CCSBT-ESC/2409/09
(ESC Agenda item 7.1)

Australia's National
Science Agency

Update on the SBT close-kin tissue sampling, processing and kin-finding 2024

Jessica Farley, Paige Eveson, Rasanthi Gunasekera

CCSBT-ESC/2409/09

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

2-6 September 2024

Acknowledgments

The SBT sampling program in Indonesia would not be possible without the significant cooperation of the scientist team at the 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 over many years. 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 appreciate the support and corporation of the Research Centre for Fisheries, National Research and Innovation Agency (BRIN) and the Directorate General of Capture Fisheries (DGCF) for coordinating and undertaking sampling of adult SBT in Indonesia. Finally, we thank Adam Kemp at Seatec Pty Ltd for sampling juvenile SBT in Australia, and the Australian Southern Bluefin Tuna Industry Association (ASBTIA) for their ongoing support. We thank Scott Copper for helping with data base issues and 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 Environment.

Contents

Abstract

The Close-kin project is an on-going monitoring program that provides data on the adult component of the population for use in the Cape Town management procedure and stock assessment models.

Muscle tissue samples collected from harvested juvenile southern bluefin tuna (SBT) at tuna processors in Port Lincoln, Australia, in 2022 were subsampled, and DNA was extracted and sequenced. The kin-finding analysis to identify parent-offspring pairs (POPs) and half-sibling pairs (HSPs) was updated to include the new juvenile data. The results were provided to the CCSBT in June 2024. A total of 123 POPs and 232 high-confidence HSPs have now been identified, with a false negative rate for the HSPs of 0.25.

Sampling of juvenile SBT in 2023 and 2024 is complete.

As reported previously (CCSBT-ESC/2308/07), muscle tissue samples were not collected from the Indonesian longline fishery in 2021/22 and only 148 SBT were sampled in 2022/23 due to disruptions caused by institutional changes in Indonesia. In January 2024, a training workshop was held in Benoa, Bali, for Enumerators from the Directorate General of Capture Fisheries (DGCF) on how to collect SBT muscle tissue and otolith samples, to enable the SBT monitoring and sampling in Benoa to recommence. Sampling recommenced in January 2024 and 236 muscle tissue samples were collected. A second meeting with the enumeration team and DGCF monitoring program managers was held in August 2024 to review the season's data and prepare for the upcoming sampling season, which we anticipate will recommence in September 2024.

Introduction

In 2013, the Extended Scientific Committee (ESC) developed a new Scientific Research Plan (SRP) for southern bluefin tuna (SBT). High priority items identified in the work plan included the continued collection and genotyping of tissue samples for close-kin mark-recapture (CKMR) genetics to assess the abundance of adult SBT. The CCSBT has funded the collection and archiving of SBT muscle tissue since 2014/15, and DNA extraction and sequencing of these tissue samples since 2015/16. These samples and data have subsequently contributed to the completion of a second 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 has also funded the analysis of sequencing data to find POPs and HSPs (close-kin identification) on an annual basis, and this was identified as an on-going monitoring program when establishing the 2023-2027 SRP in 2022 (see Anon, 2022, ESC report Attachment 8). Table 1 shows the work completed by spawning season since 2014/15. In 2019, the CCSBT agreed to increase the number of tissue samples genotyped from ~2000 to 3100 annually (i.e., the total number of adults and juveniles collected annually) to increase the number of "POPs per cohort comparison" (Anon 2019). In this paper we provide an update on progress of activities in 2023.

Table 1. Summary of SBT close-kin work undertaken as part of CCSBT projects since 2015.

Sampling season	No samples collected Adults/Juveniles	Genotyping completed	Kin-finding completed
2014/15	1500/1600	Yes	Yes
2015/16	1500/1600	Yes	Yes
2016/17	1500/1600	Yes	Yes
2017/18	1500/1600	Yes	Yes
2018/19	1500/1600	Yes	Yes
2019/20	1500/1600	Yes	Yes
2020/21	1500/1600	Yes	Yes
2021/22	0/1600	Yes	Yes
2022/23	148/1600	Planned for 2024 (juveniles and a small number of adults)	Planned for 2025
2023/24	236/1600	Planned for 2025 (juveniles and a small number of adults)	Planned for 2026

Muscle tissue collection

Targeted sampling of adult SBT was reduced again at the Benoa Fishing Port in Indonesia in the 2023/24 spawning season due to disruptions caused by institutional changes in Indonesia. In January 2024, a training workshop was held for DGCF Enumerators in Benoa on how to collect SBT muscle tissue and otolith samples to enable the monitoring and sampling in Benoa to recommence (Figure 1). Sampling recommenced in January 2024 and 236 muscle tissue samples were collected. We anticipate full sampling will recommence in September for the 2024/25 spawning season (1500 samples) and we propose to collect additional muscle tissue samples to compensate for the lack of Indonesian muscle tissue sampling in the previous two seasons. See paper CCSBT-ESC/2409/12 for more information on the CCSBT SRP project to rebuild the catch monitoring program in Benoa.

The target number of samples from juveniles was met in 2024 with 1600 samples collected at the tuna processors during harvest operations in Port Lincoln, South Australia, in April-July. Tissue was obtained from fish ranging from 89-105 cm fork length (FL) to cover the size range of 3-year-olds (see Preece et al. 2024). The tissue samples were placed in 2ml tubes and frozen according to protocols provided by CSIRO and will be transported to Hobart. The frozen muscle tissue samples are stored in consecutively labelled boxes with 100 positions (10 by 10) in each box (A01 through J10). Individual samples are given a unique identification label (e.g., SbPL2014_Bx01_A01).

Figure 1. Trained DGCF Enumerators preparing to collect SBT muscle tissue and otolith samples at processors in Benoa, Bali.

DNA extraction and sequencing

Muscle tissue samples from juveniles collected in Port Lincoln (Australia) in 2022were subsampled and DNA extracted. The 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). Archive plates of extracted DNA are stored in a dedicated -80°C freezer located at CSIRO Hobart. All sequencing data were sent to CSIRO Hobart in early 2024 for inclusion in the kin-finding (below). There were no new adult data because very few adults were sampled in 2023/24 in Indonesia and the samples are currently in Benoa. These samples will be sequenced and included in kin-finding in 2025.

Kin-finding

The kin-finding analysis database used for identification of POPs and HSPs was updated to include the DArT sequencing data for juvenile samples collected in 2022.

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 complicated algorithms developed by CSIRO specifically for DArTcap sequencing data and 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 (Farley et al. 2019). This ensured that, for a given locus, the average count across all samples on a plate was the same for each 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, 11,261 adults and 19,333 juveniles remained for kin-finding (Table 2).

POP-finding

We used the genotype data to identify POPs using the same method as previous 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 comparing across all possible genotyped adult-juvenile pairs (19,333 juveniles x 11,261 adults = 217.7 million comparisons). The POPs are visible as a small bump to the left of the dashed blue line. 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 of

unrelated pairs drops off quickly to the left of ~0.075, and the "tail" between ~0.045-0.075 will contain a number of adult/juvenile HSPs and grandparent-grandchild pairs. The gap between this "tail" and the POPs bump has become less distinct over the years, which is to be expected given the number of comparisons has increased exponentially. In future, additional analyses may be required to help determine whether pairs in this uncertain zone between ~0.04-0.05 are POPs or less related pairs.

The number of POPs identified in this data set is 80. 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 123 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 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
2020	1421	1512
2021	1431	1384
2022	0	1176
Total	**11261**	**19333**

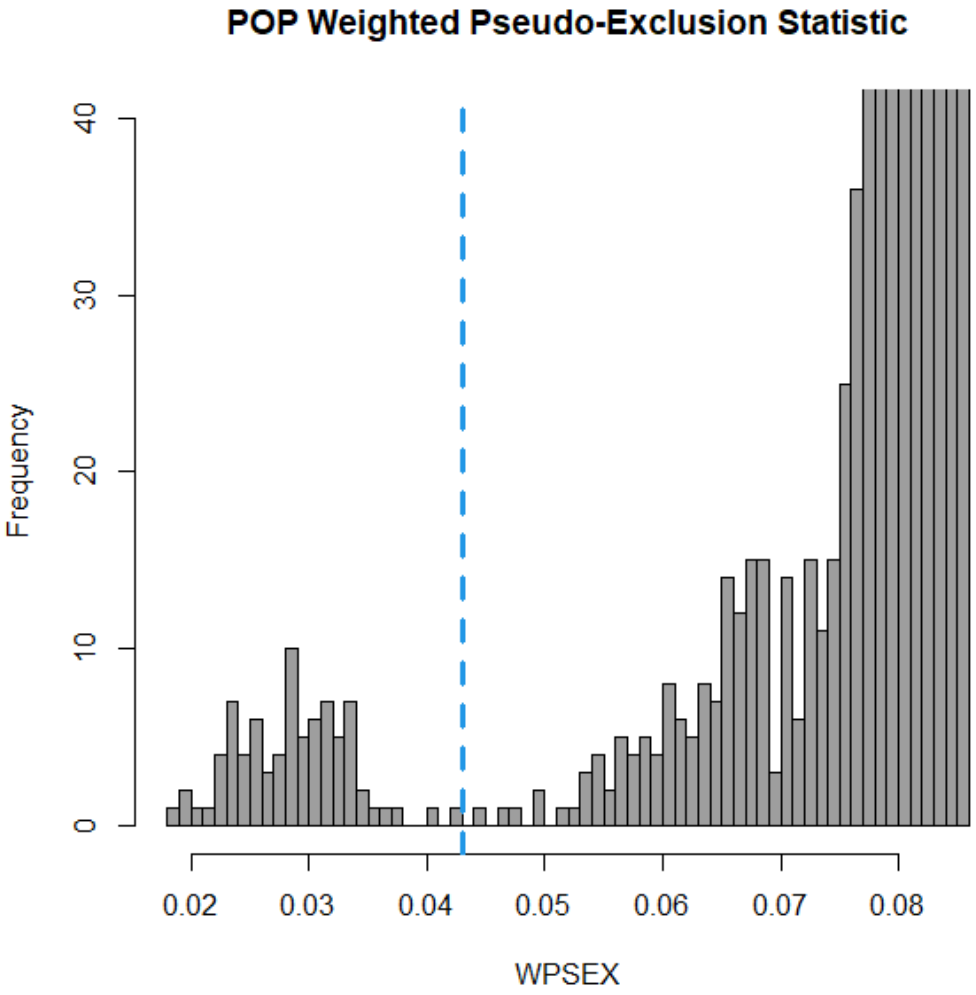

Figure 2. Histogram of the weighted-pseudo-exclusion (WPSEX) statistic for identifying parent-offspring pairs (POPs). Low values (below the vertical blue dashed line) indicate POPs. The x-axis is right-truncated to omit the gigantic peak of unrelated pairs to the right.

Table 3. Number of parent-offspring pairs (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	2020	2021
2002	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2003	0	5	1	2	0	0	0	1	0	1	0	0	0	0	0
2004	0	2	0	0	3	0	0	0	0	0	0	0	0	1	0
2005	1	4	5	4	1	0	0	1	2	0	0	0	0	0	0
2006	NA	4	3	2	0	0	0	0	0	0	0	0	0	0	2
2007	NA	NA	3	4	1	3	2	0	2	0	1	0	0	1	1
2008	NA	NA	NA	NA	0	1	1	1	0	0	0	2	0	0	0
2009	NA	NA	NA	NA	0	1	1	1	0	0	0	0	1	1	0
2010	NA	NA	NA	NA	0	0	1	4	0	2	0	0	1	0	0
2011	NA	NA	NA	NA	NA	0	1	2	1	2	0	0	0	0	0
2012	NA	NA	NA	NA	NA	NA	0	1	1	0	0	1	0	0	1
2013	NA	NA	NA	NA	NA	NA	NA	0	0	1	1	3	1	1	0
2014	NA	NA	NA	NA	NA	NA	NA	NA	0	0	1	0	0	1	0
2015	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	0	0	0	0
2016	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	2	1	0	1	1
2017	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	0	1	2	0
2018	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	2	3	0
2019	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	0	2

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 (see Appendix C of Bravington et al. 2017).

When using the PLOD statistic to compare all possible pairs of juveniles (19,333 x 19,333 = 373.8 million comparisons), we do not get a clear separation between the distribution for HSPs and that for unrelated/less related fish, in particular half-thiatic pairs (HTPs, e.g. uncle and half nephew). This has been noted as a problem in the past several years, and the overlap keeps increasing as the number of comparisons increases. We dealt with this overlap issue using the same approach as in 2021 and 2022 (e.g. Farley et al. 2022), limiting 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). Since HTPs are likely to be greater apart in age than 9 years, this reduces the number of comparisons between fish that are potentially HTPs, thus reducing the size of the problematic HTP bump while not excluding too many potential HSPs. To address this issue long term, a high-quality genome assembly for SBT was developed in 2021 in collaboration with the Wellcome Sanger Institute (UK). New algorithms are being developed that leverage linkage information between genetic markers gained from the genome assembly to improve the accuracy of kin-finding. These algorithms are currently being progressed but require further development and testing before applying to SBT.

The PLOD statistic for comparisons between juveniles born less than 9 years apart is shown in Figure 3. The four pairs with PLOD values to the right of 150 are full sibling pairs (FSPs)[1]. The division between HSPs and less related pairs is still not as clear as we would like however, setting a PLOD cut-off value of 50 for HSPs results in only 3 expected false positive HSPs and 232 pairs that we are quite confident are HSPs. Note that we used the theoretical means and approximate variances of the PLOD distributions for HSPs and unrelated/less related pairs (calculated as part of the kin-finding software) to determine the expected number of false positives. The breakdown in numbers of identified HSPs by birth year 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 232 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).

Table 4. Number of half-sibling pairs (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	2017	2018	2019
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	2		
2010								2	1	2	1	0	1	1	0	0	
2011									3	2	1	0	3	4	1	1	0
2012										3	2	1	1	1	2	1	0
2013											2	4	1	0	1	0	0
2014												2	2	1	4	3	1
2015													4	2	1	3	0
2016														6	4	5	6
2017															5	2	4
2018																5	3
2019																	4

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

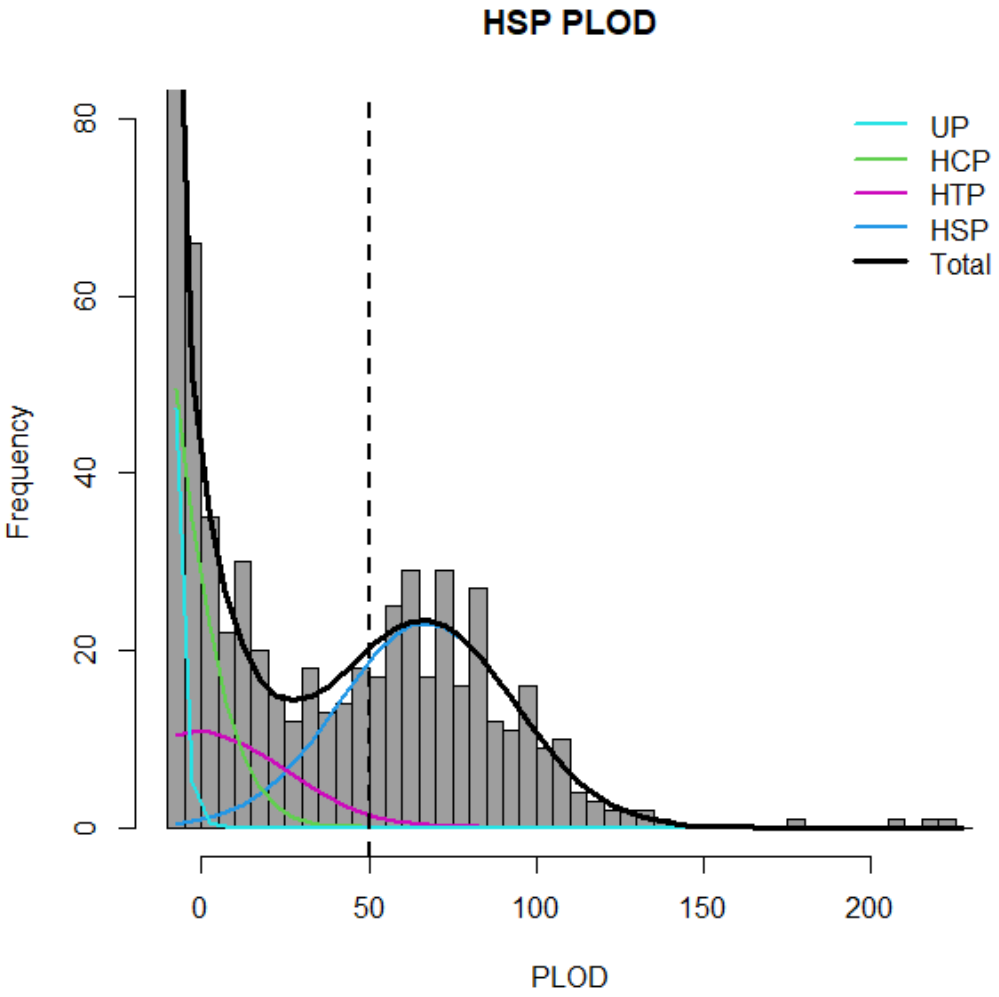

Figure 3. Histogram of the pseudo-log-odds-ratio (PLOD) statistic for pairwise comparisons of juveniles born less than 9 years apart. The approximate PLOD distributions for unrelated (UP), half-cousin (HCP), half-thiatic (HTP) and half-sibling (HSP) pairs are shown. With a lower PLOD cut-off value of 50 for HSPs, we expect ~3 false-positive HTPs. Note that the x-axis is left-truncated to omit the gigantic peak of UPs to the left.

Summary

This year we successfully completed:

1) The 2024 tissue sampling in Australia (juveniles) but unfortunately only 236 samples were collected in Indonesia in 2023/24 (adults) due to disruptions caused by the institutional changes in Indonesia.
2) Two training workshops for DGCF Enumerators to enable the SBT monitoring and sampling in Benoa to recommence in January 2024 (2023/24 season) and August (2024/25 season).
3) Tissue subsampling and DNA extraction and sequencing of juvenile samples collected in Port Lincoln in 2022.
4) Kin-finding (POPs and HSPs) to include the 2022 juvenile samples.

An updated dataset of identified SBT POPs and HSPs was provided to the CCSBT in June 2024. To date, a total of 123 POPs and 232 high-confidence HSPs have been identified with the false negative rate for HSPs estimated to be 0.25.

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, 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.

Farley J, Eveson P, Gunasekera R. 2022. Update on the SBT close-kin tissue sampling, processing and kin finding 2022. CCSBT-ESC/2208/09.

Preece AL, Bradford R, Eveson JP, Hartog J. 2024. Update on the Gene-tagging program 2024 and RMA request. CCSBT/2409/10.

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

Environment
Jessica Farley
+61 6 6232 5189
jessica.farley@csiro.au
www.csiro.au

Environment
Paige Eveson
+61 3 6232 5015
paige.eveson@csiro.au
www.csiro.au