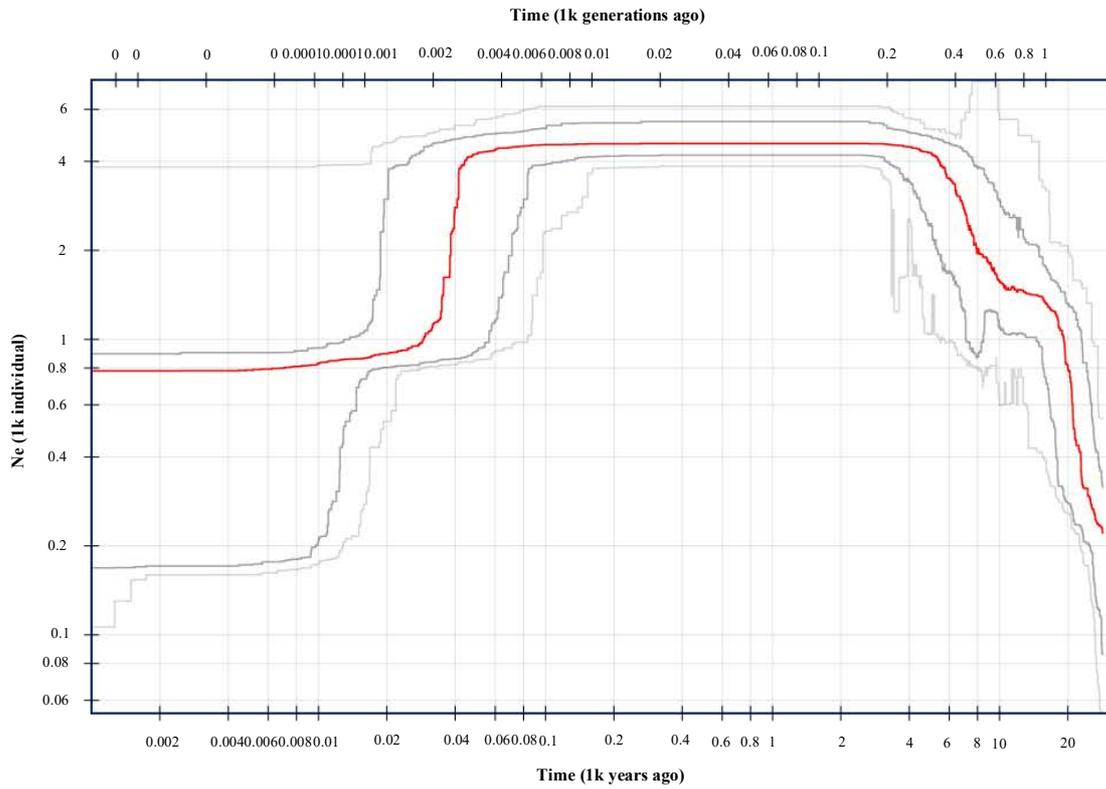


Figure 13. Stairway plot presenting estimated changes in effective population size (N_e) for the South-cluster from 28,447 - 0 years ago. Comparisons of changes in N_e over time are made between 1k generations ago and 1k years ago on the top x-axis and bottom x-axis respectively. Dark grey lines denote 75% confidence intervals and light grey lines indicate 95% confidence intervals.

Table 6. Summary of major points of change in effective population size (N_e) for the South-cluster as estimated in Stairway plot. Year denotes the time (1k years ago) estimates were calculated for, while percentages show corresponding confidence intervals around the median approximation of N_e .

Year	N_e _median	75% CI lower	75% CI upper	95% CI lower	95% CI upper
28,447	221.4731	54.2405	541.4438	85.9181	316.2093
10,470	1508.6058	602.0633	5551.7492	1048.5307	2738.1653
5,032	4250.9126	1048.4006	5146.1036	2420.0440	4812.6896
59	4382.5106	902.5508	5609.5863	1150.6696	4957.2429
44	3970.6057	837.8459	5294.3316	870.1175	4840.6208
10	841.6526	179.2778	3875.6861	234.5035	963.7001
0	782.5996	106.1823	3828.6646	168.4107	893.7577

4. Discussion

While marine ecosystems seemingly provide the right conditions for widespread dispersal and gene flow, species ranges are often subdivided into distinct populations (Ovenden 2013; Puckett & Eggleston 2016). Evidence of genetic partitioning into two stocks of Gummy shark was well supported. “Stocks” are defined here by the identified genetic clusters. One genetic stock included individual Gummy sharks sampled from both NSW sites (EC; NSW-IN & NSW-OFF Fig. 2, Tab.1) and those sampled at all sites along the South coast (SC) including Tasmania (VIC, TAS-NW, TAS-SE, GAB and WA; Fig. 2, Tab. 1) formed the second stock. The spatial structure of genetic variation within each of these stocks suggests widespread gene flow with some evidence for mild natal-site fidelity. Distinct populations form in response to a variety of biological and physical factors such as a species life history strategy, dispersal capacity, thermal tolerance, habitat requirements and the presence of paleo-land connections within their range (Miller et al. 2013; Ovenden 2013).

A complex interaction of historical and contemporary factors may have influenced the observed genetic structuring of the Gummy shark (Chabot et al. 2015). This is supported by the analysis of ancestral gene frequencies which showed little mixing between clusters in admixture results (Fig. 7). In the context of genetic structure, this indicates that the observed pattern of moderate genetic differentiation between individuals from the East and South coast of Australia has been long-standing. The physical and biological factors separating the Peronian province on the south-east coast and the Maugean province which encompasses Tasmania and the Victorian coast are likely influential (Waters & Roy 2003; Waters 2008). Such factors include the presence of a historical land bridge (the Bassian Isthmus) which connected mainland Australia to Tasmania in the Pleistocene, generating an east-south barrier to gene flow (Waters & Roy 2003; Waters 2008; Miller et al. 2013). The influence of this barrier on the historical distribution of the Gummy shark is expressed in estimates of the demographic history of each cluster (Fig. 12 & 13) which show a gradual increase in N_e from the earliest estimates concurrent with the population expansion estimated to have occurred for the species following a divergence event early in the Pleistocene (Boomer et al. 2012). Whilst the early Pleistocene occurred prior to the demographic estimates presented here, the gradual increase in N_e that occurred in both clusters from ~30,000 ya may be the result of residual founder effects occurring from the expansion of the *Mustelus* genus (Boomer et al.

2012). The species grew in size at a time concurrent with the gradual dissipation of the land bridge meaning that the disruption to gene flow occurred at a time when populations of the Gummy shark were at their largest (Fig. 12 & 13), promoting vicariance (Mirams et al. 2011). However, whether signals of genetic differentiation that result from historical disruptions to geneflow are subject to erosion overtime through post-Pleistocene dispersal, or are able to be retained until the present day is unknown (Waters 2008).

It is known, however, that prevailing ocean currents contributing to anomalous biogeographic boundaries can perpetuate historical patterns of genetic differentiation (Mirams et al. 2011; Briggs & Bowen 2013). This can occur even if a species has the capacity to disperse over large distances, because dispersal may be reduced within and around the biogeographic boundary due to the strength of currents (Mirams et al. 2011; Colgan 2016). This is the case for many species that exist between the Peronian-Maugean biogeographic boundary because this is where the south-flowing East Australian Current (EAC) on the Eastern Australian coast meets the south-east flowing Leeuwin current that functions along the Southern coast of Australia (Waters & Roy 2003; Colgan 2016). The interplay of these currents, particularly the strength of the EAC travelling Southwards, may discourage the dispersal of the Gummy shark between each coastline (Dawson 2014). The influence of the ocean currents that mark the Peronian-Maugean boundary on the genetic structuring of active dispersers has received very little research. A recent review compiling examples of species in which Australian biogeographic boundaries influenced the genetic structure of marine species reported on 17 Chordata species in total (at the Peronian-Maugean boundary), 16 of which were larval dispersers (Colgan 2016). Research by Gardner and Ward (1998) on the genetic structure of the Gummy shark was the only example of an active-dispersing, ovoviviparous chordate included in this review. Further research regarding the influence of ocean currents on the genetic structure of marine vertebrates with large dispersal potential is needed to understand this process. In addition, sample sites within this study do not include Southern NSW or Victoria on the Eastern coast. Further sampling that encompasses these areas is needed in order to clarify the point in which genetic structure begins for the species, and the proximity of this to the Peronian-Maugean boundary.

Demographic modelling of both the EC and SC revealed fluctuations in N_e whereby a rapid increase in N_e from the earliest projections is interrupted by a long-standing plateau and followed by a rapid decline in the more recent past (Fig. 12, Fig. 13). While both clusters

exhibit a comparably rapid decline, each cluster has a unique genetic history whereby the projected timing and magnitude of these estimates varied. Within the EC, the first estimated decrease (of 4,005.6898 individuals) in N_e began 155 ya and has continued to the present day (Fig. 12). It can be assumed that a gradual decrease in census size contributed to this trend as the key variable contributing to reductions in N_e is population loss (Frankham 1995). As the targeting of Gummy sharks for commercial harvest preceded the onset of this decline (Walker & Shotton 1999; Pribac et al. 2005), this observation may have potentially resulted from an intensification of the EAC (Ridgway 2007). A long-term observational study by Ridgway (2007) showed a significant change in the circulation of the EAC since 1944, with its waters becoming warmer and saltier over the 60 year period. Previous research linking choral chemistry to oceanographic and climate variability by Thresher et al. (2004) supported this finding and presented further evidence that variation in the strength of the EAC has persisted for up to 300 years. Shifts in the abundance and distribution of teleosts (the preferred food source of the Gummy shark) alongside changes in the availability of appropriate temporal habitat are likely to have resulted from the persistent variability of the EAC, potentially leading to the observed decline in N_e (van Putten et al. 2013). The declines of N_e estimated for the SC occurred during a time when potential influencing factors could be directly quantified. A rapidly declining N_e was estimated to have occurred as recently as 44 ya in the SC, interrupting a long-standing plateau 4,972 years long. During this decline, N_e is observed to drop from 3970.6057 to 841.6526 individuals (10 ya; Fig. 13, Tab. 6). Throughout this time period an estimated one-third decrease in Gummy shark catch along the South coast was observed by the Australian Southern Shark Fishery (recorded between 1973-76 and 1998-2001), signalling widespread population loss across their targeted range (Fig. 1; Walker & Shotton 1999).

The current estimate of N_e for the SC falls between the recommended 500-1000 individuals required to maintain a level of genetic variation that supports adaptive evolution in a changing environment (Lynch & Lande 1998). An N_e of approximately 1000 individuals is estimated to be genetically equivalent to a population of infinite size (Lynch & Lande 1998). As such, the present-day estimate of N_e for the EC of 1394.7342 individuals suggests that this cluster is also unlikely to suffer losses of genetic variation. This is further supported by calculations of F (EC = 0.0351, SC = 0.0451) which do not indicate a significant deficit of heterozygosity in either cluster (Tab. 3). However, estimates of allelic richness which are more adept at demonstrating genetic drift on short time-scales due to their sensitivity to rare

alleles (Pinsky & Palumbi 2014), were smaller for the SC than the EC (Tab. 3; A_r : EC= 0.0833, SC= 0.0294). Although the rate of decline in N_e has slowed in the past 10 years within the SC (Fig. 13, Tab. 6) it is still declining with evidence to suggest that genetic variation is beginning to slowly erode due to previous rapid population loss. Whilst conclusions on the rate and timing of the observed decline in N_e across each cluster should be interpreted as estimates only based on the broad confidence intervals presented (Fig. 12, Tab. 5, Fig. 13, Tab. 6), there is evidence to suggest that population declines have occurred for the species to the point where genetic variation was lost overtime (Charlesworth 2009).

Conclusions

The presence of genetic structure between East and South coast individuals suggests that management of the Gummy shark as one stock across their entire range would be unsuitable for their long-term sustainability. Reasoning supporting this suggestion is highlighted by the unique demographic history estimated for each cluster, detailing the relative impact of fishing pressure on the N_e of more heavily targeted areas. Although the genetic effects of overfishing on the South-cluster of the Gummy shark are not presently harmful with current estimates of N_e sufficient to offset genetic drift and support the current level of fishing pressure, the continued decline in N_e over the past ~44 years suggests that this may need to be re-examined in future.

References

- Allendorf FW, Berry O, Ryman N. 2014. So long to genetic diversity, and thanks for all the fish. *Molecular ecology* **23**:23-25.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**:81.
- Antao T, Pérez-Figueroa A, Luikart G. 2011. Early detection of population declines: high power of genetic monitoring using effective population size estimators. *Evolutionary Applications* **4**:144-154.
- Archer FI, Adams PE, Schneiders BB. 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* **17**:5-11.
- Banks SC, Cary GJ, Smith AL, Davies ID, Driscoll DA, Gill AM, Lindenmayer DB, Peakall R. 2013. How does ecological disturbance influence genetic diversity? *Trends in ecology & evolution* **28**:670-679.
- Benestan L. 2019. Population Genomics Applied to Fishery Management and Conservation. Boomer J. 2013. Genetic panmixia for two commercially important sharks (*Mustelus antarcticus* and *M. lenticulatus*). Biological sciences. Macquarie University.
- Boomer JJ, Harcourt RG, Francis MP, Stow AJ. 2012. Genetic divergence, speciation and biogeography of *Mustelus* (sharks) in the central Indo-Pacific and Australasia. *Molecular phylogenetics and evolution* **64**:697-703.
- Boomer JJ, Stow AJ. 2010. Rapid isolation of the first set of polymorphic microsatellite loci from the Australian gummy shark, *Mustelus antarcticus* and their utility across divergent shark taxa. *Conservation Genetics Resources* **2**:393-395.
- Briggs JC, Bowen BW. 2013. Marine shelf habitat: biogeography and evolution. *Journal of Biogeography* **40**:1023-1035.
- Cadrin SX, Kerr LA, Mariani S 2013. Stock identification methods: applications in fishery science. Academic Press.
- Chabot CL, Espinoza M, Mascareñas-Osorio I, Rocha-Olivares A. 2015. The effect of biogeographic and phylogeographic barriers on gene flow in the brown smoothhound shark, *Mustelus henlei*, in the northeastern Pacific. *Ecology and evolution* **5**:1585-1600.
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**:195.
- Colgan D. 2016. Marine and estuarine phylogeography of the coasts of south-eastern Australia. *Marine and Freshwater Research* **67**:1597-1610.
- Collie JS, Botsford LW, Hastings A, Kaplan IC, Largier JL, Livingston PA, Plagányi É, Rose KA, Wells BK, Werner FE. 2016. Ecosystem models for fisheries management: finding the sweet spot. *Fish and Fisheries* **17**:101-125.
- Dabney A, Storey JD, Warnes G. 2004. Q-value estimation for false discovery rate control. *Medicine* **344**:48.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156-2158.
- DArT. 2018. DNA Extraction Protocol for DArT, Available from https://ordering.diversityarrays.com/files/DArT_DNA_isolation.pdf (accessed 8.11.18 2018).

- Davidson LN, Krawchuk MA, Dulvy NK. 2016. Why have global shark and ray landings declined: improved management or overfishing? *Fish and Fisheries* **17**:438-458.
- Dawson MN. 2014. Biogeography and complex traits: dispersal syndromes, in the sea. *Frontiers of Biogeography* **6**.
- Dray S, Dufour A-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* **22**:1-20.
- Dulvy NK, Fowler SL, Musick JA, Cavanagh RD, Kyne PM, Harrison LR, Carlson JK, Davidson LN, Fordham SV, Francis MP. 2014. Extinction risk and conservation of the world's sharks and rays. *elife* **3**:e00590.
- Duncan K, Martin A, Bowen B, De Couet H. 2006. Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular ecology* **15**:2239-2251.
- Eaton D, Overcast I. 2016. ipyrad: interactive assembly and analysis of RADseq data sets. <http://github.com/dereneaton/ipyrad>.
- Edwards A. 1971. Distances between populations on the basis of gene frequencies. *Biometrics*:873-881.
- Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**:285.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics* **1**:117693430500100003.
- Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**:977-993.
- Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* **66**:95-107.
- Frankham R. 2005. Genetics and extinction. *Biological conservation* **126**:131-140.
- Frankham R. 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological conservation* **143**:1919-1927.
- Frankham R, Briscoe DA, Ballou JD 2002. *Introduction to conservation genetics*. Cambridge university press.
- Frichot E, François O. 2015. LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6**:925-929.
- Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**:973-983.
- Galván-Tirado C, Díaz-Jaimes P, García-de León FJ, Galván-Magaña F, Uribe-Alcocer M. 2013. Historical demography and genetic differentiation inferred from the mitochondrial DNA of the silky shark (*Carcharhinus falciformis*) in the Pacific Ocean. *Fisheries Research* **147**:36-46.
- Gardner M, Ward R. 1998. Population structure of the Australian gummy shark (*Mustelus antarcticus* Günther) inferred from allozymes, mitochondrial DNA and vertebrae counts. *Marine and Freshwater Research* **49**:733-745.
- Gordon A, Hannon G. 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab.cshl.edu/fastx_toolkit **5**.
- Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology* **25**:438-449.
- Harmon LJ, Glor RE. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution: International Journal of Organic Evolution* **64**:2173-2178.

- Hawkins S, Bohn K, Sims D, Ribeiro P, Faria J, Presa P, Pita A, Martins G, Neto A, Burrows M. 2016. Fisheries stocks from an ecological perspective: Disentangling ecological connectivity from genetic interchange. *Fisheries Research* **179**:333-341.
- Helidoniotis F, Emery, T, Woodhams, J, Curtotti, R. 2019. Chapter 8: Southern and Eastern Scalefish and Shark Fishery in Agriculture Do, editor.
- Hilborn R, Ovando D. 2014. Reflections on the success of traditional fisheries management. *ICES journal of Marine Science* **71**:1040-1046.
- Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**:3070-3071.
- Karl SA, Castro AL, Garla RC. 2012. Population genetics of the nurse shark (*Ginglymostoma cirratum*) in the western Atlantic. *Marine Biology* **159**:489-498.
- Keenan K, McGinnity P, Cross T, Crozier W, Prodöhl P. 2013. diveRcity: an R package for the estimation of population genetics parameters and their associated errors. *Methods Ecol Evol* **4**: 782–788.
- Keeney D, Heist E. 2006. Worldwide phylogeography of the blacktip shark (*Carcharhinus limbatus*) inferred from mitochondrial DNA reveals isolation of western Atlantic populations coupled with recent Pacific dispersal. *Molecular Ecology* **15**:3669-3679.
- Kelly RP, Port JA, Yamahara KM, Crowder LB. 2014. Using environmental DNA to census marine fishes in a large mesocosm. *PloS one* **9**:e86175.
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. Pages 67-89. *Data production and analysis in population genomics*. Springer.
- Legendre P, Fortin MJ, Borcard D. 2015. Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6**:1239-1247.
- Linck E, Battey C. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular ecology resources* **19**:639-647.
- Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nature genetics* **47**:555.
- Lowe WH, Allendorf FW. 2010. What can genetics tell us about population connectivity? *Molecular ecology* **19**:3038-3051.
- Lynch M, Lande R. 1998. The critical effective size for a genetically secure population. *Animal Conservation* **1**:70-72.
- MacDonald CM. 1988. Genetic variation, breeding structure and taxonomic status of the gummy shark *Mustelus antarcticus* in southern Australian waters. *Marine and Freshwater Research* **39**:641-648.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer research* **27**:209-220.
- Marandel F, Lorange P, Andrello M, Charrier G, Le Cam S, Lehuta S, Trenkel VM. 2017. Insights from genetic and demographic connectivity for the management of rays and skates. *Canadian Journal of Fisheries and Aquatic Sciences* **75**:1291-1302.
- Maunder MN, Piner KR. 2014. Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science* **72**:7-18.
- Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC. 2013. Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia—implications for predicting the movement of passive dispersers across a marine biogeographic barrier. *Ecology and evolution* **3**:1248-1261.
- Miramis A, Trembl EA, Shields J, Liggins L, Riginos C. 2011. Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. *Coral Reefs* **30**:937-949.

- Ortuño Crespo G, Dunn DC. 2017. A review of the impacts of fisheries on open-ocean ecosystems. *ICES Journal of Marine Science* **74**:2283-2297.
- Ovenden JR. 2013. Crinkles in connectivity: combining genetics and other types of biological data to estimate movement and interbreeding between populations. *Marine and Freshwater Research* **64**:201-207.
- Ovenden JR, Dudgeon C, Feutry P, Feldheim K, Maes GE. 2018. Genetics and Genomics for Fundamental and Applied Research on Elasmobranchs. *Shark Research: Emerging Technologies and Applications for the Field and Laboratory*.
- Overcast I. 2017. easySFS, GitHub.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2012. nlme: Linear and nonlinear mixed effects models. R package version **3**.
- Pinsky ML, Palumbi SR. 2014. Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology* **23**:29-39.
- Pope N. 2018. CorMLPE: a correlation structure for symmetric relational data, Available from <https://github.com/nspope/corMLPE/commit/80100fcf02c86a85431e72bf61df9368da822772>.
- Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP. 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular ecology* **25**:527-541.
- Pribac F, Punt AE, Walker T, Taylor B. 2005. Using length, age and tagging data in a stock assessment of a length selective fishery for gummy shark (*Mustelus antarcticus*). *Journal of Northwest Atlantic Fishery Science* **35**:267-290.
- Puckett B, Eggleston D. 2016. Metapopulation dynamics guide marine reserve design: importance of connectivity, demographics, and stock enhancement. *Ecosphere* **7**:e01322.
- Raymond M, Rousset F. 1995. GENEPOP on the Web (Version 3.4). URL: <http://wbiomed.curtin.edu.au/genepop/> Updated from Raymond & Rousset.
- Raymond M, Rousset F. 2004. GENEPOP on the web. Montpellier, FR: Lab. Genet. Environ.
- Ridgway K. 2007. Long-term trend and decadal variability of the southward penetration of the East Australian Current. *Geophysical Research Letters* **34**.
- Rubinas M. 2017. Using Knowledge of Shark Biology and Behavior to Inform Management and Conservation Efforts for the Globally Distributed and Fishery Exploited Shortfin Mako (*Isurus oxyrinchus*).
- Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. 2011. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. Page P54. BMC proceedings. BioMed Central.
- Schultz J, Feldheim K, Gruber S, Ashley M, McGovern T, Bowen B. 2008. Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Molecular Ecology* **17**:5336-5348.
- Selkoe KA, Aloia CC, Crandall ED, Iacchei M, Liggins L, Puritz JB, von der Heyden S, Toonen RJ. 2016. A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series* **554**:1-19.
- Simpfendorfer CA, Dulvy NK. 2017. Bright spots of sustainable shark fishing. *Current Biology* **27**:R97-R98.
- Smith SE, Au DW, Show C. 1998. Intrinsic rebound potentials of 26 species of Pacific sharks. *Marine and Freshwater Research* **49**:663-678.
- Smouse RPP, Peakall R. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**:2537-2539.

- Stow A, Zenger K, Briscoe D, Gillings M, Peddemors V, Otway N, Harcourt R. 2006. Isolation and genetic diversity of endangered grey nurse shark (*Carcharias taurus*) populations. *Biology Letters* **2**:308-311.
- Stow AJ, Sunnucks P, Briscoe D, Gardner M. 2001. The impact of habitat fragmentation on dispersal of Cunningham's skink (*Egernia cunninghami*): evidence from allelic and genotypic analyses of microsatellites. *Molecular ecology* **10**:867-878.
- Swartz W, Sala E, Tracey S, Watson R, Pauly D. 2010. The spatial expansion and ecological footprint of fisheries (1950 to present). *PloS one* **5**:e15143.
- Taillebois L, Castelin M, Ovenden JR, Bonillo C, Keith P. 2013. Contrasting genetic structure among populations of two amphidromous fish species (Sicydiinae) in the Central West Pacific. *PloS one* **8**:e75465.
- Thresher R, Rintoul SR, Koslow JA, Weidman C, Adkins J, Proctor C. 2004. Oceanic evidence of climate change in southern Australia over the last three centuries. *Geophysical Research Letters* **31**.
- Valencia LM, Martins A, Ortiz EM, Di Fiore A. 2018. A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. *PloS one* **13**:e0201254.
- van Putten IE, Jennings S, Frusher S, Gardner C, Haward M, Hobday AJ, Nursey-Bray M, Pecl G, Punt A, Revill H. 2013. Building blocks of economic resilience to climate change: a south east Australian fisheries example. *Regional Environmental Change* **13**:1313-1323.
- Vasconcelos RP, Eggleston DB, Le Pape O, Tulp I. 2013. Patterns and processes of habitat-specific demographic variability in exploited marine species. *ICES Journal of Marine Science* **71**:638-647.
- Walker T, Shotton R. 1999. Southern Australian shark fishery management.
- Walker TI, Hudson RJ, Gason AS. 2005. Catch evaluation of target, by-product and by-catch species taken by gillnets and longlines in the shark fishery of south-eastern Australia. *Journal of Northwest Atlantic Fishery Science* **35**:505-530.
- Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature methods* **9**:808.
- Waters JM. 2008. Marine biogeographical disjunction in temperate Australia: historical landbridge, contemporary currents, or both? *Diversity and Distributions* **14**:692-700.
- Waters JM, Roy MS. 2003. Marine biogeography of southern Australia: phylogeographical structure in a temperate sea-star. *Journal of Biogeography* **30**:1787-1796.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *evolution* **38**:1358-1370.
- Whitlock MC, Lotterhos KE. 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. *The American Naturalist* **186**:S24-S36.
- Woodhams J, Green, C, Lyle, J, Braccini, M, Rogers, P, Peddemors, V. 2018. Gummy shark (2018), Available from <http://fish.gov.au/report/181-Gummy-Shark-2018> (accessed 11/3/19).
- Wright D, Bishop JM, Matthee CA, von der Heyden S. 2015. Genetic isolation by distance reveals restricted dispersal across a range of life histories: implications for biodiversity conservation planning across highly variable marine environments. *Diversity and Distributions* **21**:698-710.
- Zhang L, Yang X, Qi X, Guo C, Jing Z. 2018. Characterizing the transcriptome and microsatellite markers for almond (*Amygdalus communis* L.) using the Illumina sequencing platform. *Hereditas* **155**:14.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**:3326-3328.

Supplementary Material

Appendix 1.1: Sample list

Species	Sample name	Sample location	Tissue type
<i>MustelusAntarcticus</i>	WPB_J21M	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J20F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J1F	Western Port Bay (VIC)	fin
<i>MustelusAntarcticus</i>	WPB_J3F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J6F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J11F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J26F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J10F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J12F	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	WPB_J24M	Western Port Bay (VIC)	muscle
<i>MustelusAntarcticus</i>	TAS13M	South-East Tasmania	muscle
<i>MustelusAntarcticus</i>	TAS19M	South-East Tasmania	muscle
<i>MustelusAntarcticus</i>	TAS3M	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS2M	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS6M	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS7F	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS8F	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS10F	South-East Tasmania	muscle
<i>MustelusAntarcticus</i>	TAS4M	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS20F	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS1M	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS21F	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	TAS18F	South-East Tasmania	fin
<i>MustelusAntarcticus</i>	NSWOF9F	NSW offshore	muscle (all)
<i>MustelusAntarcticus</i>	WA61M	Albany (WA)	muscle (all)
<i>MustelusAntarcticus</i>	NSWIN_JM190308-01F	NSW inshore	muscle (all)
<i>MustelusAntarcticus</i>	GAB_S04324M	Great Australian Bite (SA)	muscle (all)
<i>MustelusAntarcticus</i>	NWT_S03235M	North-West Tasmania	muscle (all)

Appendix 1.2: ipyrad parameters input file

```
----- ipyrad params file (v.0.7.29)-----
mustelus_strict      ## [0] [assembly_name]: Assembly name. Used to name output
directories for assembly steps
~/Desktop/ipyRAD_mustelus_denovo_strict      ## [1] [project_dir]: Project dir
(made in curdir if not present)
      ## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
      ## [3] [barcodes_path]: Location of barcodes file
~/Desktop/renamed/trimmed/trimmed2/*.FASTQ.gz      ## [4]
[sorted_fastq_path]: Location of demultiplexed/sorted fastq files
denovo      ## [5] [assembly_method]: Assembly method (denovo, reference,
denovo+reference, denovo-reference)
      ## [6] [reference_sequence]: Location of reference sequence file.
rad      ## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
      ## I have rad data.
TGCAG      ## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2).
      ## Standard for the enzyme used (pst1).
5 in a read      ## [9] [max_low_qual_bases]: Max low quality base calls.
      ## Standard, allows for up to 5 low quality bases in a read.
33      ## [10] [phred_Qscore_offset]: phred Q score offset.
      ## Default, standard for illumina data.
6      ## [11] [mindepth_statistical]: Min depth for statistical base calling.
      ## Standard, for most reasonable error rate estimates 6 is approximately
      ## the minimum depth at which a heterozygous base call can be
      ## distinguished from a sequencing error.
10      ## [12] [mindepth_majrule]: Min depth for majority-rule base calling.
      ## Upped it from the standard 6 reads to improve confidence of
      ## homozygosity. The chance is low at >10 (~1/2) because we analysed
      ## more alleles.
500      ## [13] [maxdepth]: Max cluster depth within samples.
      ## Default is 10000 which is considered quite high, chose to set it lower
      ## to avoid over-representation within reads.
0.9      ## [14] [clust_threshold]: Clustering threshold for de novo assembly.
      ## 0.85-0.90 is considered a fairly reliable range, balancing over-splitting
      ## of loci vs over-lumping.
0      ## [15] [max_barcode_mismatch]: Max number of allowable mismatches
      ## in barcodes.
      ## Want no mismatches between barcodes file and sequenced reads.
2      ## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter).
      ## This option searches for the common Illumina adapter, plus the reverse
      ## complement of the second cut site (if present), plus the barcode (if
      ## present), and this part of the read is trimmed.
35      ## [17] [filter_min_trim_len]: Min length of reads after adapter trim.
      ## Standard for ipyRad.
2      ## [18] [max_alleles_consens]: Max alleles per site in consensus
      ## sequences.
      ## Default is 2 which is appropriate for diploids. At this setting any locus
      ## which has a sample with more than 2 alleles detected will be
      ##excluded/filtered out.
```

```

3      ## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus (R1,
      ## R2).
      ## Default is 5, however that is quite large for our sequences of 69bp
      ## reads. So chose 3 instead because we don't want more than 3 bases in
      ## our reads where there are uncalled bases.
8      ## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1,
      ## R2).
      ## Standard, helps to remove poor alignments which will tend to have an
      ## excess of Hs.
105    ## [21] [min_samples_locus]: Min # samples per locus for output.
      ## 105 samples out of the original 117 (including replicates) equates to
      ## only 10% missing data allowed, any more than that and the loci
      ## wouldn't be very informative.
5      ## [22] [max_SNPs_locus]: Max # SNPs per locus (R1, R2).
      ## Default is 20, although setting a lower value is helpful for extra
      ## filtering in case of a messy dataset.
5      ## [23] [max_Indels_locus]: Max # of indels per locus (R1, R2).
      ## Default for single end data.
0.5    ## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus (R1,
      ## R2).
      ## Default for single end data.
0, 0, 0, 0  ## [25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see
      ## docs).
      ## Data already trimmed.
0, 0, 0, 0  ## [26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2).
      ## Data already trimmed.
v      ## [27] [output_formats]: Output formats (see docs).
      ## [28] [pop_assign_file]: Path to population assignment file.

```

Appendix 1.3: Stairway plot blueprint file – East-cluster

```
#East-pop
#input setting
popid: east-pop # id of the population (no white space)
nseq: 48 # number of sequences
L: 2433699 # total number of observed nucleic sites, including polymorphic and
monomorphic
whether_folded: true # whether the SFS is folded (true or false)
SFS: 4339 2045 940 688 430 361 264 222 214 160 137 137 118 135 110 95 78 101 76 93 95
76 73 42 # snp frequency spectrum: number of singleton, number of doubleton, etc.
(separated by white space)
#smallest_size_of_SFS_bin_used_for_estimation: 1 # default is 1; to ignore singletons,
change this number to 2
#largest_size_of_SFS_bin_used_for_estimation: 24 # default is n-1; to ignore singletons,
change this number to nseq-2
pct_training: 0.67 # percentage of sites for training
nrand: 12 23 35 46 # number of random break points for each try (separated by
white space)
project_dir: east-pop # project directory
stairway_plot_dir: stairway_plot_es # directory to the stairway plot files
ninput: 200 # number of input files to be created for each estimation
#output setting
mu: 0.0000001 # assumed mutation rate per site per generation
year_per_generation: 16 # assumed generation time (in years)
#plot setting
plot_title: east-pop # title of the plot
xrange: 0.01,10000 # Time (1k year) range; format: xmin,xmax; "0,0" for default
yrange: 0,0 # Ne (1k individual) range; format: xmin,xmax; "0,0" for default
xspacing: 2 # X axis spacing
yspacing: 2 # Y axis spacing
fontsize: 12 # Font size
```

Appendix 1.4: Stairway plot blueprint file – South-cluster

```
#South-pop
#input setting
popid: south-pop # id of the population (no white space)
nseq: 140 # number of sequences
L: 2433699 # total number of observed nucleic sites, including polymorphic and
monomorphic
whether_folded: true # whether the SFS is folded (true or false)
SFS: 3155 1754 695 601 370 308 246 196 165 147 103 90 83 78 68 70 54 49 41 29 24 37 39
28 28 33 28 20 31 21 27 28 22 18 17 23 24 13 12 13 16 8 17 14 16 9 8 12 13 11 11 13 7 11 8
7 9 10 7 5 5 9 3 5 8 7 5 4 4 7 # snp frequency spectrum: number of singleton, number of
doubleton, etc. (separated by white space)
#smallest_size_of_SFS_bin_used_for_estimation: 1 # default is 1; to ignore singletons,
change this number to 2
#largest_size_of_SFS_bin_used_for_estimation: 70 # default is n-1; to ignore singletons,
change this number to nseq-2
pct_training: 0.67 # percentage of sites for training
nrand: 35 69 103 138 # number of random break points for each try (separated by
white space)
project_dir: south-pop # project directory
stairway_plot_dir: stairway_plot_es # directory to the stairway plot files
ninput: 200 # number of input files to be created for each estimation
#output setting
mu: 0.0000001 # assumed mutation rate per site per generation
year_per_generation: 16 # assumed generation time (in years)
#plot setting
plot_title: south-pop # title of the plot
xrange: 0.001,10000 # Time (1k year) range; format: xmin,xmax; "0,0" for default
yrange: 0,0 # Ne (1k individual) range; format: xmin,xmax; "0,0" for default
xspacing: 2 # X axis spacing
yspacing: 2 # Y axis spacing
fontsize: 12 # Font size
```

Appendix 2.1: Loci identified under positive selection

HWE ($p < 0.05$)	Outflank (FDR < 0.05)		Arlequin (FDR < 0.05)	Bayescan (FDR < 0.05)	
locus_599	locus_168502	locus_405	locus_7138	locus_137675	locus_405
locus_926	locus_169250	locus_462	locus_10308	locus_138200	locus_3608
locus_1705	locus_169372	locus_1898	locus_10649	locus_139999	locus_3739
locus_2240	locus_170141	locus_2621	locus_11714	locus_140622	locus_5052
locus_3573	locus_171290	locus_3265	locus_15025	locus_141194	locus_7138
locus_3718	locus_171490	locus_3300	locus_15243	locus_147777	locus_7294
locus_4058	locus_171556	locus_3608	locus_17870	locus_148423	locus_9394
locus_8123	locus_171570	locus_3633	locus_18375	locus_148612	locus_9789
locus_8777	locus_171869	locus_3739	locus_24828	locus_152482	locus_9825
locus_8962	locus_114677	locus_4981	locus_32071	locus_152877	locus_10308
locus_9846	locus_115953	locus_5052	locus_34299	locus_153244	locus_10649
locus_10236	locus_116193	locus_5714	locus_39622	locus_154759	locus_11714
locus_10541	locus_117209	locus_6115	locus_42062	locus_155018	locus_11750
locus_12060	locus_117386	locus_7138	locus_42764	locus_155616	locus_13331
locus_13754	locus_117998	locus_7294	locus_50670	locus_155880	locus_14069
locus_13837	locus_118195	locus_7319	locus_51336	locus_157227	locus_14351
locus_16256	locus_119561	locus_7346	locus_56544	locus_158191	locus_15025
locus_16360	locus_119811	locus_7433	locus_57110	locus_158344	locus_15243
locus_16431	locus_119898	locus_7884	locus_62949	locus_158630	locus_15469
locus_18049	locus_120093	locus_7909	locus_64212	locus_159010	locus_16192
locus_18774	locus_121161	locus_8454	locus_66222	locus_159069	locus_16660
locus_19028	locus_121842	locus_8768	locus_66618	locus_159126	locus_17759
locus_19489	locus_122462	locus_8822	locus_66839	locus_159202	locus_17870
locus_19564	locus_122516	locus_9394	locus_72175	locus_159609	locus_18375
locus_20971	locus_123109	locus_9789	locus_73029	locus_160471	locus_18408
locus_21750	locus_123110	locus_9825	locus_77507	locus_160874	locus_21582
locus_23157	locus_123298	locus_9956	locus_81026	locus_162521	locus_24615
locus_23889	locus_124479	locus_10308	locus_81268	locus_162787	locus_24828
locus_24481	locus_125267	locus_10313	locus_82649	locus_166527	locus_27040
locus_25929	locus_125535	locus_10649	locus_87931	locus_167107	locus_28412
locus_26190	locus_126328	locus_11184	locus_88222	locus_167621	locus_30231
locus_27103	locus_127396	locus_11203	locus_96014	locus_168460	locus_31003
locus_27328	locus_127522	locus_11714	locus_99143	locus_168502	locus_31680
locus_28591	locus_128045	locus_11750	locus_99783	locus_169250	locus_31880
locus_28602	locus_129219	locus_12307	locus_101490	locus_169372	locus_32071
locus_31676	locus_130350	locus_12478	locus_101873	locus_171869	locus_33340
locus_35151	locus_130616	locus_13004	locus_103105	locus_137675	locus_34299
locus_35286	locus_131423	locus_13073	locus_104141	locus_138200	locus_35615
locus_35566	locus_132334	locus_13331	locus_105044	locus_139999	locus_35691
locus_36703	locus_132983	locus_13961	locus_105448	locus_140622	locus_38232
locus_38800	locus_133000	locus_14069	locus_105699	locus_141194	locus_38996
locus_40467	locus_133322	locus_14351	locus_110223	locus_147777	locus_39622
locus_42679	locus_133543	locus_15021	locus_113778	locus_148423	locus_39779
locus_46782	locus_133904	locus_15025	locus_116193	locus_148612	locus_42062

locus_47430	locus_135731	locus_15243	locus_120093	locus_152482	locus_42567
locus_48034	locus_136528	locus_15469	locus_127396	locus_152877	locus_42764
locus_49322	locus_137293	locus_15510	locus_129219	locus_107439	locus_43154
locus_51705	locus_137455	locus_16192	locus_148612	locus_107516	locus_47382
locus_52607	locus_137675	locus_16660	locus_152877	locus_108318	locus_47861
locus_58238	locus_138200	locus_17759	locus_153244	locus_108567	locus_49719
locus_58498	locus_139010	locus_17870	locus_155018	locus_108754	locus_50007
locus_60868	locus_139704	locus_18207	locus_155880	locus_110127	locus_50595
locus_64673	locus_139737	locus_18228	locus_158344	locus_110223	locus_50670
locus_67091	locus_139999	locus_18375	locus_159010	locus_112760	locus_50900
locus_67100	locus_140622	locus_18408	locus_159202	locus_112975	locus_51336
locus_67539	locus_141194	locus_18841	locus_159609	locus_113778	locus_52114
locus_72348	locus_141586	locus_20590	locus_162521	locus_114619	locus_52787
locus_75312	locus_144069	locus_21582	locus_162787	locus_115953	locus_52952
locus_78090	locus_145794	locus_21864	locus_166527	locus_116193	locus_53325
locus_79289	locus_147777	locus_24364	locus_167107	locus_117998	locus_54330
locus_81020	locus_148423	locus_24615	locus_169250	locus_118195	locus_55161
locus_81605	locus_148612	locus_24828		locus_119811	locus_55833
locus_82895	locus_151399	locus_25861		locus_119898	locus_56221
locus_84105	locus_151844	locus_26383		locus_120093	locus_56544
locus_84761	locus_152482	locus_27040		locus_121161	locus_56562
locus_85054	locus_152877	locus_27709		locus_122462	locus_56632
locus_85161	locus_153244	locus_28110		locus_122516	locus_56709
locus_85358	locus_154759	locus_28412		locus_123109	locus_57110
locus_85636	locus_155018	locus_28487		locus_124479	locus_57618
locus_86415	locus_155328	locus_29071		locus_125267	locus_62471
locus_87721	locus_155616	locus_29331		locus_127396	locus_62949
locus_87834	locus_155880	locus_29449		locus_129219	locus_64212
locus_88097	locus_157227	locus_30231		locus_130350	locus_64408
locus_89366	locus_157310	locus_31003		locus_132334	locus_65558
locus_90919	locus_157627	locus_31498		locus_132983	locus_65732
locus_91721	locus_157921	locus_31680		locus_133322	locus_65891
locus_93664	locus_158191	locus_31880		locus_136528	locus_66222
locus_94562	locus_158344	locus_32071		locus_137293	locus_66618
locus_95360	locus_158630	locus_32167		locus_100812	locus_66839
locus_96169	locus_159010	locus_33340		locus_100973	locus_68431
locus_97862	locus_159069	locus_33823		locus_101189	locus_69105
locus_98317	locus_159126	locus_34299		locus_101490	locus_70967
locus_100996	locus_159202	locus_34580		locus_101873	locus_71233
locus_101040	locus_159263	locus_35396		locus_102269	locus_72175
locus_101738	locus_159609	locus_35615		locus_103105	locus_72385
locus_101831	locus_159664	locus_35691		locus_103811	locus_72767
locus_101917	locus_159764	locus_36649		locus_104141	locus_73029
locus_101975	locus_160133	locus_36920		locus_104981	locus_73630
locus_102092	locus_160471	locus_37941		locus_105044	locus_74900
locus_102138	locus_160874	locus_38232		locus_105448	locus_74967
locus_104323	locus_161531	locus_38996		locus_105496	locus_77507
locus_104659	locus_162521	locus_39400		locus_105699	locus_79226

locus_104664	locus_162787	locus_39622		locus_106787	locus_79517
locus_105015	locus_163273	locus_39779		locus_94066	locus_81026
locus_105056	locus_164124	locus_40464		locus_96014	locus_81201
locus_105552	locus_165624	locus_40576		locus_96342	locus_81268
locus_105718	locus_165688	locus_41311		locus_96478	locus_81483
locus_105873	locus_166527	locus_42062		locus_96657	locus_82560
locus_107464	locus_166692	locus_42567		locus_97009	locus_82649
locus_107817	locus_167107	locus_42764		locus_97032	locus_83265
locus_108580	locus_167237	locus_43154		locus_97114	locus_84537
locus_109980	locus_167621	locus_47382		locus_97489	locus_86288
locus_111985	locus_168460	locus_47861		locus_97851	locus_87931
locus_112891	locus_105448	locus_49719		locus_99085	locus_87966
locus_114572	locus_105496	locus_50007		locus_99143	locus_88222
locus_115906	locus_105699	locus_50595		locus_99732	locus_88274
locus_121278	locus_106787	locus_50670		locus_99783	locus_88451
locus_124758	locus_107439	locus_50831		locus_100731	locus_88576
locus_125623	locus_107516	locus_50900		locus_100744	locus_88969
locus_125990	locus_108025	locus_51257		locus_92080	locus_91135
locus_126922	locus_108049	locus_51276		locus_92500	locus_91279
locus_127096	locus_108302	locus_51336		locus_93043	locus_91900
locus_132575	locus_108318	locus_52114			
locus_133084	locus_108567	locus_52787			
locus_134574	locus_108661	locus_52952			
locus_134634	locus_108754	locus_52960			
locus_137539	locus_110127	locus_53325			
locus_137969	locus_110223	locus_53782			
locus_138261	locus_112760	locus_53969			
locus_139996	locus_112975	locus_54330			
locus_140188	locus_113778	locus_54369			
locus_140698	locus_114619	locus_55161			
locus_147868	locus_97489	locus_55833			
locus_148101	locus_97521	locus_56102			
locus_148198	locus_97851	locus_56221			
locus_150255	locus_98153	locus_56544			
locus_151007	locus_98389	locus_56562			
locus_155373	locus_99085	locus_56632			
locus_156710	locus_99143	locus_56709			
locus_159622	locus_99483	locus_57110			
locus_162092	locus_99485	locus_57618			
locus_162093	locus_99732	locus_57890			
locus_163693	locus_99783	locus_60051			
locus_164400	locus_100731	locus_60382			
locus_170509	locus_100744	locus_62003			
locus_171045	locus_100776	locus_62471			
locus_172009	locus_100812	locus_62949			
	locus_100973	locus_63986			
	locus_101189	locus_64032			
	locus_101316	locus_64212			

	locus_101490	locus_64408			
	locus_101873	locus_65558			
	locus_102269	locus_65732			
	locus_103105	locus_65891			
	locus_103811	locus_66222			
	locus_103959	locus_66618			
	locus_104120	locus_66839			
	locus_104141	locus_67561			
	locus_104429	locus_68233			
	locus_104975	locus_68431			
	locus_104981	locus_69105			
	locus_105044	locus_70967			
	locus_105318	locus_71233			
	locus_90598	locus_72175			
	locus_91135	locus_72385			
	locus_91279	locus_72767			
	locus_91738	locus_73029			
	locus_91900	locus_73615			
	locus_91987	locus_73630			
	locus_92027	locus_74900			
	locus_92080	locus_74967			
	locus_92135	locus_76979			
	locus_92388	locus_77118			
	locus_92500	locus_77507			
	locus_92912	locus_78324			
	locus_93043	locus_78412			
	locus_93626	locus_79226			
	locus_93771	locus_79517			
	locus_94066	locus_81026			
	locus_94614	locus_81201			
	locus_95760	locus_81268			
	locus_95876	locus_81483			
	locus_95904	locus_81643			
	locus_96014	locus_82054			
	locus_96342	locus_82560			
	locus_96478	locus_82649			
	locus_96657	locus_83168			
	locus_97032	locus_83265			
	locus_97114	locus_83441			
	locus_97360	locus_83576			
	locus_88222	locus_84537			
	locus_88274	locus_85113			
	locus_88451	locus_85362			
	locus_88576	locus_85746			
	locus_88868	locus_86288			
	locus_88969	locus_87931			
	locus_89042	locus_87966			
	locus_90404				

Appendix 2.2: PCA - loci under positive selection

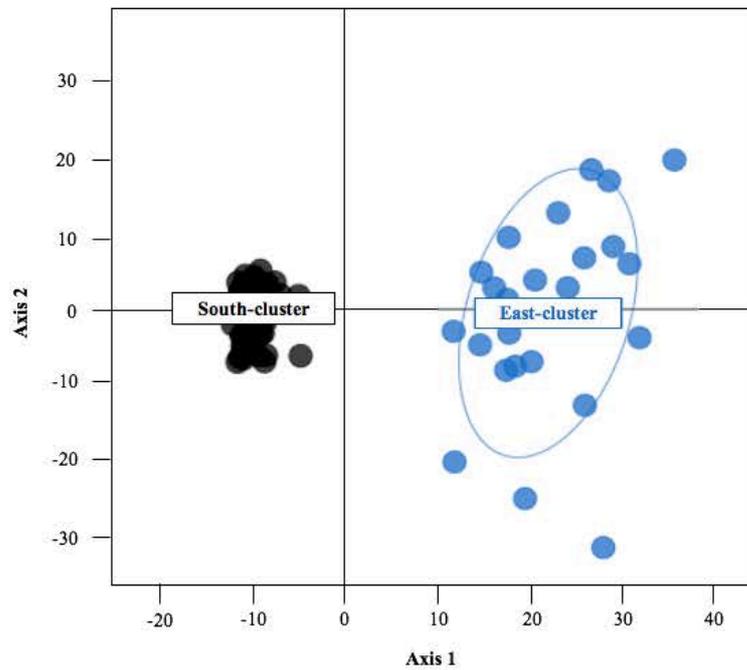


Figure X. Principal component analysis (PCA) of loci under positive selections ($n = 376$). PCA shows the separation of individuals ($n=94$; indicated by circle points) from 0 into two distinct groups. Ellipses indicate individuals with a normal distribution (within) and those that are outliers (outside) within each group.

Appendix 2.3: Mantel test – results

No signal of isolation by distance (IBD) was identified by the mantel test for the SC of *Mustelus antarcticus*. The mantel correlation (Fig. X) showed no spatial pattern of genetic variation, with no linear relationship observed between genetic and geographic distances within the SC. There is a distinction in comparisons of genetic distance split between a higher genetic distance range (0.23-0.25) and a lower range (0.10-0.14) with no comparisons found outside of these ranges. Comparison points within these upper and lower ranges are evenly spread across the sample sites contributing to the geographic distance of the test (Fig. X).

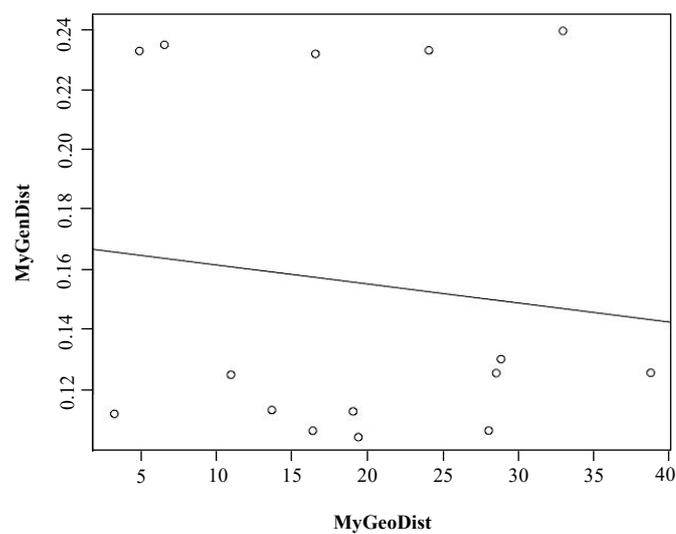


Figure X. Mantel correlation plot comparing the genetic distance of all individuals in the South-cluster ($n=70$) to the geographic location of their associated sample sites. Circles indicate points of comparison (GenDist/GeoDist) based on the number of sample locations ($n=6$) in the South-cluster.

Mantel tests for both female and male individuals (Fig. XXab) of *M. antarcticus* in the SC showed a lack of spatial structure in the distribution of genetic variation concurrent with results from comparisons of the entire SC (Fig. X). Alongside this, a similar pattern of separation between comparisons of genetic distance divided into a higher genetic distance range (0.23-0.25, females and males) and a lower range (0.15-0.18) is shown for both of the sexes with no comparisons observed in between (Fig. XXab).

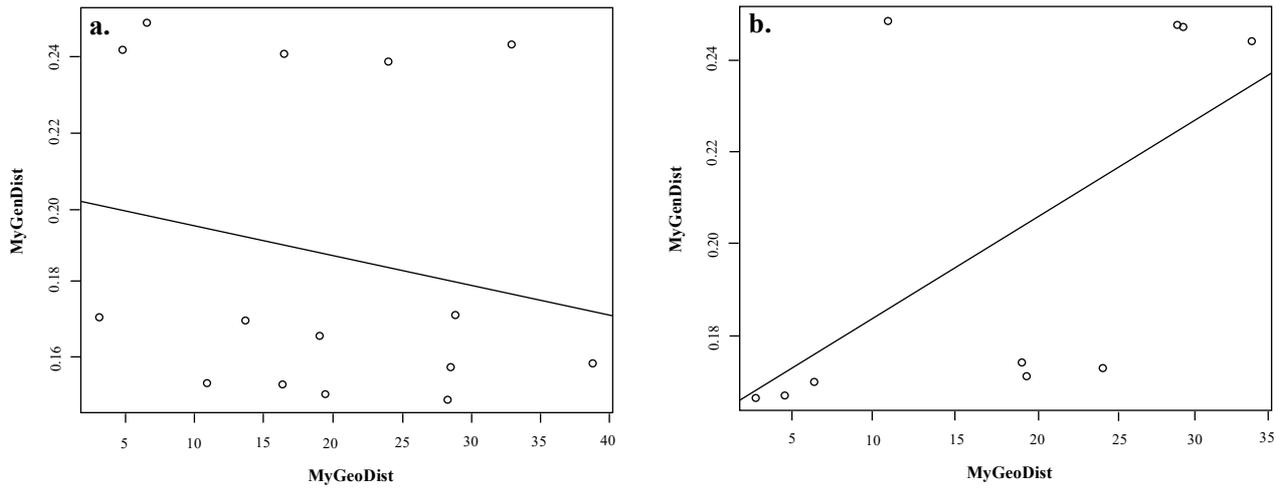


Figure XX. Mantel correlation plot comparing the genetic distance and geographic location of individuals in the South-cluster ($n=70$). a) Comparisons of female individuals only ($n=41$). b) Comparison of male individuals only ($n=29$). Circles indicate points of comparison (GenDist/GeoDist) based on the number of locations each sex was sampled in (F $n=6$, M $n=5$).