CCSBT-ESC/0909/21

Update on the close-kin genetics project for estimating the absolute spawning stock size of SBT

Mark Bravington
Peter Grewe
Campbell Davies

Prepared for the CCSBT 14th Meeting of the Extended Scientific Committee
5-11 September 2009, Busan, Korea

Abstract

We describe progress on estimating SBT spawner abundance using close-kin data, following on from the study proposed in 2007 and updated last year. The main points are: continued collection of adult and juvenile samples; refinement of protocols and locus selection, to ensure reliable and replicable genotyping of very large samples; continued selection of further loci; preliminary examination of sib and half-sib incidence among the juvenile samples. We do not yet have enough fish genotyped to make any abundance estimate, but are on track to have an estimate available by CCSBT 2010.

Update on SBT close-kin abundance estimation

This paper is a short update on progress with SBT close-kin abundance estimation, following on from the study proposed in CCSBT-SC/0709/18 (Bravington and Grewe, 2007; Bravington and Grewe, 2008).

The project has a Steering Committee, including international expertise on population genetics, mark-recapture, and fisheries assessment, which met by phone in May. It was agreed that the next stage of the project should be to check sibling incidence amongst a subsample of juveniles (see below), since the CV of the adult abundance estimate could theoretically become excessive if a high proportion of juveniles are sibs or half-sibs. This check needs to be done prior to embarking on large-scale genotyping of adults and juveniles, and hence before any abundance estimate can be made. A preliminary check for siblings on 100 juveniles did not suggest any problems, although the sample size was limited. Following the Steering Committee meeting, we have genotyped 500 juveniles, enough to do a thorough sib-incidence check. We have also genotyped a number of adults, as part of the need to carefully co-ordinate and cross-check lab protocols between CSIRO (where the preparatory genetic studies have been done) and the Australian Genome Research Facility (where the bulk of the genotyping will be done). We now hold over 20,000 samples in total, with tissue subsampling complete for over 6000 fish, and DNA extraction into bar-coded storage for over 4000 fish; we are therefore close to finishing the preparations for genotyping our planned sample size of 7500. Once we have finished selecting loci, and assuming the sib-incidence check does not indicate any problems, we will begin mass genotyping and abundance estimation in time to produce an abundance estimate for CCSBT 2010.

We are currently using 11 loci for parent-offspring identification, selected on the basis of very "clean" scoring, high power to exclude unrelated pairs, and no evidence of genetic artefacts (e.g. good adherence to Hardy-Weinberg equilibrium). Enough fish have been genotyped for us to estimate allele frequencies reliably at these loci; given these frequencies and the large number of comparisons ($\sim 10^7$) between unrelated fish that will be made, about another 5 loci of equal power and reliability will be required before embarking on mass genotyping, in order to reduce false positives to negligible levels (about 3 more loci required for this) and to safeguard against false negatives that could theoretically arise through genotyping error. Work is in hand to identify suitable loci, and we expect to have a full set available by the end of 2009. Meanwhile, the 11 loci are sufficient to assess sibling incidence, as described below.

The table of samples collected and genotyped to date now stands as follows:

Year (Jul-Jun)	Place	Samples held	Subsampled	Extracted
2005-6	Indo	216	216	216
	PL	4000	500	700
2006-7	Indo	1520	700	700
	PL	4000	800	0
2007-8	Indo	1594	1594	1594
	PL	4000	1200	900
2008-9	Indo	1637	1637	0
	PL	3500+	0	0
TOTAL		20000+	6647	4110

Table 1: Samples collected and stored up to August 2009. Indo - Indonesia, via Benoa sampling program; PL = Port Lincoln via "freezer boat" processing during harvest. 2009 PL still being collected.

Checking for sib- and half-sib incidence

Each time a juvenile is genotyped, two adults are marked, which can then be recaptured amongst the genotyped adults. If two juveniles are siblings, then the marks are duplicated. As noted in Bravington and Grewe, 2007, this does not affect the expected number of *matches* (and therefore does not bias the overall estimate), but it does affect the expected number of *matching adults* because the potential matches are concentrated on a smaller number of adults. Consequently, the variance of the estimate will be increased if there are substantial numbers of full or half-sibs amongst the genotyped juveniles.

To check whether this is the case, we can use the juvenile genotype information by itself to look sibs and half-sibs, and to estimate the total number of parents that contributed to the entire pool of J genotyped juveniles. This could be anything between 2 (all genotyped juveniles are full sibs, from the same mating event) and $2J$ (no adult is a parent of more than one *sampled* juvenile); to get useful CVs, the number of contributing parents should be closer to $2J$. This sib-incidence check is an important staging post in the whole close-kin abundance estimation project, because if the sib incidence is too high, there would be no point in going through the costly process of genotyping the adults as the achievable precision would be very low. By the same token, though, we do not want to have to genotype the entire set of 10000+ juveniles merely in order to assess feasibility. We therefore need a procedure that can analyse a modest subsample of juveniles, check for sibs and half-sibs, and extrapolate to the entire juvenile sample.

This turns out to be a tough problem statistically— considerably harder than the estimation of adult spawner abundance— for two main reasons. First, there is a combinatorial explosion in the number of possible ways that a given number of sibling relationships can be distributed amongst a set of juveniles, and the pattern of those relationships has a major bearing on the number of contributing parents; computational efficiency is paramount, and existing algorithms for studying kinship assume datasets far smaller than we need to handle. Second, genotype data are not as informative about sibs, and particularly half-sibs, as they are about parent-offspring relationships; hence it is necessary to deal with uncertainty in the sib-status of a pair of juveniles, whereas for parent-offspring status we

will work with enough loci that false positives and false negatives are essentially impossible. However, we have now developed an algorithm which overcomes these difficulties.

We have already run a part of the algorithm on a set of 96 juveniles (2-year-olds from Port Lincoln, 2007) at 11 loci. Encouragingly, there was no evidence of sib- or half-sib incidence above what would be expected by chance using those loci on truly unrelated juveniles (i.e. about 4 apparent half-sibs, consistent with expected false positives, and no apparent full-sibs). If sibs or half-sibs do occur, they must surely result from high survivorship and subsequent persistent schooling associated with particular mating events. There is only a remote chance of (half-)sibs being found in different juvenile cohorts, or even from the same juvenile cohort caught in different years, so it suffices to look at a single year and cohort. The next step is to apply the algorithm in full to the 500 3-yr-olds from 2006 that we now have genotyped. These constitute about 1/6 of the entire samples that we have for that cohort in that year, and should provide a reasonable basis for extrapolating to the full set of samples from the cohort. We expect to have this prepared for publication by November 2009, in time for the Steering Committee to consider at its next meeting.

Acknowledgements

Particular thanks are due to Craig Proctor (CSIRO), Ms Retno Andamari (Research Institute for Mariculture, Gondol, Bali), Mr Kiroan Siregar, and Mr Rusjas Mashar (sampling enumerators) for their continued diligent efforts in ensuring the successful collection of Indonesian tissue samples.

We appreciate the assistance of the the Australian SBT industry (Port Lincoln), the Indonesian tuna fishing industry (Benoa), and the Research Centre for Capture Fisheries (Jakarta).

Funding for this work is provided by CSIRO Wealth from Oceans flagship and by the Fisheries Research and Development Corporation.

References

[Bravington and Grewe, 2007] Mark Bravington and Peter Grewe. A method for estimating the absolute spawning stock size of sbt, using close-kin genetics. Scientific Committee Report CCSBT-SC/0709/18, Commission for the Conservation of Southern Bluefin Tuna, 2007.

[Bravington and Grewe, 2008] Mark Bravington and Peter Grewe. Update on sbt close-kin abundance estimation, 2008. Scientific Committee Report CCSBT-SC/0809/29, Commission for the Conservation of Southern Bluefin Tuna, 2008.