CCSBT-ESC/1909/08
(ESC Agenda item 9.1)

OCEANS AND ATMOSPHERE
www.csiro.au

Update on the SBT close-kin tissue sampling, processing, and kin finding

Jessica Farley, Paige Eveson, Mark Bravington, Peter Grewe

CCSBT-ESC/1909/08

Prepared for the Extended Scientific Committee for the Twenty Fourth Meeting of the Scientific Committee, Cape Town, South Africa, 2-7 September 2019

Citation

Farley J, Eveson P, Bravington M, Grewe P (2019). Update on the SBT close-kin tissue sampling, processing, kin-finding and long term sample storage. CCSBT-ESC/1909/08, Twenty Fourth Meeting of the Scientific Committee, Cape Town, South Africa, 2-7 September 2019.

Acknowledgments

The SBT sampling program in Indonesia would not be possible without the significant corporation of the the scientist team at Research Institute for Tuna Fisheries (Bali), and in particular that of Mr Kiroan Siregar and Mr Rusjas Mashar and other staff involved in measuring SBT and collecting tissue (and otolith) samples in Indonesia. The cooperation of the longline tuna industry (coordinated through Asosiasi Tuna Longline Indonesia) and the individual processing companies in providing access and facilities to carry out the sampling is much appreciated. We also thank Adam Kemp at Seatec Pty Ltd for sampling juveniles in Australia, and the Australian Southern Bluefin Tuna Industry Association (ASBTIA) for their ongoing support. Finally, we thank Rasanthi Gunasekera for coordinating the sample processing in the laboratory. This work was funded by the Commission for the Conservation of Southern Bluefin Tuna and CSIRO Oceans and Atmosphere.

Contents

1 Abstract

Muscle tissue samples were collected from SBT landed by the Indonesian longline fishery in Bali, Indonesia (adults; n=1500) and from harvested SBT at tuna processors in Port Lincoln, Australia (juveniles; n=1600) in 2018/19. Samples collected in Indonesia are stored at -20°C at the RIMF facility during the harvest season (Sep-Apr). They will be transported frozen to Hobart and held at -20°C until they are processed.

Muscle samples from the 2017/18 season were subsampled and the DNA subsequently extracted. A portion of the DNA was sent to DArT for genotype sequencing. The remaining tissue and extracted DNA samples were moved to -80°C archive freezer, where they currently remain.

DNA extracts from the 2016/17 muscle tissue samples selected for genotyping (Farley et al. 2018) were processed by DArT and the genotype data sent to CSIRO in early 2019. The kin-finding analyses to identify parent-offspring pairs (POPs) and half-sibling pairs (HSPs) were updated to include these data, and the identified POPs and HSPs were provided to the CCSBT in April 2019. Improvements were made this year to the procedure used for genotype calling to deal with inconsistencies in loci performance between different DArTcap sequencing batches, as encountered the past two years. Furthermore, substantial improvements were made to our kin-finding processes to ensure that false-positive kin pairs do not become a problem. To date, a total of 82 POPs and 167 “high confidence” HSPs have been identified, with the false negative rate for HSPs estimated to be 0.16.

2 Introduction

In 2013, the Extended Scientific Committee (ESC) developed a new Scientific Research Plan for southern bluefin tuna (SBT). The specific projects and priorities for the SRP were considered in 2014 and again in 2015. Several items were identified as high priority in the work plan including the continued collection and genotyping of tissue samples for ‘close-kin mark recapture’ genetics to assess the abundance of adult southern bluefin tuna (SBT). The CCSBT has funded the collection and archiving of SBT muscle tissue (since the 2014/15 season) and DNA extraction & sequencing of the tissue samples (since the 2015/16 season). These samples and data subsequently contributed to the completion of a second CKMR abundance estimation project that incorporated both POP and HSP which was reported to the ESC in 2018 (Davies et al 2018). Since 2018, the CCSBT have also funded the analysis of the sequencing data to find parent-offspring and half-sibling pairs in the samples (close kin identification). Table 1 shows the work undertaken in each project since 2015. In this paper we provide an update on progress of activities in 2019.

Table 1. Summary of SBT close-kin work undertaken each year since 2015. For the genotyping and kin-finding analysis, the season in which the fish were sampled is given.

Project	Muscle tissue collection	DNA extraction & genotyping	Close kin finding	ESC paper
2015	2014/15	NA[1]	NA[1]	CCSBT-ESC/1509/15
2016	2015/16	2014/15	NA[1]	CCSBT-ESC/1609/08
2017	2016/17	2015/16	NA[1]	CCSBT-ESC/1708/09
2018	2017/18	2016/17	2015/16	CCSBT-ESC/1809/08
2019 (current project)	2018/19	2017/18	2016/17	Current paper

[1] Genotyping & close kin finding undertaken in FRDC project 2016-044 (see Bravington et al. 2017; Davies et al 2018).

3 Muscle tissue collection

In Indonesia, targeted sampling of SBT occurred at Benoa Fishing Port in the 2018/19 spawning season using the existing Indonesia-CSIRO monitoring system for the longline fishery (e.g. see Proctor et al, 2006). Length measurements and muscle tissue samples were obtained for 1500 SBT ranging from 134-209 cm fork length (FL). The same fish are also sampled for otoliths (see Sulistyaningsih et al, 2019).

In Australia in June-July 2019, muscle tissue samples were collected from juvenile SBT at the tuna processors during harvest in Port Lincoln, South Australia. Tissue was obtained from 1600 fish ranging from 98 to 109 cm FL to ensure the full size range of 3 year-olds is being sampled. The muscle tissue was frozen according to protocols provided by CSIRO.

The frozen muscle samples are stored frozen in consecutively labelled boxes with 100 positions (10 by 10) in each box (A01 through J10). Individual sample are given a unique identification label (e.g., SbPL2014_Bx01_A01) and are stored in -80°C freezers at CSIRO laboratories in Hobart.

4 Close kin genotyping

A total of 2024 muscle tissue samples collected in the 2017/18 season were selected for genotyping. Of these, 1012 were from fish caught by the Australian surface fishery in the Great Australian Bight (juveniles) and 1012 from fish caught by the Indonesian longline fishery and landed in Bali, Indonesia (adults).

The samples from Australia were selected for analysis based on fish length. Of the 1012 selected, all samples from lengths 98-106 cm FL were selected and 100 and 53 were randomly selected from the 107 and 108 cm length classes (Figure 1). This selection process reduced the chance that 4 yo fish were selected for analysis.

The samples for genotying from from Indonesia were selected based on size of fish. Only fish ≥150 cm FL were included to avoid potential of including immature fish. All large fish ≥175cm were

selected plus additional fish selected randomly within each of the remaining 1-cm length class (150 -209 cm) up to a total sample size of 1012 fish.

DNA was extracted from a 10mg sub-sample of tissue for all fish. For most samples, a magnetic bead-based extraction protocol (Machery Nagel Nucleomag) kit was used on an Eppendorf EP motion robot to produce a 150uL archive and 50uL working stock of DNA in micro-titre format plates.

Archive plates of extracted DNA are stored in dedicated -80'C freezers located at CSIRO Hobart. Working stock plates of extracted DNA were shipped to Diversity Arrays Technology (DArT) in Canberra for genotype sequencing, referred to as "DArTcap", of approximately 2000 single nucleotide polymorphic loci (SNPs). When completed, the sequencing information will be transmitted to CSIRO Hobart.

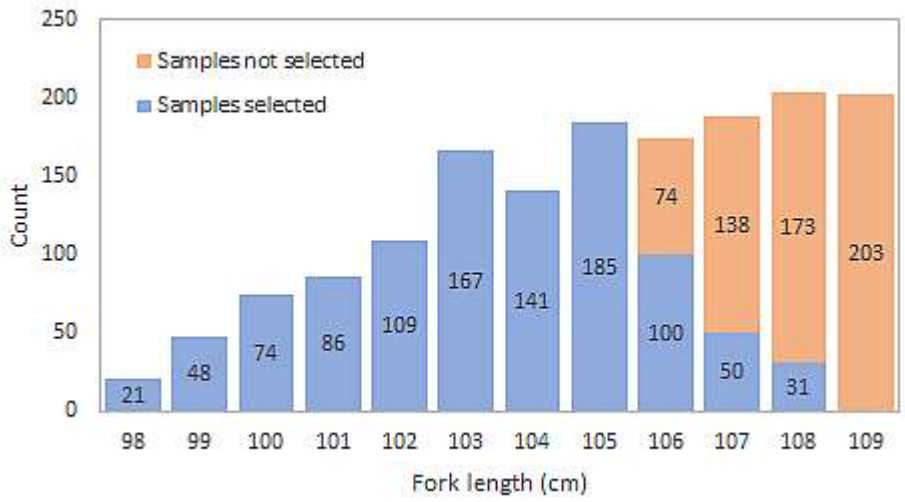

Figure 1. Length frequency of SBT selected for close-kin mark recapture genotyping from the 2017/18 samples collected in Australia Port Lincoln.

5 Close kin finding

DNA extracts from the 2016/17 muscle tissue samples selected for genotyping (Farley et al. 2018) were processed by DArT and the genotype data sent to CSIRO in early 2019. The kin-finding analysis database used for identification of parent-offspring pairs (POPs) and half-sibling pairs (HSPs) was updated to include the 2016/17 data.

Prior to kin-finding, we had to "call the genotype" for each fish and locus in the new data (i.e., use the sequencing data to infer the pair of alleles present). This genotype-calling entails quite complicated algorithms developed by CSIRO specifically for DArTcap sequencing data, and also estimates the genotyping error-rates for each locus, which is important in the identification of half-sibling pairs. We noted last year in Farley et al. (2018) that "the DArTcap data added to the analysis this year was not entirely consistent with the previous data (e.g., the sequence counts for some loci were significantly higher or lower on average than before). […] As a consequence, some

modification to the genotype-calling process was required (which is being investigated further)." The DArTcap data for 2016/17 had the same issue. After further investigation, we concluded that the differences in loci performance were largely at a plate level (each plate of 96 samples sent to DArT gets processed in the exact same manner). As such, we applied a plate-level standardization to the sequence count data from all years before calling the genotypes. This ensured that, for a given loci, the average count across all samples on a plate was the same for every plate; this greatly improved the genotype calls, as evidenced by QC checks and statistical tests of observed versus expected genotype frequencies.

Similar to past years, a series of quality control (QC) steps were applied to the genotyped data to remove fish with unreliable genotype calls; this includes a test of whether a fish has an unexpectedly high number of heterozygous loci, which could be an indication of cross-contamination of DNA between individuals. This is an evolving process, and some tweaks were made to thresholds used in the QC steps to improve performance. Note that, unlike last year, the same QC criteria were applied for both POP- and HSP-finding since improvements we made to the genotype-calling process meant that the more stringest criteria used for HSP-finding last year was no longer necessary. After applying the QC steps to the entire dataset, 6,269 adults and 13,056 juveniles remained for kin-finding (Table 2), noting that only the juveniles are used in identifying HSPs.

Table 2. Number of fish used in the kin-finding analyses this year after quality control (QC) checks were applied (note that the same QC checks were used in both POP- and HSP-finding this year). For the adults, samples were collected from Indonesia in the fishing season ending in the year shown (i.e., samples collected over the 2005/06 fishing season are referred to as year 2006).

Year	Adults	Juveniles
2006	0	1317
2007	0	1325
2008	0	1356
2009	0	1347
2010	972	1315
2011	958	963
2012	536	876
2013	959	903
2014	922	899
2015	0	953
2016	951	854
2017	971	948
Total	6269	13,056

POP-finding

We used the genotype data to identify POPs using the same method as in 2017, which is a modified Mendelian-exclusion statistic referred to as the Weighted-PSeudo-EXclusion (WPSEX) statistic (see Appendix B of Bravington et al. 2017). Figure 2 shows part of the histogram of the WPSEX statistic, across all genotyped adult-juvenile pairs (13,056 juveniles x 6,269 adults = 81.8 million comparisons). The POPs are visible as a small bump on the left side, and are clearly separated from non-POPs. Most of the histogram (to the right) has been truncated, because otherwise the POPs are too few compared to the gigantic bump of unrelated pairs (the peak of which is around 0.116 – exactly where theory predicts it should be based on allele frequencies of each locus) and could not be visualized. The giant bump drops off very quickly to the left of ~0.08, and the flattish tail around 0.055-0.075 will contain a number of adult/juvenile HSPs or grandparent-grandoffspring pairs, which should be somewhat rarer than POPs on demographic grounds.

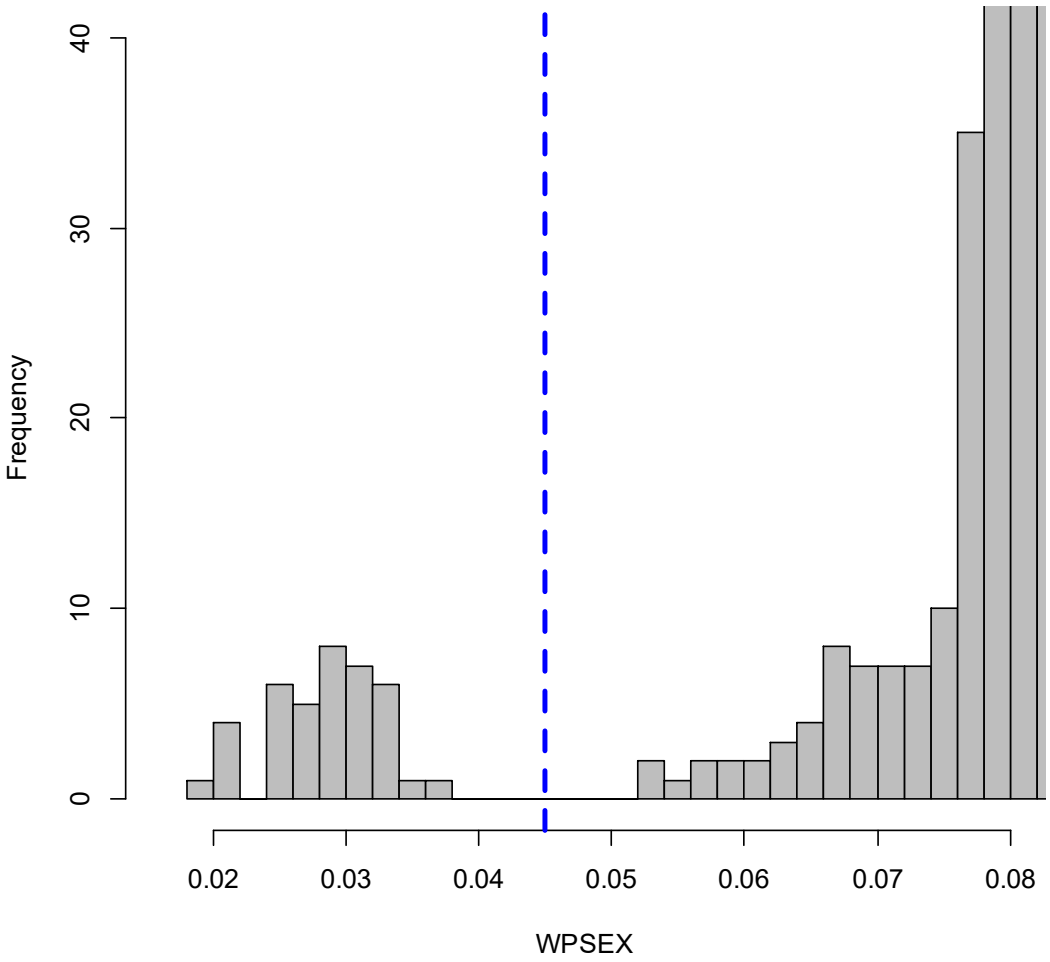

Figure 2. Histogram of the weighted-pseudo-exclusion (WPSEX) statistic for identifying parent-offspring-pairs (POPs). Low values (below the vertical blue dashed line) indicate POPs. The x-axis is right-truncated to omit the gigantic peak of unrelated pairs to the right.

The number of POPs identified in this data set is 39. Including the POPs that were identified previously using microsatellites (recall that the genotyping method changed after 2015 from using microsatellites to DArTcap sequencing; see Bravington et al. 2015, 2017), we now have a total of 82 pairs[1]. The breakdown by juvenile birth year and adult capture year is given in Table 3.

Table 3. Number of POPs (including those identified using microsatellites and DArTcap data) broken down by juvenile birth year (rows) and adult capture year (columns).

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2016	2017
2002	0	0	0	0	0	NA	NA	NA	NA	NA	NA
2003	0	5	1	2	0	0	0	1	0	1	0
2004	0	2	0	0	3	0	0	0	0	0	0
2005	1	4	5	4	1	0	0	1	2	0	0
2006	NA	4	3	2	0	0	0	0	0	0	0
2007	NA	NA	3	4	1	3	2	0	2	0	1
2008	NA	NA	NA	NA	0	1	1	1	0	0	0
2009	NA	NA	NA	NA	0	1	1	1	0	0	0
2010	NA	NA	NA	NA	0	0	1	4	0	2	0
2011	NA	NA	NA	NA	0	0	1	2	1	2	0
2012	NA	NA	NA	NA	0	0	0	1	1	0	0
2013	NA	NA	NA	NA	0	0	0	0	0	1	1
2014	NA	NA	NA	NA	0	0	0	0	0	0	1

HSP-finding

HSPs were identified using the same method as in 2017, which uses a pseudo-log-odds-ratio (PLOD) statistic to measure the relative probability of a pair of fish having their observed genotypes if they are HSPs compared to if they are unrelated. The details are provided in Appendix C of Bravington et al. (2017).

Among 13,056 juveniles included in the HSP-finding analysis (i.e., 13,056 *13,056 /2 = ~85.2 million pairwise comparisons), we found 167 that we are quite confident are HSPs (and 4 that are full-sibling-pairs (FSPs)) based on the PLOD test statistic (Figure 3). The observed PLOD distributions for unrelated pairs and HSPs match the predictions of genetic theory (Figure 3, left), which gives us confidence in using this statistic to identify HSPs. Unlike the WPSEX statistic for identifying POPs, the PLOD statistic does not give a clear separation between the bump for HSPs and that (to the left) for unrelated/less-related fish – and as total sample sizes increase, the potential for overlap between true HSPs and unrelated/less-related pairs becomes greater. Thus, we invested significant research effort to make improvements to our kin-finding processes, to ensure that false-positive kin pairs do not become a problem.

[1] Note that 45 pairs were identified using microsatellites and 39 identified using DArTcap data, which gives a total of 84, but 20 of these are the same because the 2010 adult samples were run using both microsatellites and DArTcap and we are finding 2 POPs involving an adult captured in 2010 using both data sets.

In addition to making improvements to the kin-finding process, we also developed a more statistical method to determine the PLOD value to use as the lower cut-off for HSPs (in previous years, this was done by visual inspection of the PLOD histogram). In brief, we made use of the theoretical means and approximate variances of the PLOD distributions for HSPs and unrelated/less-related pairs to choose a cutoff value that ensures that the number of false-positives is unlikely to be more than 1 or 2 (ie ensuring that any bias in the signal from HSPs about abundance is under 1%, based on 2-in-167) whilst maintaining a large enough number of HSPs for the estimate to have good precision. Using this method, we chose the lower cut-off value for HSPs to be PLOD=40 (Figure 4). An inevitable consequence of ensuring that false positives are rare is that a reasonable number of false negatives will be present; using the expected PLOD distribution for HSPs, we estimated the true number of HSPs to be about 16% higher than 167 because of false-negatives. The false-negative rate is allowed for in the population modelling, so is not a problem as long as we have a good estimate of it (Bravington et al. 2017). The division between PLOD values for HSPs and FSPs (the isolated individuals to the right of PLOD=150) was clear (Figure 3). The breakdown in numbers of identified HSPs by birth years is given in Table 4.

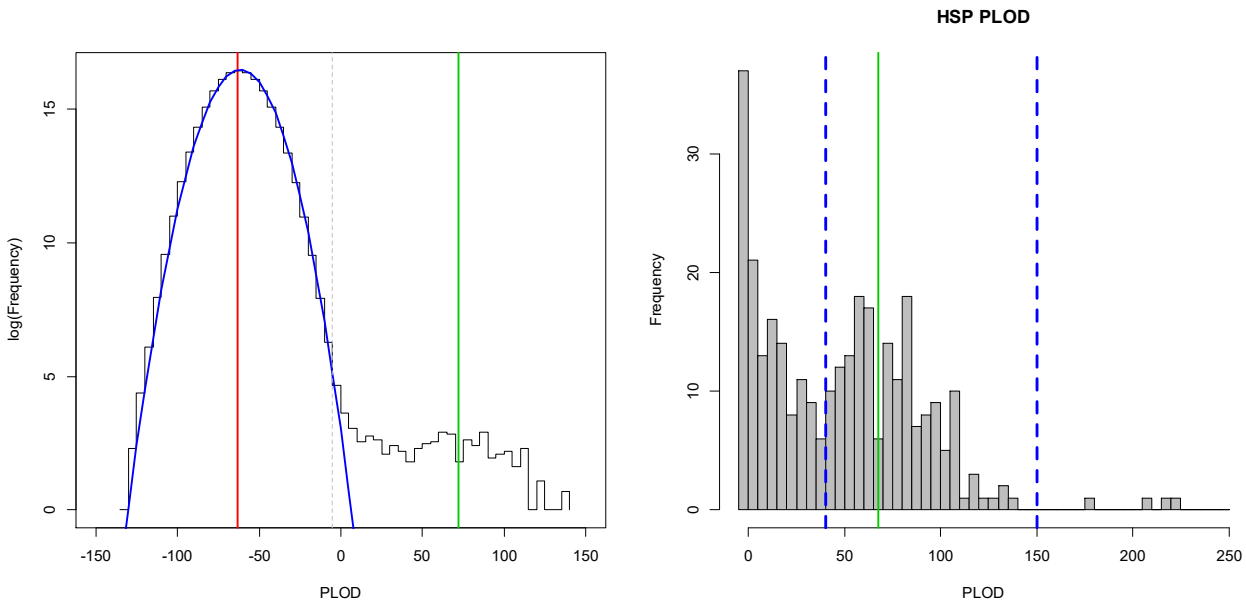

Figure 3. (left) Log histogram showing the pseudo-log-odds-ratio (PLOD) statistic for every pairwise comparison of juvenile SBT (~85.2 million comparisons). The solid blue line shows the theoretical distribution for unrelated pairs (UPs), and the red and green vertical lines are the theoretical means for UPs and HSPs respectively. (right) Histogram of PLOD values that are above -5. Values between the two vertical blue dashed lines indicate almost certain HSPs (see text and footnote 2). Higher values (>150) indicate full-sibling-pairs (FSPs), and lower values (<40) indicate unrelated and less-related pairs, but will also contain some false-negative HSPs (see text and footnote 2).

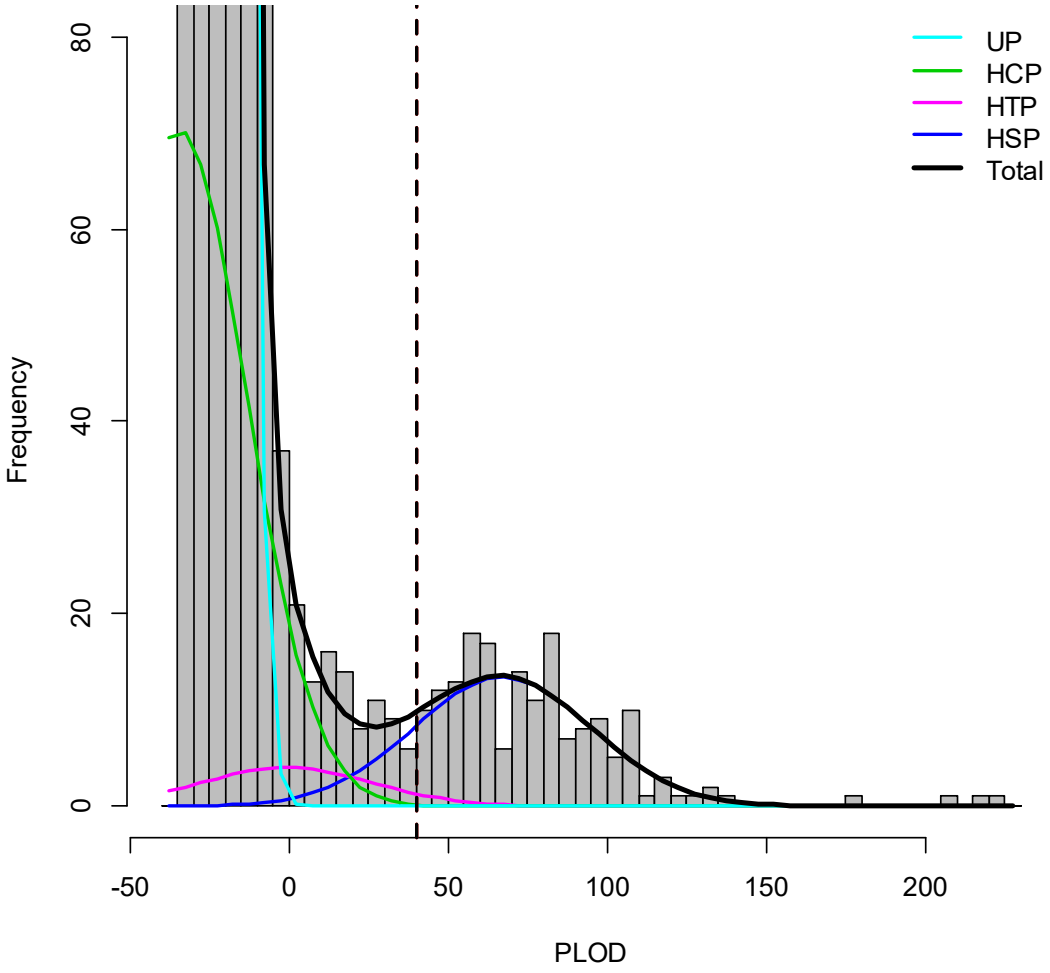

Figure 4. Approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs. By picking a lower cutoff of PLOD = 40 for HSPs, we expect no false-positive UPs or HCPs and a minimal number of false-positive HTPs.

Table 4. Number of HSPs broken down by birth year of younger sibling (rows) and older sibling (columns).

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
2003	2	4	4	3	0	2	1	0	0	2	0	2
2004		6	3	6	2	2	1	0	0	2	0	0
2005			5	4	3	4	0	5	1	1	0	2
2006				11	5	1	3	7	4	0	1	1
2007					3	5	3	2	2	2	2	1
2008						5	1	1	3	3	0	1
2009							1	2	1	0	0	0
2010								2	1	2	1	1
2011									3	3	2	0
2012										3	2	1
2013											2	4
2014												2

6 Summary

The project successfully completed:

1) 2018/19 tissue sampling in Australia and Indonesia (juveniles aand adults);
2) 2017/18 tissue subsampling and DNA extraction. DArT will complete the genotyping before the end of the project.
3) 2016/17 kin finding (POPs and HSPs).

An updated dataset of identified SBT parent-offspring pairs and half-sibling was provided to the CCSBT in April 2019. To date, a total of 82 POPs and 167 "high confidence" HSPs have been identified, with the false negative rate for HSPs estimated to be 0.16. Significant improvements were made this year to the procedures used for genotype calling and kin-finding to improve the consistency and accuracy of the genotype calls and to ensure that false-positive kin pairs do not become a problem in the future as sample sizes increase.

Although the total number of POPs is substantial, we note that there are rather few corresponding to recent juvenile cohorts (only 5 where the juvenile was born in 2012-2014). Thus there is not much direct information about adult stock size in those recent years. As the adult stock continues to rebuild, there will be even fewer "POPs per cohort per comparison" in future. Consequently, it may be necessary to increase annual sample sizes somewhat, in order to maintain robust and up-to-date information on adult stock size. The MP-testing process is a way to explore what sample sizes might be appropriate in future, but that investigation may take some time to settle down (since the MP process has other higher-priority issues to consider). As a common-sense precaution, we suggest it would be worthwhile increasing the annual number of genotypes for CKMR from current value of 2000 to around 3000, which is the number actually collected; the marginal cost of doing this should be quite small.

7 References

Baust, J. G. (2008). Strategies for the storage of DNA. *Cell Preservation Technology*, 6, 251-252.

Bonnet, J., Colotte, M., Coudy, D., Couallier, V., Portier, J., Morin, B., & Tuffet, S. (2009). Chain and conformation stability of solid-state DNA: implications for room temperature storage. *Nucleic acids research*, 38, 1531-1546.

Bravington M, Eveson P, Grewe P, Davies C. 2015. SBT Close-Kin Mark-Recapture: options for the medium term. CCSBT-ESC/1509/19.

Bravington MV, Eveson JP, Grewe PM, Davies CR. 2017. SBT close-kin mark-recapture with parent-offspring and half-sibling pairs: update on genotyping, kin-finding and model development. CCSBT-ESC/1709/12.

Davies CR, Bravington M, Bravington MV, Grewe PM, Eveson JP. 2018. Close-kin project Report. CCSBT-ESC/1809/14.

Farley J, Eveson P, Bravington M, Aulich J, Grewe P. 2018. Update on the SBT close-kin tissue sampling, processing, kin finding and long-term sample storage. CCSBT-ESC/1809/08.

Lee, S. B., Clabaugh, K. C., Silva, B., Odigie, K. O., Coble, M. D., Loreille, O., ... & Parsons, T. J. (2012). Assessing a novel room temperature DNA storage medium for forensic biological samples. *Forensic Science International,* Genetics, *6*, 31-40.

Nagy, Z. T. (2010). A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms, Diversity and Evolution,* 10, 91-105

Roder B, Fruhwirth K, Vogl C, Wagner M and Rossmanith P (2010).Impact of long term storage on stability of standard DNA for nucleic acid based methods. *Journal of Clinical Microbiology*, 48, 4260-4262

Shewale, J. G., & Liu, R. H. (Eds.). (2013). *Forensic DNA Analysis: Current Practices and Emerging Technologies*. CRC Press.

Smith, S., & Morin, P. A. (2005). Optimal storage conditions for highly dilute DNA samples: a role for trehalose as a preserving agent. *Journal of Forensic Science*, 50, JFS2004411-8.

Sulistyaningsih R, Farley J, Proctor C. 2019. Update on the length and age distribution of southern bluefin tuna (SBT) in the Indonesian longline catch. CCSBT-ESC/1909.

Wasko, A. P., Martins, C., Oliveira, C. and Foresti, F. (2003). Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas,* 138, 161–165

Yuanzheng Y and Angell C (2005). Clarifying the glass transition behaviour of water by comparison with hyperquenched inorganic glasses. *Nature*, 427, 717

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.csiro.au

AT CSIRO, WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

Our innovations contribute billions of dollars to the Australian economy every year. As the largest patent holder in the nation, our vast wealth of intellectual property has led to more than 150 spin-off companies.

With more than 5,000 experts and a burning desire to get things done, we are Australia's catalyst for innovation.

CSIRO. WE IMAGINE. WE COLLABORATE. WE INNOVATE.

FOR FURTHER INFORMATION

Oceans & Atmosphere
Jessica Farley
t +61 3 6232 5189
e Jessica.farley@csiro.au
w www.csiro.au

Oceans & Atmosphere
Paige Eveson
t +61 3 6232 5015
e **paige.eveson**@csiro.au
w www.csiro.au