CCSBT-ESC/2008/07
(ESC Agenda item 7.1)

OCEANS AND ATMOSPHERE
www.csiro.au

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

Jessica Farley, Paige Eveson, Rasanthi Gunasekera, Peter Grewe, Mark Bravington

CCSBT-ESC/2008/07

Prepared for the Extended Scientific Committee for the Twenty Fifth Meeting of the Scientific Committee, Online, 31 August - 7 September 2020

Citation

Farley J, Eveson P, Gunasekera R, Grewe P, Bravington M (2020). Update on the SBT close-kin tissue sampling, processing, kin-finding and long term sample storage. CCSBT-ESC/2008/07, Twenty Fifth Meeting of the Scientific Committee, Online, 31 August - 7 September 2020

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. We also thank Shane Baylis for work on the R package used for kin-finding ('kinference'). 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 2019/20. Samples collected in Indonesia are stored at -20°C 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 2018/19 season were subsampled and DNA extracted. A portion of the DNA was sent to DArT for genotype sequencing. The remaining tissue and extracted DNA samples were moved to a -80°C archive freezer, where they currently remain.

DNA extracts from the 2017/18 muscle tissue samples selected for genotyping (Farley et al. 2019) were processed by DArT and the genotype data sent to CSIRO in early 2020. The kin-finding analyses to identify parent-offspring pairs (POPs) and half-sibling pairs (HSPs) were updated to include these data using the same methods for genotype calling and kin-finding as last year, and the identified POPs and HSPs were provided to the CCSBT in April 2020. The total number of POPs to date is 89, and the total number of HSPs for which we have high confidence is 161, with a false negative rate estimated at 0.26. In order to keep the risk of false positives very low, we needed to increase the lower cut-off on the "PLOD" statistic used for determining HSPs. This has resulted in fewer HSPs, and a higher false negative rate, than last year. In future, we aim to make use of a genome assembly for SBT to improve the separation and "reclaim" some of the HSPs currently being excluded; however, in the meantime, the number of HSPs is sufficiently large to provide reliable information for the MP.

2 Introduction

In 2013, the Extended Scientific Committee (ESC) developed a new Scientific Research Plan (SRP) 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; 2020). 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 2019, the CCSBT agreed to increase the number of tissue samples genotyped from ~2000 to 3,100 (the number actually collected) to increase the number of "POPs per cohort comparison" (Anon 2019). In this paper we provide an update on progress of activities in 2020.

Table 1. Summary of SBT close-kin work undertaken as part of CCSBT projects 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	2018/19	2017/18	2016/17	CCSBT-ESC/1908
2020 (current project)	2019/20	2018/19	2017/18	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 2019/20 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 134-199 cm fork length (FL). The same fish are also sampled for otoliths (see Sulistyaningsih et al. 2020). Samples are stored at -20°C during the harvest season (Sep-Apr). They will be transported frozen to Hobart and held at -20°C until they are processed.

In Australia in June-July 2020, 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 98-109 cm FL to ensure the full size range of 3 year-olds is being sampled. The tissue samples were frozen according to protocols provided by CSIRO and will be transported frozen to Hobart and held at -20°C until they are processed.

The frozen muscle tissue samples are stored 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).

4 Close kin genotyping

DNA extracts from the 2017/18 muscle tissue samples selected for genotyping (Farley et al. 2019) were processed by DArT in 2019/20 and the sequencing data sent to CSIRO Hobart in early 2020.

Compared to previous years where a subsample of tissue samples were selected for genotyping, all tissue samples collected in the 2018/19 season were selected. Of these, 1599 were from fish caught by the Australian surface fishery in the Great Australian Bight (juveniles) and 1500 from fish

caught by the Indonesian longline fishery and landed in Bali, Indonesia (adults). The length distribution of fish genotyped is shown in Figure 1.

In previous years when fewer fish were genotyped, only fish ≥150 cm FL were selected from the Indonesian catch to avoid potential of including immature fish. Reproductive studies, however, have shown that all SBT samples from the Indonesian catch on the spawning ground are mature (Farley and Davis 1998, 2015; Hartaty et al. 2019; Hartaty and Zulkarnaen 2020); since the population model for the close kin data accounts for differential fecundity with size, inclusion of these small fish should not bias the results.

DNA was extracted from a 10mg sub-sample of tissue for all fish. For all samples, a magnetic bead-based extraction protocol (Machery Nagel Nucleomag) kit was used on an Eppendorf EP motion robot to produce a 90uL archive and 30uL 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 sequencing, referred to as "DArTcap", of approximately 2000 single nucleotide polymorphic loci (SNPs). When completed, the sequencing information will be transmitted to CSIRO Hobart.

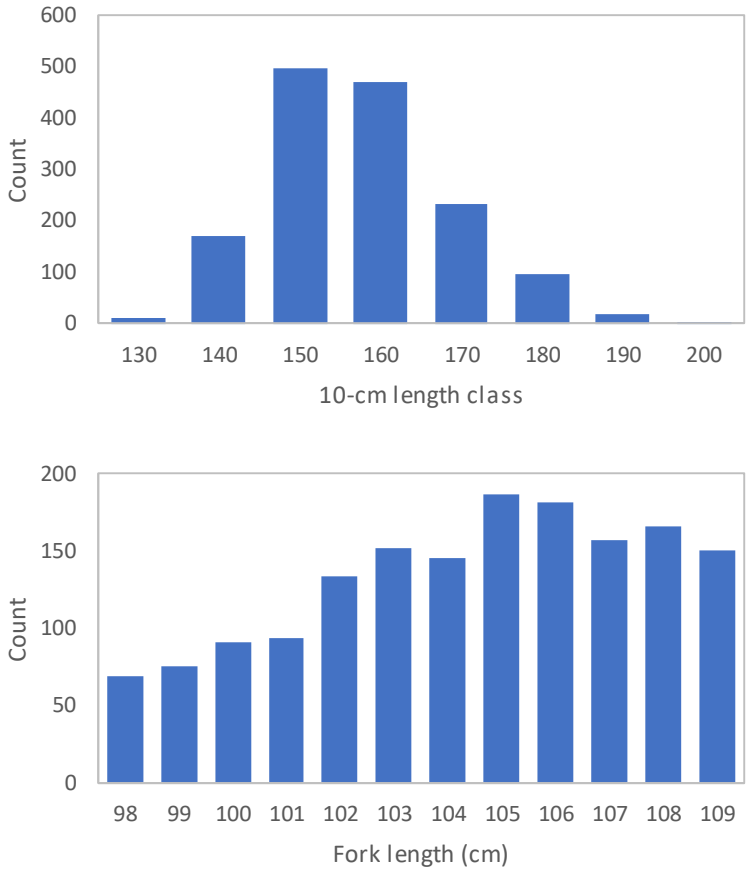

Figure 1. Length frequency of SBT selected for close-kin mark recapture genotyping from the 2018/19 samples collected in Indonesia (top) and Port Lincoln, Australia (bottom).

5 Close kin finding

The kin-finding analysis database used for identification of parent-offspring pairs (POPs) and half-sibling pairs (HSPs) was updated to include the 2017/18 data.

Prior to kin-finding, the sequencing data are used to "call the genotype" for each fish and locus in the data (i.e., 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 HSPs. Same as last year, a plate-level standardization was applied to the sequence count data from all years before calling the genotypes (see Farley et al. 2019). This ensured that, for a given loci, the average count across all samples on a plate was the same for every plate..

Similar to past years, a series of quality control (QC) steps were applied to the genotyped data to remove fish with unreliable genotype calls. These include: a test for heterogeneity to remove fish with an unexpectedly high number of heterozygous loci, which could be an indication of cross-contamination of DNA between individuals; a test of whether an individual's genotype could plausibly have been drawn from the 'stock' represented by the rest of the samples to remove fish potentially mis-identified as SBT; and a test for an over-representation of null alleles in each individual genotype to remove degraded samples. After applying the QC steps to the entire dataset, 6,969 adults and 13,833 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. 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
2018	700	777
Total	6969	13,833

POP-finding

We used the genotype data to identify POPs using the same method as the previous two years, 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 (13,833 juveniles x 6,969 adults = 96.4 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, 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 0.05-0.075 will contain a number of adult/juvenile HSPs or grandparent-grandoffspring pairs, which should be somewhat rarer than POPs on demographic grounds.

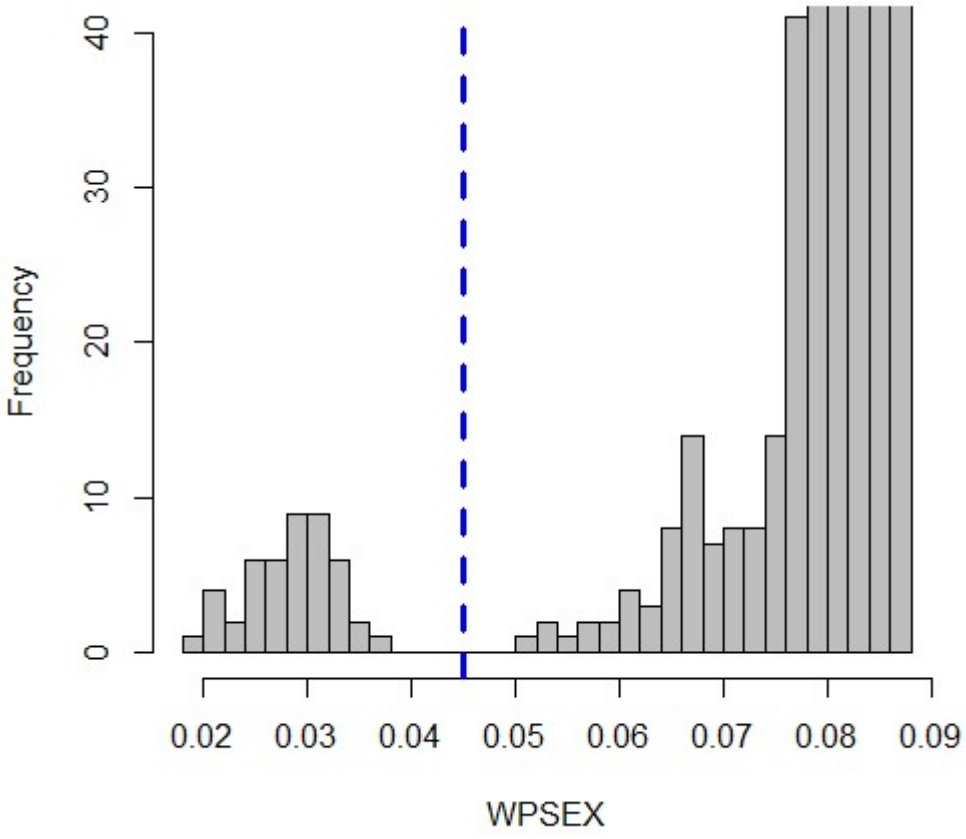

Figure 2. Histogram of the weighted-pseudo-exclusion (WPSEX) statistic for identifying parent-offspring-pairs (POPs). Low values (below the vertical blue dashed line) indicate POPs. The x-axis is right-truncated to omit the gigantic peak of unrelated pairs to the right.

The number of POPs identified in this data set is 46. 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 89 pairs. 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	2018
2002	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA
2003	0	5	1	2	0	0	0	1	0	1	0	0
2004	0	2	0	0	3	0	0	0	0	0	0	0
2005	1	4	5	4	1	0	0	1	2	0	0	0
2006	NA	4	3	2	0	0	0	0	0	0	0	0
2007	NA	NA	3	4	1	3	2	0	2	0	1	0
2008	NA	NA	NA	NA	0	1	1	1	0	0	0	2
2009	NA	NA	NA	NA	0	1	1	1	0	0	0	0
2010	NA	NA	NA	NA	0	0	1	4	0	2	0	0
2011	NA	NA	NA	NA	0	0	1	2	1	2	0	0
2012	NA	NA	NA	NA	0	0	0	1	1	0	0	1
2013	NA	NA	NA	NA	0	0	0	0	0	1	1	3
2014	NA	NA	NA	NA	0	0	0	0	0	0	1	0
2015	NA	NA	NA	NA	0	0	0	0	0	1	0	0

HSP-finding

HSPs were identified using the same method as the previous two years, which is based on 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 13,833 juveniles included in the HSP-finding analysis (i.e., 13,833 *13,832 /2 = ~95.7 million pairwise comparisons), we found 161 that we are quite confident are HSPs (and four 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.

The division between PLOD values for HSPs and FSPs (the four isolated individuals to the right of PLOD = 150) is clear[1] (Figure 3). However, 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.

[1] Note that all four FSPs were within-cohort pairs, as one would expect for a large adult stock.

We used the same method described last year (Farley et al. 2019) to determine the PLOD value to use as the lower cut-off for HSPs. 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 cut-off value that ensures that the number of false-positives is unlikely to be more than 2 (=1.2% based on 161 potential HSPs), whilst maintaining as many HSPs as possible. Using this method, we determined the lower cut-off value for HSPs to be 50 (Figure 4). This cut-off is larger than the value of 40 used last year, but was necessary in order to keep the risk of false positives equally low.

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 26% higher than 161 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 breakdown in numbers of identified HSPs by birth years is given in Table 4.

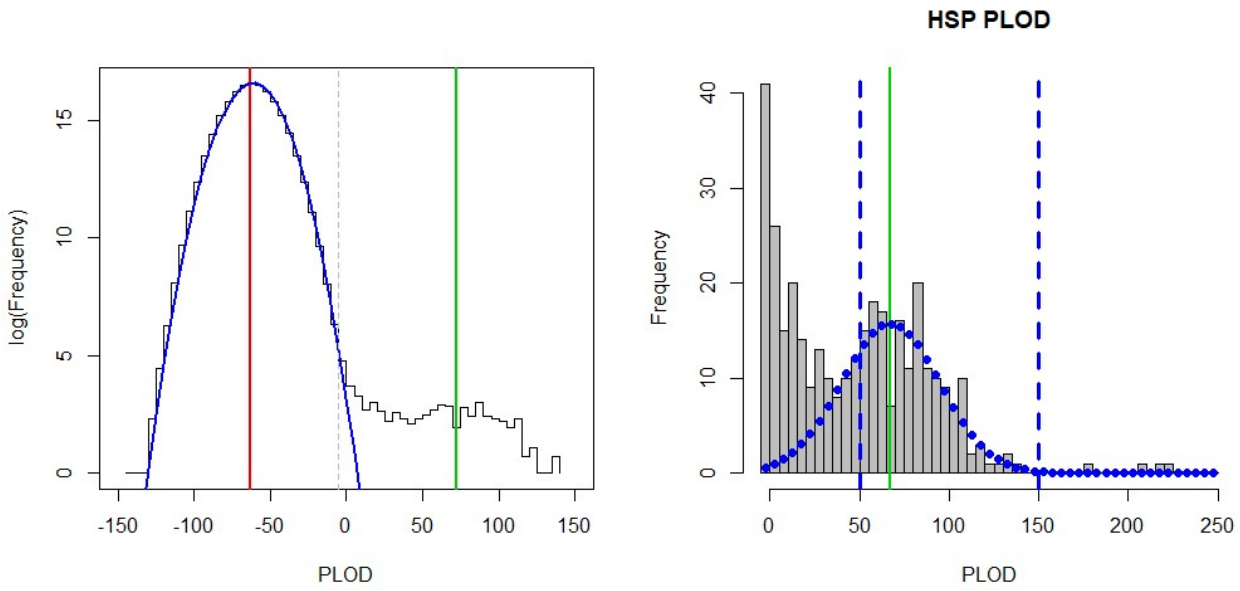

Figure 3. (left) Log histogram showing the pseudo-log-odds-ratio (PLOD) statistic for every pairwise comparison of juvenile SBT (~95.7 million comparisons). The solid blue line shows the theoretical distribution for unrelated pairs (UPs), and the red and green vertical lines are the theoretical means for UPs and HSPs respectively. (right) Histogram of PLOD values above -5. Values between the two vertical blue dashed lines indicate almost certain HSPs (see text). Higher values (>150) indicate full-sibling-pairs (FSPs), and lower values (<50) indicate unrelated and less-related pairs, but will also contain some false-negative HSPs (see text).

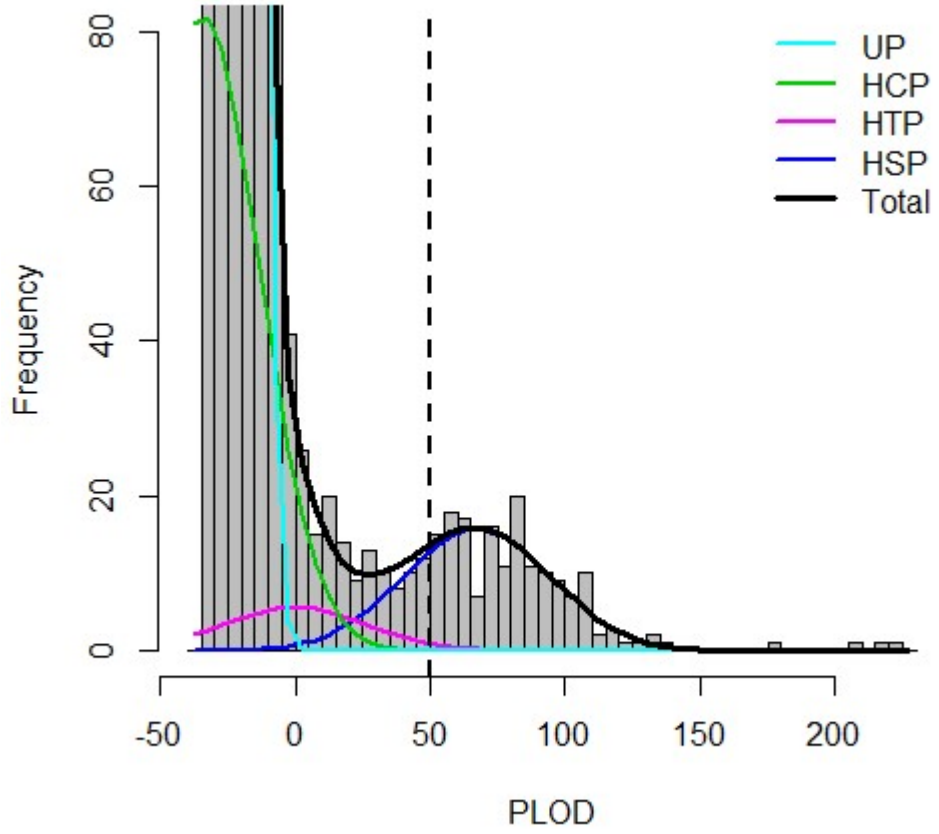

Figure 4. Approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs. By picking a lower cutoff of PLOD = 50 for HSPs, we expect no false-positive UPs or HCPs and a minimal number of false-positive HTPs.

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	2015
2003	2	4	2	1	0	0	1	0	0	2	0	2	1
2004		6	3	6	2	2	1	0	0	2	0	0	0
2005			5	3	3	3	0	5	1	1	0	2	0
2006				8	4	1	3	5	3	0	1	1	1
2007					3	3	2	2	2	2	2	1	2
2008						5	1	1	2	3	0	1	0
2009							1	2	1	0	0	0	0
2010								2	1	2	1	0	1
2011									3	2	1	0	3
2012										3	2	1	1
2013											2	4	1
2014												2	2
2015													4

6 Summary

The project successfully completed:

1) 2019/20 tissue sampling in Australia and Indonesia (juveniles aand adults);
2) 2018/19 tissue subsampling and DNA extraction. DArT will complete the genotyping before the end of the project.
3) 2017/18 kin finding (POPs and HSPs).

An updated dataset of identified SBT POPs and HSPs was provided to the CCSBT in April 2020. To date, a total of 89 POPs and 161 "high confidence" HSPs have been identified, with the false negative rate for HSPs estimated to be 0.26. As noted in past reports, the overlap between true HSPs and unrelated (or weakly-related) pairs will increase as the total sample sizes increase. Thus, in order to keep the risk of false positives very low, it was necessary this year to increase the lower cut-off on the "PLOD" statistic used for determining HSPs. This has resulted in fewer HSPs, and a higher false negative rate, than last year. In future, we aim to make use of a genome assembly for SBT to improve the separation and "reclaim" some of the HSPs currently being excluded; however, in the meantime, the number of HSPs is sufficiently large to provide reliable information for the MP.

7 References

Anonymous (2019). Report of the Twenty Sixth Annual Meeting of the Commission, Commission for the Conservation of Southern Bluefin Tuna, 17 October 2019, Cape Town, South Africa.

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.

Davies C, Bravington M, Bravington M, Eveson P, Lansdell, M, Aulich J, Grewe P. 2020. Next-generation Close-kin Mark Recapture: Using SNPs to identify half- sibling pairs in Southern Bluefin Tuna and estimate abundance, mortality and selectivity. Final Report, FRDC Project No 2016-044.

Farley JH, Davis TLO 1998. Reproductive dynamics of southern bluefin tuna, *Thunnus maccoyii.* Fishery Bulletin. 96(2): 223-236.

Farley JH, Davis TLO, Bravington MV, Andamari R, Davies CR. 2015. Spawning dynamics and size related trends in reproductive parameters of southern bluefin tuna, *Thunnus maccoyi*i. PLoS ONE 10(5): e0125744. doi:10.1371/journal.pone.0125744

Farley J, Eveson P, Bravington M, Aulich J, Grewe P. 2019. Update on the SBT close-kin tissue sampling, processing, kin finding and long-term sample storage. CCSBT-ESC/1909/08.

Hartaty H, Zulkarnaen F. 2020. Updated study of the reproductive activity of SBT caught in Indonesian tuna fisheries. CCSBT-ESC/2008/Info03.

Hartaty H, Fahmi Z, Farley J. 2019. Study of the reproductive activity of SBT caught in Indonesian tuna fisheries. CCSBT-ESC/1909/42.

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.

CONTACT US

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

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