Australia's National
Science Agency

Update on the SBT close-kin tissue sampling, processing and kin-finding 2021

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

CCSBT-ESC/2108/06

Prepared for the CCSBT Extended Scientific Committee for the Twenty Sixth Meeting of the Scientific Committee

23-31 August 2021

Acknowledgments

The SBT sampling program in Indonesia would not be possible without the significant corporation of 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

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 2020/21. Samples collected in Indonesia are stored at -20°C and will be transported frozen to Hobart, when COVID-19 restrictions allow. Note that muscle tissue samples collected last year (2019/20) are also currently in Benoa as COVID-19 travel restrictions prevented transportation of the frozen samples to Australia as planned. Therefore, it was not possible to extract DNA from the tissue for genotype sequencing. However, muscle samples from the 2019/20 season collected in Port Lincoln (juveniles) were subsampled and DNA extracted. A portion of the DNA was sent to DArT for genotype sequencing and the remaining tissue and extracted DNA samples were moved to a new dedicated close-kin -80°C archive freezer (funded by CCSBT), where they currently remain.

DNA extracts from the 2018/19 muscle tissue samples selected for genotyping last year (see Farley et al. 2020) were processed by DArT and the sequencing data sent to CSIRO in early 2021. 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 2021. The total number of POPs to date is 95, and the total number of HSPs for which we have high confidence is 174, with a false negative rate estimated at 0.25. In order to keep the risk of false positives very low (e.g., to minimise the number of less-related pairs, in particular half-thiatic pairs (HTPs), incorrectly identified as HSPs), we limited our HSP comparisons to pairs of juveniles born less than 9 years apart. This greatly reduces the number of comparisons between fish that are potentially HTPs (since HTPs are likely to be further apart in age), while not excluding too many potential HSPs. While this was an adequate solution for this year, in future, we will make use of a new genome assembly for SBT to improve the separation and "reclaim" some of the HSPs currently being excluded.

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 close kin mark recapture (CKMR) abundance estimation project that incorporated both parent-offspring pairs (POPs) and half-sibling pairs (HSPs), 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 POPs and HSPs in the samples (close-kin identification) on an annual basis. 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 2021.

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	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/1909/08
2020	2019/20	2018/19	2017/18	CCSBT-ESC/2008/07
2021 (current project)	2020/21	2019/20	2018/19	Current paper

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

Muscle tissue collection

In Indonesia, targeted sampling of SBT occurred at Benoa Fishing Port in the 2020/21 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. The same fish are also sampled for otoliths. Samples are stored at -20°C and will be transported frozen to Hobart, when COVID-19 restrictions allow. Note that muscle tissue samples collected in 2019/20 are also currently in Benoa as COVID-19 travel restrictions prevented transportation to Australia.

In Australia in June-July 2021, muscle tissue samples were collected from juvenile SBT at the tuna processors during harvest operations in Port Lincoln, South Australia. Tissue was obtained from 1600 fish ranging from 98-109 cm FL to ensure the complete 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).

DNA extraction and sequencing

Muscle tissue samples collected in Indonesia last year (2019/20 season) are currently in Benoa as COVID-19 travel restrictions prevented transportation of the frozen samples to Australia.

Therefore, it was not possible to extract DNA from the tissue for genotype sequencing as planned. However, all muscle samples collected last year in Port Lincoln (juveniles) were subsampled. DNA was extracted using a magnetic bead-based extraction protocol (Machery Nagel Nucleomag) kit on an Eppendorf EP motion robot to produce a 90uL archive and 30uL working stock of DNA in micro-titre format plates. 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. Archive plates of extracted DNA are stored in a dedicated -80°C freezer located at CSIRO Hobart.

DNA extracts from the 2018/19 muscle tissue samples selected for sequencing last year (see Farley et al. 2020) were processed by DArT and the sequencing data sent to CSIRO Hobart in early 2021.

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 2018/19 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. 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.

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, 8,409 adults and 15,261 juveniles remained for kin-finding (Table 2), noting that only the juveniles are used in identifying HSPs.

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 (15,261 juveniles x 8,409 adults = 128.3million 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.

The number of POPs identified in this data set is 52. 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 95 pairs. The breakdown by juvenile birth year and adult capture year is given in Table 3.

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	756
2019	1440	1449
Total	**8409**	**15261**

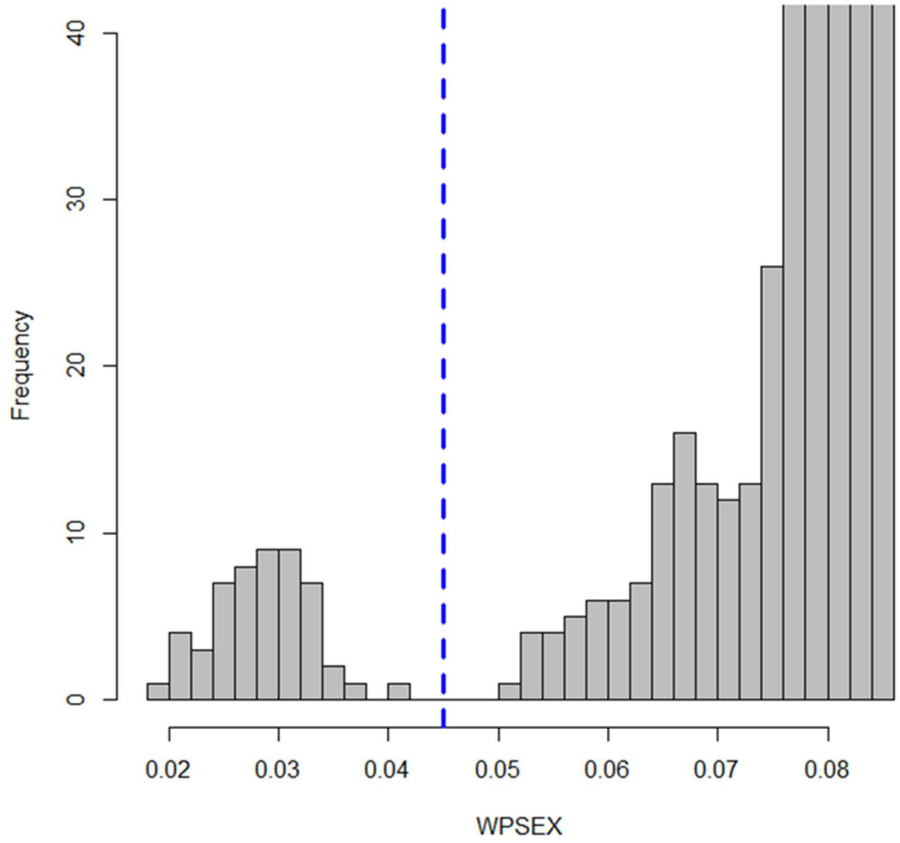

Figure 2. Histogram of the weighted-pseudo-exclusion (WPSEX) statistic for identifying parent-offspring-pairs (POPs). Low values (below the vertical blue dashed line) indicate POPs. The x-axis is right-truncated to omit the gigantic peak of unrelated pairs to the right.

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). NA indicates that no POPs were possible either because no samples exist for that combination of years, or the adult capture year is before the juvenile birth year.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2016	2017	2018	2019
2002	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
2003	0	5	1	2	0	0	0	1	0	1	0	0	0
2004	0	2	0	0	3	0	0	0	0	0	0	0	0
2005	1	4	5	4	1	0	0	1	2	0	0	0	0
2006	NA	4	3	2	0	0	0	0	0	0	0	0	0
2007	NA	NA	3	4	1	3	2	0	2	0	1	0	0
2008	NA	NA	NA	NA	0	1	1	1	0	0	0	2	0
2009	NA	NA	NA	NA	0	1	1	1	0	0	0	0	1
2010	NA	NA	NA	NA	0	0	1	4	0	2	0	0	1
2011	NA	NA	NA	NA	NA	0	1	2	1	2	0	0	0
2012	NA	NA	NA	NA	NA	NA	0	1	1	0	0	1	0
2013	NA	NA	NA	NA	NA	NA	NA	0	0	1	1	3	1
2014	NA	NA	NA	NA	NA	NA	NA	NA	0	0	1	0	0
2015	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	0	0
2016	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	2	1	0

HSP-finding

HSPs were again identified using 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).

The PLOD statistic for comparisons between all pairs of juveniles is shown in Figure 3. The division between PLOD values for HSPs and full-sibling pairs (FSPs) is clear, with the FSPs being the four pairs with PLOD values to the right of 150[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, in particular half-thiatic pairs (HTPs). As noted in previous years, the expected overlap becomes greater as the total sample size increases. This was apparent last year and meant that the PLOD value chosen as a lower cut-off for HSPs needed to be set at 50 (Farley et al. 2020), compared to 40 in the year before (Farley et al. 2019), in order to keep the risk of false positives equally low (<2). This year, when all pairs of juveniles are compared, setting the PLOD cut-off value at 50 resulted in 4 expected false positive HSPs, which we consider too high. To achieve less than 2 false positives would require setting the cut-off value at close to 60, and thus losing an unacceptably large percentage (~40%) of true HSPs. (See Farley et al. 2019 for the method used to determine the expected number of expected false positive, given a particular PLOD cut-off value.)

To deal with this, we limited the comparisons to juveniles born less than 9 years apart (or, equivalently, sampled less than 9 years apart since all juveniles are age 3 when sampled), where we chose 9 years after trialling several values. This greatly reduces the number of comparisons between fish that are potentially HTPs since HTPs are likely to be greater apart in age, thus reducing the size of the problematic HTP bump while not excluding too many potential HSPs.

The PLOD statistic for comparisons between juveniles born less than 9 years apart is shown in Figure 4. In this case setting a PLOD cut-off value of 50 for HSPs resulted in only 1.5 expected false positive HSPs, and 174 pairs that we are quite confident are HSPs. The breakdown in numbers of identified HSPs by birth years is given in Table 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 25% higher than 174 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). In order to better address the overlap issue between unrelated/less related pairs and HSPs in the future, we have generated a high-quality genome assembly for SBT in collaboration with the Wellcome Sanger Institute (UK). We are currently developing new algorithms leveraging linkage information between genetic markers to improve the accuracy of kin-finding.

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

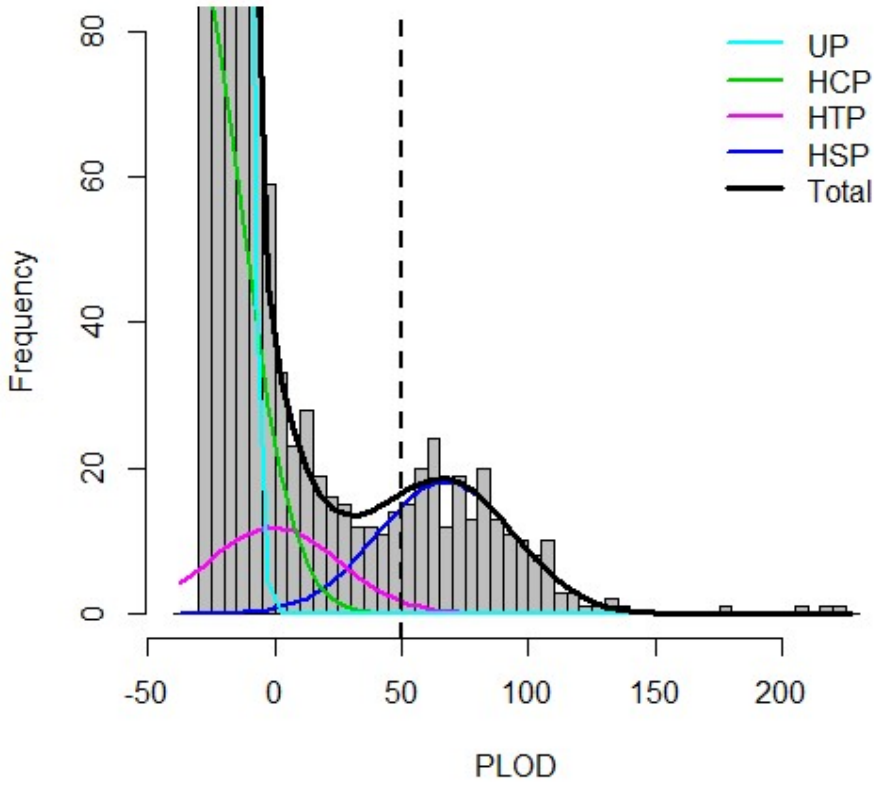

Figure 3. Histogram of the pseudo-log-odds-ratio (PLOD) statistic for all pairwise comparisons of juveniles. The approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs are shown. With a lower PLOD cut-off value of 50 for HSPs, we expect ~4 false-positive HTPs, which is higher than considered acceptable. (Note that the x-axis is left-truncated to omit the gigantic peak of UPs to the left.)

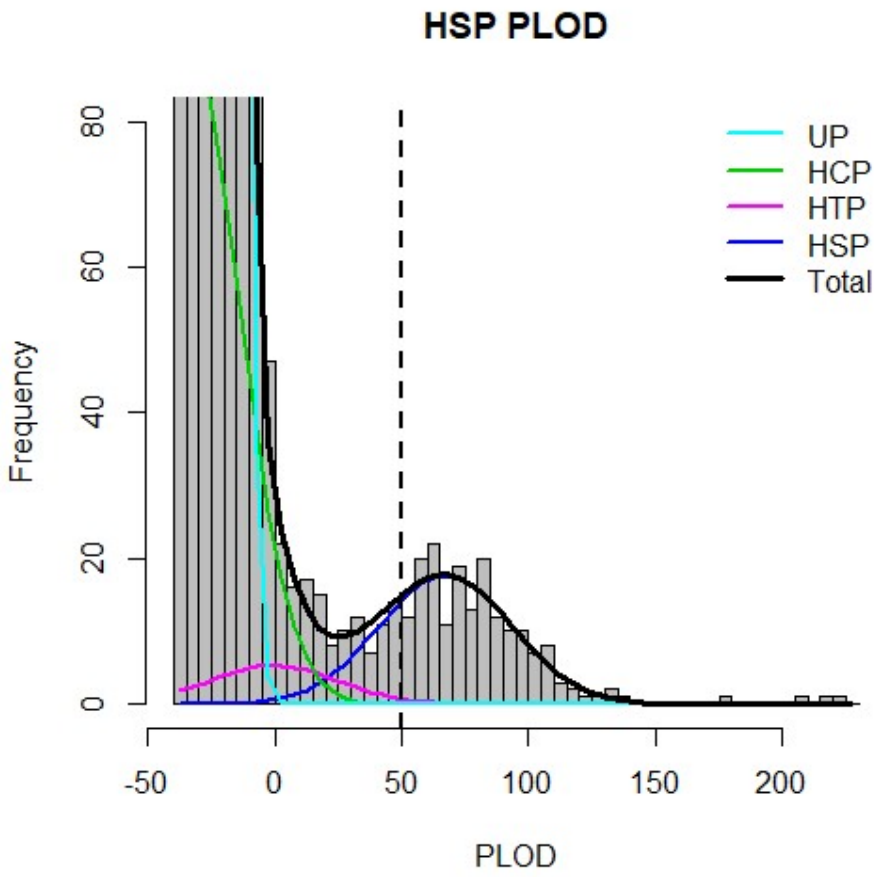

Figure 4. Histogram of the pseudo-log-odds-ratio (PLOD) statistic for pairwise comparisons of juveniles born less than 9 years apart. The approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs are shown. With a lower PLOD cut-off value of 50 for HSPs, we expect < 2 false-positive HTPs. (Note that the x-axis is left-truncated to omit the gigantic peak of UPs to the left.)

Table 4. Number of HSPs broken down by birth year of younger sibling (rows) and older sibling (columns). Note that comparisons were only made between juveniles born less than 9 years apart.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2003	2	4	2	1	0	0	1	0	0					
2004		6	3	6	2	2	1	0	0	2				
2005			5	3	3	3	0	5	1	1	0			
2006				8	4	1	3	5	3	0	1	1		
2007					3	3	2	2	2	2	2	1	2	
2008						5	1	1	2	3	0	1	0	2
2009							1	2	1	0	0	0	0	4
2010								2	1	2	1	0	1	1
2011									3	2	1	0	3	4
2012										3	2	1	1	1
2013											2	4	1	0
2014												2	2	1
2015													4	2
2016														6

Summary

The project successfully completed:

1) 2020/21 tissue sampling in Australia and Indonesia (juveniles and adults);
2) 2019/20 tissue subsampling and DNA extraction for Australian samples only (juveniles); DArT will complete the genotyping by the end of the project;
3) 2018/19 kin-finding (POPs and HSPs).

Unfortunately, COVID-19 travel restrictions prevented us from transporting the tissue samples collected last year in Indonesia to Australia for DNA extraction.

An updated dataset of identified SBT POPs and HSPs was provided to the CCSBT in April 2021. To date, a total of 95 POPs and 174 "high confidence" HSPs have been identified, with the false negative rate for HSPs estimated to be 0.25. As noted in past reports, the overlap between true HSPs and less-related pairs, in particular HTPs, continues to increase as the total sample size increases. Thus, in order to keep the risk of false positives very low, it was necessary this year to limit the HSP comparisons to juveniles born less than 9 years apart (see HSP-finding section for details). While this was an adequate solution for this year, in future, we will make use of a recently generated high-quality genome assembly for SBT to improve the separation and "reclaim" some of the HSPs currently being excluded.

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, Grewe PM, Eveson JP. 2018. Close-kin project report. CCSBT-ESC/1809/14.

Davies C, 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 J, Eveson P, Gunasekera R, Grewe P, Bravington M. 2020. Update on the SBT close-kin tissue sampling, processing, and kin finding. CCSBT-ESC/2008/07.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

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

For further information

Oceans and Atmosphere
Jessica Farley
+61 6 6232 5189
Jessica.farley@csiro.au
csiro.au

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