WEALTH FROM OCEANS
www.csiro.au

Close-kin Update

Mark Bravington, Peter Grewe, Campbell Davies
CCSBT-OMMP/1307/Info1

Prepared for the CCSBT OMMP4: Fourth Operating Model and Management Procedure Technical Workshop, 23-26 July 2013, Portland Maine, USA

Wealth from Oceans Flagship

CSIRO Marine and Atmospheric Research

Copyright and disclaimer

Important disclaimer

WEALTH FROM OCEANS
www.csiro.au

DRAFT: Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers

Mark Bravington, Peter Grewe and Campbell Davies
FRDC Project No. 2007/034
June, 2013

Australian Government
Fisheries Research and Development Corporation

Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers

Mark Bravington, Peter Grewe and Campbell Davies

FRDC Project No. 2007/034

Australian Government
Fisheries Research and Development Corporation

Citation

Bravington MV, Grewe P and Davies CR (2013) Fishery-independent estimate of spawning biomass of southern bluefin tuna through identification of close-kin using genetic markers. CSIRO, Hobart.

Copyright and disclaimer

Important disclaimer

National Library of Australia Cataloguing-in-Publication entry

Contents

List of Tables

List of Figures

1 Non-Technical Summary

Principal Investigator

Mark Bravington
CSIRO Marine Laboratories, Castray Esplanade, Hobart, Tasmania 7000.
CSIRO Division of Mathematic, Informatics and Statistics.

Objectives

1. To provide a fishery-independent estimate of the number of SBT spawners.
2. To provide direct estimates of age-specific fecundity and an improved definition of spawning stock biomass

Outcomes Achieved

The planned outcomes achieved are:

1. provided an independent check of the assessment model, which are entirely reliant on fishery-dependent data;
2. provided for the incorporation of the SSB estimates (from this project) into the assessment;
3. reduced the uncertainty in the current assessment, and;
4. provided an independent benchmark to measure rebuilding of the stock.

It was also an expectation that the work on improving the definition of "spawning stock biomass" would lead to a better understanding of stock productivity, the relative importance of different age classes to total reproductive capacity of the stock and likelihood of different rebuilding trajectories. This has also been achieved.

Non-technical Summary

Estimation of abundance, either absolute or relative, is one of the main objectives of stock assessment. It is, however, one of the most difficult parameters to estimate. Nowhere is this more the case than for highly migratory stocks, such as tuna. Typical ground fish surveys are not possible. While other survey techniques, such as aerial spotting, can provide reasonable estimates of relative abundance, they are generally only practical for a component of the stock, are logistically difficult to maintain over the long-term, and are generally expensive relative to other methods. In the case of Southern Bluefin Tuna (SBT), there are other sources of abundance information included in the operating model that have been used to assess the status of the stock and evaluate Management Procedures (MP) since the early 2000s. Nevertheless, the dominant influence on the estimated trends in abundance is the CPUE series derived from Japanese

longline catch and effort data. Changes in the spatial dynamics of the stock and the fleet over time, combined with the large-scale unreporting of longline catches revealed in 2006, meant that it is no longer possible to conduct a stock assessment for SBT, in the traditional sense; rather scenario modelling has been used to assess the impact of the unreported catches and to evaluate and select a Management Procedure for setting global catches. The genesis for this project was the desire to develop a robust and practical method to estimate the abundance of spawning SBT, in particular, and highly migratory or "hard to observe" animals, in general, that was independent of fishery CPUE; and in doing so reduce the uncertainty in the state of the SBT stock and likely rebuilding rates.

The concept of close-kin abundance estimation, the rapidly declining cost of large-scale geneotying, and reliable access to subsamples of juvenile (Port Lincoln) and spawning adult SBT (Benoa, Indonesia) provided the prospect of being able to estimate the absolute abundance without requiring catch and effort data and to do so in a relatively short period (a few years). The basic idea is very simple: take a random sample of juvenile fish and a random sample of spawners, compare the genetic makeup of each juvenile and each spawner to see if the spawner could be a parent, and count the number of juveniles that appear to have a parent in the sample of spawners. If the population size of spawners is very large, then only a small proportion of juveniles will make a "match" (i.e. a parent in the random sample of spawners). The estimate of absolute spawning stock numbers is inversely proportional to the number of juvenile-parent matches. The application to any particular population, and SBT in this case, is somewhat more complicated. It requires the ability to cost-effectively sample adults and juveniles; unambiguously match juveniles to their parents using DNA profiling; undertake large-scale geotyping (14,000 individuals) while implementing and maintaining rigorous quality control and data management procedures; develop rigorous statistical procedures for testing fundamental assumptions of the method, defining appropriate data sets for analysis and estimating probabilities of false matches and no-matches; and construct and specify an appropriate abundance estimation model given the population in question and the available ancillary data.

In the course of this project we have genotyped approximately 14,000 individual SBT caught between 2006 and 2010 in the GAB (juveniles) and off Indonesia (mature adults). We demonstrated that the incidence of siblings in the samples is negligible and not a barrier to implementation of the method for SBT (a "go/no go" milestone in the early stage of the project). Formal statistical methods were developed for excluding potentially ambiguous comparisons and setting probabilities of mistakenly identifying unrelated juveniles and adults as Parent-Offspring-Pairs (POPs). In all, we found 45 POPs from about 38,000,000 "appropriate" comparisons. The quality control and exclusion procedures resulted in less than a 1% bias while only incurring a $\approx 2\%$ increase in standard error compared to what we would have gotten from "perfect" genotyping (where every pairwise comparison is usable). That is, we have kept the probability of mistakenly identfying POPs very low with negligible affect on

precision. This reflects very well on the tissue quality, the processing, and the selection of powerful, reliable loci.

Combining data from the POPs (the number found, plus their age, size, sex, and date of capture) with fecundity-at-size studies and Indonesian length, sex, and age-frequency data, we constructed a self-contained estimation model of absolute adult abundance of SBT that does not require any catch or CPUE data. We have demonstrated that the close-kin approach can provide cost effective and precise (CV of ~ 20% or less) estimates of the spawning abundance of SBT and, in the process, estimates of mortality and age specific spawning potential. The results indicate that the absolute spawning biomass is considerably higher (~3x) than the current estimate from the CCSBT OM. Direct comparisons between the two model estimates are not strictly legitimate, however, given the structural differences between the two. Preliminary results from including the close-kin data (POPs and their associated data) in the CCSBT OM (Hillary et al., 2012b) indicate that they are likely to substantially reduce the uncertainty in the trend of spawning biomass over the most recent decades and that the estimated level of depletion is likely to be reduced (from ~ 3-7% without the close-kin data to ~6-11% with the close-kin data included). The quality of these results clearly demonstrate the potential of the method for ongoing monitoring of the SBT spawning stock and the potential to extend the utility of the population genetic methods and protocols developed through this project to genetagging of the harvested components of the stock (i.e. 2-10 yr-olds). This potential to provide fishery independent monitoring of the each of the each of the main components of the population would remove the uncertainty associated with longline CPUE as an index of abundance and improve the confidence of all stakeholders in the monitoring and assessment of this valuable stock.

2 Acknowledgments

This project has run for over 5 years with contributions from many people, too many to thank individually. Particularly heroic efforts on genotyping and logistics have come from Peta Hill, Rasanthi Gunasekara, Danielle Lalonde and Matt Lansdell of CSIRO, with Jess Farley of CSIRO providing invaluable help with otolith and fecundity data. Special thanks for their diligent efforts in ensuring the successful collection of Indonesian tissue samples are due to Mr Kiroan Siregar, Mr Rusjas Mashar, and other scientists at the Tuna Fisheries Research Institute of Benoa, Ms Retno Andamari (Research Institute for Mariculture, Gondol, Bali) and Craig Proctor (CSIRO) and to ASF, Tony's Tuna and Protech Marine for Port Lincoln samples. We greatly appreciate the continued assistance of the Australian SBT industry (Port Lincoln), the Indonesian tuna fishing industry (Benoa), and the Research Centre for Fisheries Management and Conservation (Jakarta). We also thank the early sceptics and supporters whose questions of the concept forced us to think harder about our ideas, and in doing so, honed them and increased our motivation to demonstrate it could be done; the members of the Steering Committee, the Extended Scientific Committee of the CCSBT and Lindsay Pender for contributions to the rigour of the review process over the life of the project; and Ann Preece, Jessica Farley and Toni Cracknell for contributions to finalising this report. Funding has been provided by CSIRO Wealth from Oceans Flagship and the Fisheries Research and Development Corporation of Australia, with in-kind support from the Ministry of Marine Affairs and Fisheries, Indonesia.

3 Background

Estimation of abundance, either absolute or relative, is one of the main objectives of stock assessment. It is, however, one of the most difficult parameters to estimate. Nowhere is this more the case than for highly migratory stocks such as tuna. Typical ground fish surveys are not possible. While other survey techniques, such as aerial spotting, can provide reasonable estimates of relative abundance, they are generally only practical for a component of the stock, are logistically difficult to maintain over the long-term, and are generally expensive relative to other methods, such as mark recapture. Even in the case of mark recapture methods, there can be difficulties in meeting the assumptions of the estimation methods and obtaining reliable estimates of important ancillary parameters such as tag loss and reporting rates (Polacheck et al., 2005). As a result, stock assessments of highly migratory tunas, such as southern bluefin tuna (SBT), generally rely on catch per unit effort (CPUE) from commercial longliners as the sole, or principal, abundance index (Maunder et al., 2006, Polacheck 2006; Sibert et al., 2006).

In the case of SBT, there are other sources of abundance information included in the operating model (OM) that have been used to assess the status of the stock and evaluate Management Procedures (MP) since the early 2000s. These include conventional tagging data from the 1990s (Polacheck et al., 2005) and, more recently, an index of relative abundance of 2-4 year olds from the scientific aerial survey (Eveson et al., 2012; Hillary et al., 2009). Nevertheless, the dominant influence on the estimated trends in abundance of animals four years and older in the stock is the CPUE series derived from Japanese longline catch and effort data (Anon. 2008, 2009). It is the most extensive abundance series included in the assessment, dating from the 1950s, and covers the greatest proportion of the area of the fishery. There have been concerns, however, about how well this CPUE index reflects the underlying abundance of the stock and how this relationship may have changed over the history of the fishery as a result of changes in the fishery and the spatial dynamics of the stock. More recently, the uncertainty in the relationship between the CPUE index and abundance of the stock was heightened by the revelation of large-scale underreporting of longline catches over an extended period (Anon 2006 a and b; Polacheck and Davies 2008; Polacheck 2012). The scale, duration and resulting uncertainty of these unreported catches was such that in 2006 the Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) concluded that it was no longer able to conduct a stock assessment, in the conventional sense, and would need to rely on "scenario modelling" for the purposes of reporting on the likely state of the stock (Anon 2006a). The genesis for this project was the desire to develop a robust and practical method to estimate the abundance of spawning abundance of SBT, in particular, and highly migratory or "hard to observe" animals, in general, that was independent of fishery CPUE.

The concept of close-kin abundance estimation, developed by Skaug (2001) for Northeast Atlantic minke whales, the rapidly declining cost of large-scale geneotying, and reliable access to subsamples of juvenile (Port Lincoln) and

spawning adult SBT (Benoa, Indonesia) provided the prospect of being able to estimate the absolute abundance without requiring catch and effort data and to do so in a relatively short period (a few years). The basic idea is very simple: take a random sample of juvenile fish and a random sample of spawners, compare the genetic makeup of each juvenile and each spawner to see if the spawner could be a parent, and count the number of juveniles that appear to have a parent in the sample of spawners. If the population size of spawners is very large, then only a small proportion of juveniles will make a "match" (i.e. a parent in the random sample of spawners). The estimate of absolute spawning stock numbers is inversely proportional to the number of juvenile-parent matches.

In order for the approach to be feasible for SBT a number of technical issues needed to be clarified before embarking on a large-scale project. These included: the identification and optimisation of suitable genetic markers (in this case microsatellite loci) for SBT; evidence of sufficient mixing of eggs, larvae and juvenile SBT before sampling to ensure a reasonably random sample of juveniles from the Port Lincoln farming operations; development of large-scale processing protocols at low enough cost and with high quality control procedures to make processing of the required number of samples cost-effective relative to other approaches; development of the formal estimation model for SBT; and, most importantly in the context of the application to SBT, review of and support for the approach by Australian industry, management, policy and the Scientific Committee of the CCSBT. In 2006 CSIRO funded a proof of concept project to address some of these issues and, in 2007, following strong endorsement for the approach and project from the CCSBT Scientific Committee, the current project was funded by FRDC. Oversight of the project has been provided throughout by an international Steering Committee, which included expertise in genetics, mark-recapture and SBT assessment, and regular review by the CCSBT Scientific Committee (see sections 19 & 20).

4 Need

Management of SBT is greatly complicated by large uncertainties in the stock assessment. The adoption of a formal Management Procedure (MP) is a substantial advance (Anon. 2011a and b). It provides an agreed set of monitoring series and process for setting the global TAC (Anon., 2012, Hillary et al 2012). However, there remains considerable uncertainty in the size and productivity of the stock (Anon 2011). One key parameter is absolute spawning stock biomass (SSB), for which the only available estimates are highly uncertain and are driven entirely by fishery-derived data (e.g. longline CPUE). Furthermore, none of the current abundance indices included in the operating model are derived from direct observations of the spawning component of the stock. The majority of fish included in the CPUE series are less than 10 years old and the aerial survey only observes the 1-5 year old age classes, with the relative abundance index constructed to represent 2-4 year olds. There are age, size and sex data from spawning fish included from the monitoring of the Indonesian catches, but

these are included as catch data only. Fishery-derived data are generally much more difficult to interpret than data from designed programs, because the fishery dynamics can change over time, in ways that are difficult to quantify or even identify in a reliable manner. A fishery-independent estimate of absolute spawning stock size is therefore highly desirable, but current fishery-independent approaches such as the conventional tagging program cannot provide this for the spawning component of the population.

Recent advances in genetic and statistical methods now permit a fishery-independent estimate, using identification of parent-offspring pairs in random samples of juveniles and spawners. The same approach can also provide information on age-specific fecundity and thus on an appropriate definition of SSB (spawning stock biomass). This is another area of significant uncertainty for assessment and management because the different definitions of SSB have considerably different implications for stock projections and rebuilding times.

An absolute estimate of spawning stock biomass is particularly valuable given the estimated level of depletion of the SBT stock (Anon. 2011), and the high uncertainty about the productivity of the stock (i.e. the relationship between the parent stock and recruitment). The initial aim of this project was to provide an estimate of average SSB over 2002-2005. The project was extended to include additional samples through to 2010 to provide sufficient POPs for a robust estimate for that timeframe. Ultimately it will provide the methods to enable a time-series of SSB to be estimated, if sampling continues, which could conceivably provide a basis for direct fisheries independent monitoring of the spawning stock.

5 Objectives

1. To provide a fishery-independent estimate of the number of SBT spawners.
2. To provide direct estimates of age-specific fecundity and an improved definition of spawning stock biomass.[1]

6 Methods

6.1 Theory of Close-kin Abundance Estimation

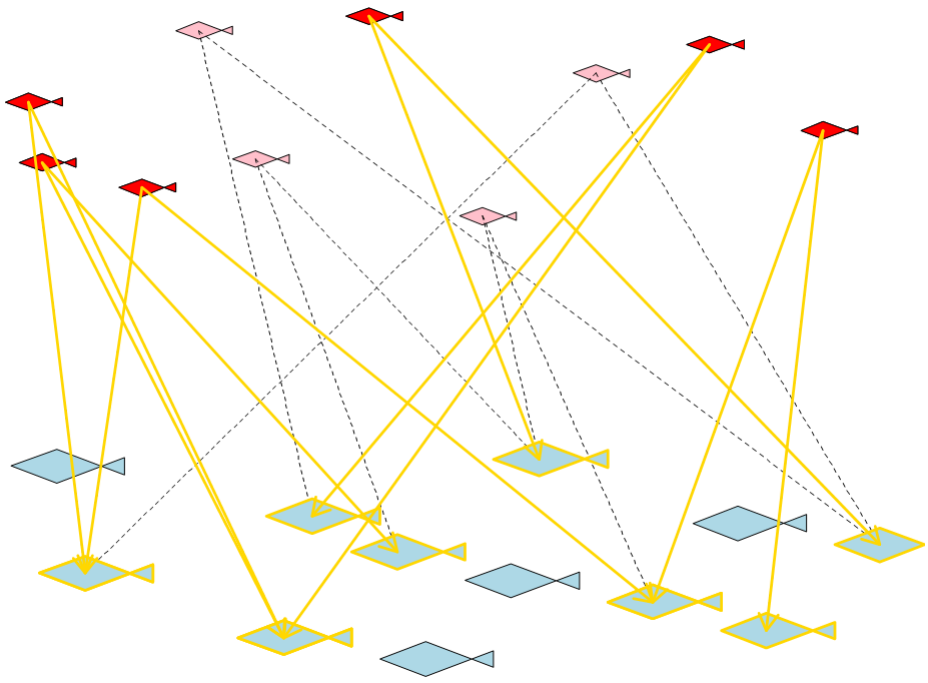
6.1.1 Simple estimator

Close-kin abundance estimation rests on two simple ideas:

- modern genetics allows us to tell whether any two fish constitute a Parent-Offspring Pair (POP), via "paternity analysis";
- all juveniles have two parents.

[1] Note: The original objectives for the project included reporting dates to the CCSBT ESC. These have been removed as they changed with the extension and rescheduling of the project.

Figure 1: "Cartoon" illustrating matching of juveniles (red) to parents (blue) by DNA profiling. The DNA profile of the juvenile "tags" two fish; each of its parents (adults at end of solid lines) in the adult population providing the basis to estimate the absolute abundance of adults from the number of matches (in this case 8) resulting from the number of comparisons made to identify them (in this case 12, the number of solid lines)



DNA tests are commonly used to test parenthood. Colloquially, for a typical "gene" with several variants in the population and two copies of the gene in each animal, a parent and its offspring must have at least one identical variant, whereas unrelated individuals might have totally different variants. Formally, a parent and its offspring must have at least one matching allele at every diploid locus. If a locus has a large number of different alleles, there is a low probability that two unrelated animals will have a matching allele at that locus just by chance. If we examine a large number of loci on each animal, the probability that two unrelated animals will have a matching allele at *every* locus (i.e. a false positive POP) is therefore extremely low. Hence, we can in principle completely rule out "false positives", i.e. apparent parent/offspring pairs that are really unrelated. False negatives (i.e. not identifying a true POP) are almost impossible if scoring is reliable, so from now on we assume that the genetic evidence is an exact indicator of a parental relationship (see section 16).

Now suppose you have a sample of m_A randomly-selected adults[2] and that, one year later, you collect a sample of m_J one-year-old juveniles. Pick one of the juveniles and one of the adults, and genotype both of them at enough loci to rule out any possibility of false-positives. What is the probability of a "hit"— i.e. that the chosen adult is actually a parent of the juvenile? Since the juvenile must have had two parents, the probability that the chosen adult is one of those two is $2/N_A$, where N_A (or just N) is the number of adults alive when the juveniles were spawned. Now repeat the comparison for the same juvenile and all the other adults. The expected number of hits between that juvenile and the entire set of m_A adults is $2m_A/N$. Now repeat this for all the juveniles: the expected total number of hits, $\mathbb{E}[H]$, is $2m_J m_A/N$. Thus, if h is the actual number of hits, we can form an approximately unbiased estimate[3] of N in the obvious way (formally, by using the "method of moments") via:

$$\hat{N} = 2m_J m_A/h$$

Note that the method cannot tell us anything about the total abundance of juveniles. The logic doesn't work in reverse: although we know that each juvenile must have had two parents, we don't know how many juveniles on average each parent would have had. In mark-recapture terms, each juvenile "marks" exactly two adults which might subsequently be recaptured, allowing us to estimate the number of adults. Looked at the other way round, though, each adult "marks" an unknown number of juveniles— which makes it impossible to use mark-recapture analysis directly to estimate the abundance of juveniles[4].

There are two crucial points to emphasize. First, the derivation of $\hat{N}$ does require that the adults are randomly sampled, but does *not* require that the juveniles are randomly sampled; in particular, the juvenile samples do not have to be mutually independent. Of course, the juveniles must be selected independently of the adults— the method breaks down if applied to mother-calf pairs (Ref: Skaug 2001), for example.

Second, the derivation of $\hat{N}$ does *not* require that all "adults" make an equal reproductive contribution. The key point is actually the random selection of adults. In fact, the "adult" population might be defined as "that set of animals which have equal probability of appearing in our m_A-sample". The trickiest part of applying the method to SBT, is correcting for unequal sampling probabilities among the "adults"; see section 6.4.

---

[2] Collected just after the spawning season, to avoid removing the very parents that we seek.

[3] As with most maximum-likelihood estimates, the estimate is only *asymptotically* unbiased, i.e. the bias disappears if the expected number of recaptures is large enough. For $h >$ ~10, the relative bias is about $1/h$, i.e. about 1.5% for the SBT project given the "target" of 70 for h.

[4] Skaug's (2001) method estimates adult and juvenile abundance together, and uses the number of half-sibling etc. matches as well as parent-offspring pairs. However, the method is less direct and requires extra assumptions which would not make sense for SBT.

6.1.2 Basic CV & sample size calculations

To get an idea of the uncertainty in $\hat{N}$, one further assumption is needed: that the numbers of hits from different juveniles are independent (see section 19.1). Then some algebra (see section 19) shows that

$$CV\left(\hat{N}\right) \approx \frac{\sqrt{2}}{m}\sqrt{N} \tag{1}$$

where m is the combined sample size (for optimality, split equally between adults and juveniles). Given some *a priori* notion of N, we can use (1) to set the sample size; e.g. a 10% CV requires about $15\sqrt{N}$ samples. For SBT, using the guesstimate from an earlier assessment of $N \approx 350,000$ (the number of fish $\geqslant 160cm$, the approximate length of 50% maturity), a target CV of 12% implies a sample size of 7000, with about 70 hits being expected and about 1% of the adults being sampled. We stress that this is only a sample-size calculation, and the achieved CV will be different for a number of reasons; see 6.4.

The remarkable thing about (1) is that it is (inversely) *linear* in sample size. By contrast, in the great majority of statistical settings, CV depends (inversely) on the *square root* of the sample size, meaning that diminishing returns usually set in as more data are collected. With close-kin abundance estimation, though, there is a quadratic gain in efficiency[5], basically because each new (juvenile) sample is compared against *all* existing (adult) samples, hence generating far more than one "data point".

6.2 Data collection

Collection of adult samples was made possible through existing collaborations between Indonesia and CSIRO and the SBT catch sampling programme in Benoa, Bali (Ref: Proctor et al., 2006). Samples for genotyping were taken throughout the SBT spawning season (October to March the following year) from all possible SBT >150cm. This size limit was chosen based on maturity data, to safely encompass all fish big enough to have been parents two years previously, when the youngest juveniles in the corresponding sample were spawned. The samples were collected from the Indonesian tropical tuna longline fleet that covers the main part of the SBT spawning grounds and largely take SBT as a by-catch from effort targeted at bigeye and yellowfin tuna (Proctor et al., 2003). The large majority of catches are landed at Benoa, although a much smaller, monitored but unsampled catch of SBT is taken further west from the Cilacap fleet, in an area of apparently lower SBT spawning density (Proctor et al., 2003). Muscle tissue was collected by a trained sampler and deep-frozen (-20 degrees C) for shipment to Australia. All fish sampled for genetics had their length measured and were sexed (by checking for residual female gonads; Farley et al., 2007), as part of the regular catch sampling programme. A portion of the fish genotyped form part of the otolith-collection set and so will be of

[5] Unless the sampling fraction becomes "large", or the period of sampling becomes so long that a high proportion of parents of "early" juveniles have died.

Table 1: Final tally of fish genotyped successfully. For adults from Indonesia (Ad) and juveniles from Port Lincoln (Jv), "year 2006" means "spawning season from November 2005 to April 2006", consistent with the definition of "SBT birthdays".

	2006	2007	2008	2009	2010	Total
Indonesia (Ad)	214	1457	1526	1394	1164	5755
Port Lincoln (Jv)	1523	1707	1448	1338	1432	7448
Total	1737	3164	2974	2732	2596	13203

known age (Ref: Farley et al., 2012). To ensure only spawners were sampled, and in the absence of precise information on fishing location, we excluded all SBT from trips with a high proportion of sub-adult fish, as in some years boats from some fishing companies fished further south, outside the SBT spawning ground (Farley et al., 2010). Coverage of the spawning grounds and spawning season was good.

Preliminary results from genotyping of ~5000 of the originally planned 7,000 fish, showed that the number of POPs found would end up considerably lower than originally expected if the original sample size was maintained. This would have meant the precision of the final results would have been very uncertain; to the extent that it was unlikely to be informative for the intended purpose (see 19). It was therefore agreed to substantially increase the sample size and additional funding was provided by industry, FRDC and CSIRO. Increasing the genotyped sample size was straightforward, given the additional funding, as many more frozen tissue samples from juveniles 2006-2010 were available than there was funding to genotype in the original project. In total nearly 14,000 SBT were processed. A summary of the genotyped fish by year and sampling location is provided in table 1.

Table 1 shows the final breakdown of 13,023 successfully[6] genotyped samples by year and site. Several hundred more were genotyped, but excluded in the end for reasons of quality control. Although the optimal scheme for a given budget would have been to genotype equal numbers of juveniles and adults (since this is likely to yield the greatest number of POPs for a fixed amount of genotyping effort), regulatory changes and delays with Indonesian export permits meant that we had to shift the balance somewhat towards juveniles. Almost all the Port Lincoln juveniles were age 3 in the year of sampling (based on clear separation of modes in the length frequency), except for a few in 2006 that were age 4. After 2006, the Indonesian samples were taken from every available fish (almost all >150cm length) alongside the existing catch monitoring and sample collection schemes. Sample collection is continuing in both Indonesia and Port Lincoln, but there are no immediate plans or funding to genotype more samples; they are simply being frozen for possible future use.

[6]Where successfully means the the fish has been genotyped and passed the subsequent genetic and statistical quality control checks to be included in the final analysis.

6.3 Development and selection of appropriate markers for close-kin abundance estimation

Every animal has two alleles at each locus, though the two may by chance be the same; one is inherited from each parent. Therefore, a POP must share at least one allele at every locus. Although two non-POP individuals could by chance share an allele at every locus compared, the probability is very low if the number of loci examined is large and the loci are individually highly variable, so that no one allele is particularly common. Therefore, the most basic and most rigid exclusion principle is: a pair is treated as a POP if, and only if, the two animals have at least one allele in common at all loci. We use the genetic data to find POPs, by first genotyping all the fish and then comparing every juvenile to every adult, eliminating non-POPs via "Mendelian exclusion". A brief guide to terminology can be found in Appendix 3, which also contains a more detailed description of the operational aspects of genotyping. The section below summarises the important attributes of the markers used to identify POPs and the criteria developed to exclude false positives and estimate the likely level of false negative POPs.

6.3.1 Marker development and selection

Loci developed for this project went through an particularly extensive checking and selection process. In short, we wanted a set of loci that: were highly variable, but not so variable that the longest alleles failed to amplify well; had a simple peak structure with minimal shoulder to the peaks and little stutter; and had clear gaps between alleles. Past experience indicated that, to be conservative and to facilitate automated genotyping, we needed to strictly focus on using tetranucleotide repeats that gave solitary, sharp, allele peaks. Over time, as more fish were scored, some of our best tetranucleotide loci turned out to have some two-base-pair insertion/deletions, which meant that some alleles were separated by only two base pairs (though usually at least one of the alleles involved was rare). This was tolerated, provided there was at least a one-base-pair gap between bins. Loci were discarded during the initial development project if they showed alleles separated by just one based pair, indicative of poly-nucleotide tracks in the amplified allele.

After genotyping 5000 fish, we selected 20 loci organized into 5 panels A-E, with very comprehensive scoring bin sets into which almost all detected alleles fell. At this point we included an additional 7 loci (total 27) which were reorganized into four multiplex panels H, I, J, and L. We scored all 27 loci where possible in the remaining 9000 fish, but used only 25 loci for finding POPs; the remaining two loci, with slightly less reliable scoring, were used only for quality control (QC) purposes. When scoring, our protocol was not to record a score if in doubt, which is safe for the purpose of POP-finding.

Excess homozygosity An important check in genetic studies, is on the proportion of homozygotes found at each locus. In theory, provided a number of

assumptions hold, this can be predicted from the allele frequencies, and the extent to which there is an excess of *apparent* homozygotes is one indication of the reliability of a locus. As shown in Table 2, all but 3 of the 25 primary loci have both low expected homozygosity (which corresponds to being highly variable, and thus powerful for POP identification), and at most a small excess observed homozygosity as given by the difference between the EXP and OBS rows; this suggests relatively few cases of failure to see the 2nd peak in a heterozygote, for example.[7]The exceptions are in the bottom right of table: D569 and D573. It appears that the excess of homozygotes in those two loci is due to "heritable nulls" (eg from a mutation in the flanking sequence so that primers don't bind), so that some alleles simply don't amplify. No loci showed appreciable evidence of Short-Allele Dominance.

To guard against the possibility of heritable nulls in *any* locus, a comparison of two different apparent homozygotes (AA in one fish vs BB in the other) was *not* used to exclude a POP even though there is ostensibly no shared allele, in case the real score was "A-null vs B-null" with the null being inherited. This relaxation has only a small effect on the false-positive probability. However, it is not feasible to relax the exclusion criterion further to allow for the commonest (but still fairly rare) scoring error whereby the second peak of a heterozygote is missed, i.e. by treating AA vs BC as not necessarily excluding. Such a weakened criterion would generate many false positives with the existing set of loci, so many more loci and more expense would have been required.

6.3.2 Genotyping

For the last 9000 of the 14000 fish genotyped (from both sites), the procedure was as follows.

1. Tissue biopsy samples from each fish were collected, labelled (location, date of capture, length as indicator of age in juveniles, and sex in adult fish), and stored fresh-frozen at -20C in boxes of 100 fish and in the case of the Indonesian fish are cross-linked to the existing otolith database.

2. DNA was obtained from a 10mg tissue subsample from the original muscle biopsy plug from each fish. Subsamples were placed in deep well microtitre plates and extraction was completed on an Eppendorf EP-motion robotics liquid handling station using Macherey-Nagel NucleoMag® 96 tissue prep kits. The remaining biopsy tissue was archived should future cross-checks be required.

3. The DNA of 92 subsamples at a time was extracted into solution supplied with the Macherey-Nagel kit. These were eluted into a 96 well micro-titre plate to be used as the PCR template plate, incorporating two control DNAs (in specified positions) and two water blank controls (in known positions, variable from plate to plate). Water blanks and control DNA were place into keyed positions that could be used to uniquely identify each plate and determine orientation from subsequent run data.

[7]Other reasons for deviation from Hardy-Weinberg, such as population structure, are unlikely for SBT, and in any case no deviation is seen for many of the loci despite the very high sample sizes and consequent high power to detect any deviation.

Table 2: Homozygote percentages, "expected" (ignoring nulls) and observed for the 25 primary loci used for bulk screening for POPs; see text.

	3D4	B5	D10	D111	D11B	D12	D122	D201	D203	D211	D225	D235	D3	D4D6
EXP	19.8	6.8	7.1	11.8	10.7	10.8	9.7	11.7	7.5	11.4	3.4	8.5	16.8	5.5
OBS	19.8	7.3	7.3	12.2	12.3	10.9	11.4	12.4	9.0	17.0	3.7	14.8	16.8	6.7

	D541	D524	D549	D570	D592A	Z3C11A	D517	D534	D582	D569	D573
EXP	14.0	12.4	11.9	7.3	9.8	13.0	3.1	9.3	7.6	9.9	4.9
OBS	14.0	13.5	11.9	7.3	10.2	13.4	3.4	10.1	7.7	45.5	30.9

4. Four micro liters was subsampled from the template plate and used in PCR amplification of four separate DNA multiplex reaction panels (H, I, J, L). Reactions were setup with a standard mix containing combinations of primers for up to 8 loci per panel. Reactions were set up in a run-plate using the Eppendorf EP-motion liquid handling station. PCR amplification was accomplished using Qiagen master mix enzyme and each plate was run on Eppendorf silver block 96 well thermocylcers.

5. Following PCR each of the first column of each run plate was sub-sampled and run on an ABI-3730-XL capillary electrophoresis sequencer running pop-7 polymer, 50cm capillaries, and Data Collection v3.1 from Applied Biosystems. Plates were processed in groups of six template plates for a total of 24 run-plates representing amplification products of all four panels. Once successful amplification and genotype of the first control DNA in position A01 was confirmed across 24 "first column" subsamples, run plates were cryovac sealed ready for shipment. Run plates are sent to the Adelaide node of AGRF (Australian Genome Research Facility) for fragment separation. There are four run plates per 96 fish, labelled H/I/J/L depending on the panel of loci involved. For each run plate, the result was four sets of 96 "FSA files". FSA files were transmitted to CSIRO Hobart where genotypes were scored using an automated calling program GeneMapper® v4.1 & Data Collection v3.1 from Applied Biosystems.

6. FSA files were scored at CSIRO by one of four experienced readers, each of whom scored several thousand samples. Results from the various scorers have been cross-validated for consistency on some plates.

(a) The check plates were sequenced at CSIRO using similar machinery to AGRF's, and genotype results were subsequently compared to the corresponding columns of the FSA files obtained from AGRF. This provides a safeguard against plates being swapped or rotated, and against faulty calibration of the sequencer.

(b) Each panel included a common locus B8B, so by comparing the B8B scores across run plates ostensibly from the same template plate, it was possible to check whether the files for each run plate really did come from their nominal sample plates.

7. A variety of QC checks are run on the FSA files, to detect plate-level phenomena such as rotation/swapping/miscalibration (see [sub:Terminology]), atypical allele frequencies, excess homozygotes, and individual-level phenomena such as duplicate genomes which could arise if samples are inadvertently double-sampled at the point of collection or during tissue sub-sampling.

For the first 5000 fish genotyped, a slightly different and less streamlined procedure was used in steps 4-5. Only 20 of the total of 27 final loci were initially examined among the first 5000 fish. The amplification reactions for 10 of these 20 were done by AGRF using single-plex PCR that were then co-plexed together for fragment separation in three panels A/B/C. One of the first 10 plus the remaining loci were amplified at CSIRO, as above, in two PCR multiplex panels (D and E) for a total of 11 loci and fragments were separated at AGRF. After the first 5000 fish had been genotyped and analysed, the A/B/C panels were subsequently reorganized into two of the final multiplex panels (I & J) used for the last 9000 fish; the D/E panels became the H & L panels

after addition 5 extra loci. After the FSA files returned to CSIRO, we used a shared locus on the D & E panels to check their "alignment", as in [enu:B8B-1]. To check alignment of the A/B/C panels with each other and with the D/E panels, we put DNA drawn from the first column of the template plates for D/E panels into one column of an extra template plate, which was then used to make run plates for the new I & J panels (containing the same loci as A/B/C, but organized differently). These were sequenced, scored, and the genotypes compared against the corresponding columns from the original A/B/C plates. Although this process was somewhat cumbersome and led to some duplication in scoring (about 10%), it provided an important safeguard for the detection of handling errors that become more likely with such large sample sizes.

6.3.3 Exclusion criteria

This project relies on the number of POPs actually identified being close to the true number of POPs in our samples. There are two possible issues. The first is false-positives: an unrelated pair might happen to share an allele at every locus just by chance, and thus look like a POP. This probability can be assessed in advance from the allele frequencies, and this step is essential in determining whether enough loci are being used. Not all loci are successfully scored for all fish, so some comparisons will involve a lot fewer than the maximum of 25 loci in our study, and those comparisons will have a substantial false-positive probability. By excluding such "weak" comparisons, we can control the overall false-positive rate so that the expected number of false positives is negligible compared to the number of true positives.

The second possible issue is false-negatives, whereby a POP *appears* not to share an allele at one or more loci. This could arise through mutation, but only very rarely; published estimates of mutation rate for the kind of loci that we used are of the order of 10^{-4} per generation, so with about 25 loci in our comparisons well under 1% of true POPs would be affected by any mutations. A more likely cause of false-negatives is scoring error, whereby the true alleles are incorrectly recorded. Scoring error rates are highly variable between studies (and to some extent between loci within a study), depending on the quality of the DNA itself (i.e. tissue preservation), how carefully the loci are chosen, how carefully protocols are followed, and how much checking is done. Careful checking can detect and eliminate large-scale scoring errors involving many fish at once (see 16.3). However, a different approach is required for small-scale errors at the level of single loci on single specimens.

Because there are so many different possible causes of scoring error, false-negative rates cannot be predicted in advance (unlike false-positive rates), and can only be inferred after the fact. This is usually done by re-scoring individuals to see how often the scores change. However, depending on the details there may be a possibility of making the same mistake twice, so re-scoring may underestimate the scoring error rate. With our POP-oriented study, we can use a more direct and robust approach; we are using so many loci that the chance of two non-POPs sharing an allele at all-but-one of (say) 25 loci is negligible,

and consequently any pairs that *seem* to share alleles at 24 of 25 loci with a mismatch at the 25th are highly likely to be false-negatives arising from scoring error[8]. The proportion of such cases compared to unambiguous true POPs (where all loci share an allele) can be used to estimate the overall false-negative rate.

6.3.4 Genetic and Data Management Quality control

The scale of genetics processing associated with this project demanded that rigorous genotyping and data management protocols were developed and implemented to minimise the risk of errors in the genetics processing, scoring of alleles and processing of the very large resulting data sets. These protocols and tests are very important outcome of the study. However, they are largely only of interest to those with a detailed knowledge of genetics, statistics or both. Hence, they are described in greater depth in Appendix 3.

6.4 Close-kin abundance estimation model for SBT

There are three main reasons why the $2m^2/P$ "cartoon" estimator ("cartoon" in the sense of Figure 1) would be seriously misleading for SBT: i) the multi-year sampling nature of the study; ii) age dependent sampling probability and, iii) non-equilibrium conditions in the spawning population. The following section describes the elaborations of the simple estimator required to address these in the case of SBT and an overview of the estimation model developed as part of the project to accommodate them. A detailed specification of the model is provided in Appendix 5.

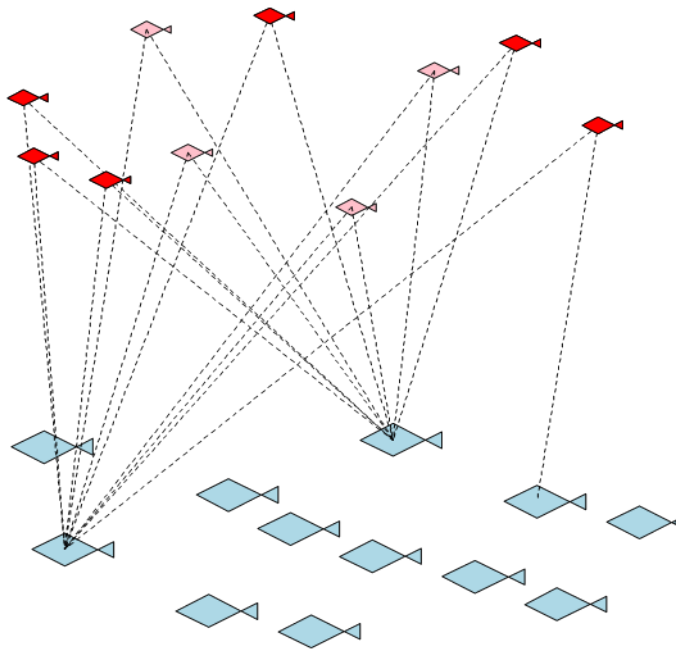
6.4.1 Necessary considerations for application to SBT

Sampling delays and multi-year sampling The first is that we cannot do comparisons only against the "parental cohort-group" of each offspring, i.e. the group of adults that were alive at its birth. Figure 2 illustrates the main point; if survival rates are the same for all adults, then the cartoon estimator would still be valid even with time lags, *provided* we could restrict comparisons to the light-grey parental cohort-group. But we cannot do so, because (i) we do not know the age of all adults sampled, (ii) maturity is not knife-edge so there is no absolute definition of the parental cohort-group, and (iii) maturity is quite likely length- rather than age-driven. If we are forced instead to sample adults from say the entire 4th column of Figure 2, after a 3-year gap, then a high proportion of comparisons will be with "impostor" adults that could not have been parents, and the cartoon estimator would be biased high. Hence, it is necessary to account for these time lags and the differential probability of an adults being the parent of a juvenile in each comparison.

[8]Note that scoring errors do not increase the false-positive probability— there is no reason to think that an error in scoring one fish will either increase or decrease the probability of it sharing an "allele" with another unrelated fish.

Figure 2: Dilution of original parent-cohort-group by incoming recruitment

Figure 3: Big SBT are more fecund *and* more likely to be caught



Age-dependent sampling probability The second, linked, reason is that adult sampling is strongly selective towards large/old fish, which are also likely to have been more fecund (even allowing for a 3-year time lag). Because they are more fecund, they have more "tags" per capita (i.e. juveniles that they are parents of), and each tag is more likely to be "recaptured" (i.e. the adult is more likely to be caught) because of selectivity in favour of larger adults. This is the close-kin analogue of "heterogeneity in capture probability", a well-known issue in mark-recapture abundance estimation. Figure 3 shows the cartoon version. The upshot for the naive $2m^2/P$ estimator would be that each comparison is more likely to yield a POP than would a comparison with a randomly-chosen adult.

Both effects concern not the number of POPs actually found, but rather the difficulty of working out how many comparisons are "relevant"[9] for abundance estimation. The two effects act in opposite directions; the time-lag dilution means that some comparisons are invalid and thus less effective than "random" comparisons, whereas the selectivity-fecundity correlation means that the valid comparisons are more effective than "random" comparisons. The time-lag dilution is also mitigated by growth, since the surviving "original" adults after 3-4

[9] where "relevant" means it is feasible that a particular adult could have been the parent of a particular juvenile

years will be bigger than the "impostors" and thus more likely to be caught. However, there is no particular reason to assume the effects will cancel out, since the time-lag effect is driven primarily by the length of the study whereas the selectivity-fecundity effect is determined by the nature of the fishery and the growth curve. And the effects can be quite large; with an adult survival of say 0.8, after 3 years only 50% of the original adults are still alive to be sampled and the impostors will (in equilibrium, and neglecting selectivity and growth) be involved in about 50% of the comparisons. To deal with these issues properly, we need a mini-assessment.

Non-steady state conditions One further issue arises from of the extended timespan of this study, which spans juvenile birth-years from 2002 and adult capture-years to 2010, as well as the initial age structure of the adults in 2002, which was determined by even earlier events. The 1990s and 2000s have been eventful decades for SBT involving historically low recruitment events and large reductions in catches and it may be such that steady-state assumptions are simply not viable.

A proper close-kin abundance estimate for SBT therefore has to deal with survival, selectivity, fecundity, and growth, and perhaps with changes in abundance over time. The requisite data come from the length and age-at-length samples from Indonesia, plus results of previous fecundity studies. While not strictly "fishery-independent", length and age data are not subject to the same problems as CPUE or total catch. It also makes sense to split the analysis by sex: the cartoon applies equally well if applied to males and females separately, where the chance of a POP comparing to a male adult is $1/N_{\text{male}}$ not $2/\left(N_{\text{male}} + N_{\text{female}}\right)$, and C is split into $\left(C_{\text{male}}, C_{\text{female}}\right)$.

6.4.2 Residence time, selectivity, and fecundity

The tropical waters off Indonesia are really no place for an adult SBT, an animal that is adapted superbly for much cooler temperate waters. Adults arrive on the spawning grounds fat, and leave thin. Of course, the longer they can stay on the grounds, the more chances to spawn they will have, so it seems reasonable to suppose that they will put up with Indonesian conditions for as long as their bodies let them. The key for disentangling the effects of fecundity, survival, and selectivity, is average **residence time** on the spawning grounds, as a function of length. A cursory glance at length distributions from Indonesia shows that few fish under 150cm, and none under 130cm, are caught on the spawning grounds, so there is obviously some link to length. As "average residence time" already factors in the probability that a fish won't be there at all in any given year (i.e. in the case that skip spawning is a real phenomenon), the estimation model specifically assumes that, given length and sex:

- Selectivity $\propto$ residence time
- Annual reproductive output $\propto$ residence time $\times$ daily reproductive output

Except as specifically noted later, we assume that length and sex are the driving influences behind the behaviour of adult SBT, rather than age.

Of course, there could be other "second-order" phenomena which slightly change the above relationships (e.g. different depth distributions by size, and thus different exposure to hooks; different egg *quality* with parental size; etc etc) but these seem likely to be small compared to the dominant effect of residence time. For the rest of this document, it may be helpful to think about selectivity and residence time as directly equivalent, at least within each sex.

We have no direct data on residence time as a function of length, so the relationship needs to be estimated indirectly from data. It is worth noting that independent data on residence time and depth distribution as a function of length, from archival tags placed on big fish, would be extremely useful: both in tightening up parameter estimates in our existing model, and in assessing whether the effects that we necessarily assume are "second-order" really are.

6.4.3 Fecundity analyses: daily reproductive output

The canonical reference for SBT (female) spawning biology and fecundity is a study from the early 2000s by Davis et al.[10]. In summary, female SBT while on the spawning grounds have an on-off cycle, consisting of several days of consecutive daily spawning (one spawning event per 24 hours), followed by several days of rest while more eggs are built up. This on-off cycle may be repeated several times. As soon as the final spawning cycle is complete the available evidence suggests they leave. The mass of eggs released per daily spawning event can be estimated from the change in gonad weight between just-about-to-spawn and just-after-spawning fish; it is approximately proportional to length to the power 2.47 (Davis et al., 2003). The average duration of each part of the cycle (and thus the proportion of days on the spawning grounds when spawning actually occurs) can also be estimated as a function of body length using histological data, because the first day of a spawning sequence can be distinguished from the other days, and similarly for a resting sequence. However, the number of cycles per season is completely unknown, and is obviously set by the residence time on the spawning ground.

To summarize, the factors involved in daily reproductive output are:

- reduction in gonad weight per spawning event
- duration of consecutive daily spawning events
- duration of consecutive resting days

A reasonable amount of data is available for all three of these, and the relationship to length can be estimated from fitting three GLMs. (This was already done for the first two factors in Ref: Davis et al., and the third factor was addressed during this study.) For now, we have treated the parameter estimates as exact in the rest of the estimation model.

[10]T. Davis, J. Farley, M. Bravington, R. Andamari (2003): *Size at first maturity and recruitment into egg production of southern bluefin tuna* FRDC project 1999:106

We have no comparable data for males, nor on the extent to which male abundance actually influences the number of fertilized eggs per year.

6.4.4 Indonesian length, sex, and age data

A substantial proportion of the Indonesian SBT catch is sampled as it passes through the main landing port of Benoa. Length (to the centimetre) and sex are always recorded, and nowadays otoliths are always extracted, although only a length-stratified subset (500 per year in the recent past) are aged on a routine basis. Between 900 and 1700 animals were measured per year between 2002 and 2010. Thus the data can be seen as

1. Random samples of length and sex from the entire adult catch
2. Random samples of age, given length and sex.

Even without the POP data, it is possible to do some steady-state analysis of the age/length/sex data (though it is obviously impossible to estimate absolute abundance), but it is impossible to completely separate selectivity (as a function of length) from average adult survival rate. When the survival rate is very high (e.g. 0.9) or very low (e.g. 0.5) it does become impossible to match the observed length-frequency distributions except by invoking a ludicrous selectivity function, but in the absence of other data reasonable fits to the age and length data can be obtained across a wide range of survival rates.

Fortunately, the POPs can help estimate survival rate, in addition to absolute abundance. The typical gap between offspring birth and adult capture— assuming that the adult is in fact captured subsequently, i.e. that the pair is an identified POP— is related to survival. If survival rates are low, very few parents will survive to be caught say 7 years later (the maximum gap possible in this study), so most of the POPs that are found will be separated by just one or two years. Growth and residence time need to be properly accounted for too, but the intuitive basis should be clear. The close-kin data thus has three vital roles: the *number* of POPs (given the number of comparison) essentially sets the scaling of absolute abundance, the age and length distribution *within* the POPs informs on selectivity/fecundity, and the distribution of time-gaps *within* the POPs essentially determines survival.

6.4.5 Model structure

The model keeps track of numbers by age and sex; each year, each fish either gets one year older or dies. However, most phenomena are driven by length, which is assumed to have a fixed distribution at age. Each fish has its own personal asymptotic length drawn from a sex-dependent log-Normal prior whose mean and variance at age are fixed, while the other von Bertalanffy parameters are the same within each sex. A plus-group is used for ages 25 up, and a minimum "recruitment" age for possible spawning also needs to be set (currently 8). There is also a plus-group for length (200cm) and, unusually for stock assessments, a

sort of "minus-group" as well, currently set to 150cm. Experience with fitting just to age and length data showed that trying to extend the fit to the small proportion of adults below 150cm gave poor results, in that this small "tail" started to "wag the dog" and distort the fit elsewhere. The focus of this study is spawners, which are mostly 160cm and up, so it is more important to get a good fit there than to squeeze a last drop of misinformation out of very small adults. However, it is necessary to somehow keep track of the small spawning contribution of fish in the minus-group, and accordingly there is some tedious book-keeping code in the model.

Most of the likelihood is quite standard; multinomial distributions for length-sex frequency data, and for age given length and sex. The effective sample sizes of the length and age data were capped at 300 per year, to avoid these data swamping the information from the POPs. The novel term is the contribution of the POPs. For each comparison made between a juvenile j and an adult i of sex (gender) g_i, the outcome (POP or not) is a Bernoulli random variable with probability given by

$$\mathbb{P}\left[j \sim i\right] = \frac{\text{expected ARO from } i\text{in year of } j\text{'s birth}}{\text{total ARO from adults of sex } g_i\text{in that year}}$$

where ARO is Annual Reproductive Output, i.e. daily fecundity multiplied by residence time as in section 6.4.2. This formula replaces the "$2/N$" probability in the simplest possible close-kin implementation.

To actually compute a likelihood, it is necessary to specify various terms:

- numbers-at-age in 2002, and for incoming recruitment (age 8) in 2003-2010;
- survival rate in each year and age;
- residence/length relationship;
- growth parameters;
- relation between daily RO and length *for males*.

The total number of potential parameters is colossal because of the numbers-at-age and survival terms, so of course one needs to specify them parsimoniously given the limited amount of data available. This is done using formulas (sensu R) for each bullet-point term above, describing what covariates are allowed to influence it, and perhaps what functional form that influence might take. For example, we might choose to make survival constant over age and time, except for the plus-group[11]. We might also make assumptions of constant "recruitment" (at age 8) in the 2000s; and/or that numbers-at-age prior to 2000 were in equilibrium with survival; and/or that von Bertalanffy k is the same for both

[11] In SBT as with other top-predators, it must be the case that natural mortality rate increases for old animals, since simple maths shows that the sea would otherwise just fill up with decrepit tuna.

sexes; and/or that the slope of the residence/length relationship (but not its midpoint) is the same by sex; etc.

The final term—male daily reproductive output as a function of length—can *in principle* be estimated provided we are willing to assume that survival rates for males are the same as for females. Without that assumption, there is nothing to anchor the selectivity/survival/fecundity triangle for males. For females, we do not need to estimate this term because we have direct data from the fecundity studies.

The likelihood itself is coded in Pascal, with derivatives computed by an automatic differentiation toolbox similar to ADMB. The overall data-handling and fitting is done in R, calling the `nlminb()` optimizer to do the fitting. Some care was needed to avoid numerical problems in calculating the log-likelihood, and because of limited time there are still starting-value problems so that some model parametrizations can't yet be fitted. However, once a starting value has been obtained, no convergence problems were encountered, at least for the fairly parsimonious specifications (say 15 parameters) included here. Detailed specification of the estimation model is provided in Appendix 5.

7 Results

7.1 Genetics: Finding Parent-Offspring Pairs

7.1.1 Limiting false positives

Barring genotyping or scoring errors, a POP must have at least one allele in common at every locus, so if a pair is unrelated we will eventually be able to rule it out as a POP by finding a locus that does not share an allele, provided that we look at enough loci. We have scored 25 loci[12] overall, but not all loci are scored for every fish, so some pairwise comparisons involve many fewer loci. If too few loci are used in a comparison between unrelated fish, there is a substantial probability that all the loci will share an allele just by chance. We therefore need to do some filtering, to exclude comparisons that are too likely to give a false positive. Table 3 shows what happens if we *don't* do any filtering. True POPs— plus false POPs, which just happen by chance to share an allele at every locus compared— are in the leftmost column "F0", i.e. with zero loci compared that do not share an allele. False POPs are obvious in the top-left of the table, where very few loci are being compared.

Note that the Table includes a small proportion of (i) impossible and (ii) useless comparisons, where the adult was (i) caught in a year before the juvenile was born, or (ii) caught in the same year. Type (ii) comparisons are biologically possible, but it's not helpful to include same-year comparisons in abundance estimation, because in the year of its capture an adult will not achieve its normal annual reproductive output. All such comparisons have been removed in subsequent summaries and results.

[12] Plus another two that showed occasional anomalies, and were therefore omitted from routine pairwise comparisons, but were used in checking ambiguous possible-POPS.

Table 3: All comparisons, broken down by #loci compared and #loci inconsistent with POPhood(see text). Hash (#) means "number of", dot means zero, plusses mean too big to fit.

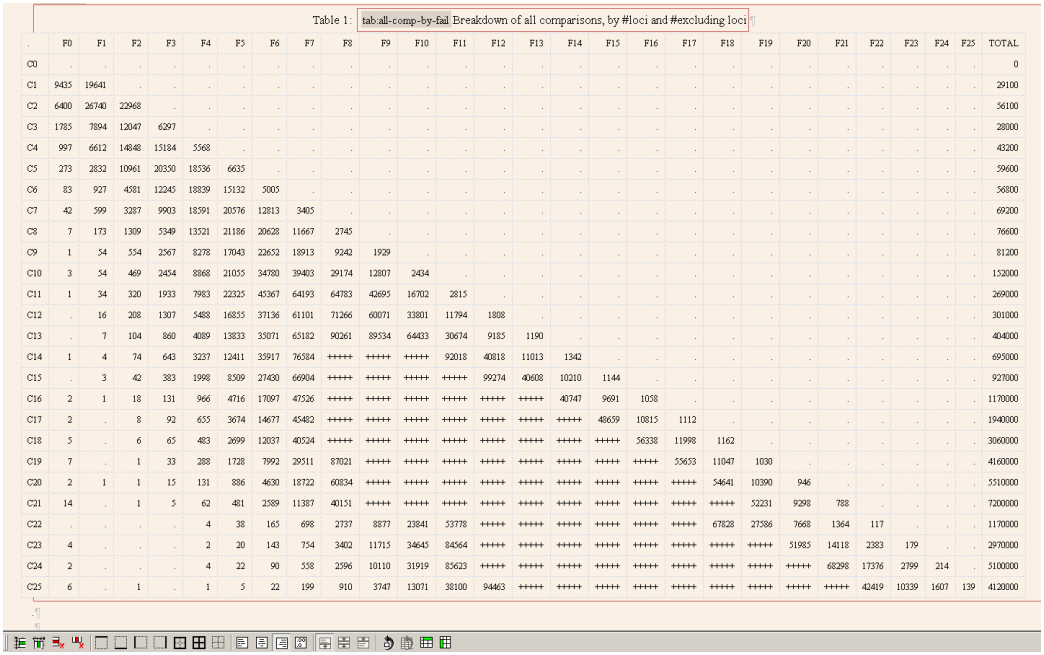


Table 1: tab:all-comp-by-fail Breakdown of all comparisons, by #loci and #excluding loci

.	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	TOTAL
C0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0
C1	9435	19641	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	29100
C2	6400	26740	22968	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	56100
C3	1785	7894	12047	6297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	28000
C4	997	6612	14848	15184	5568	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	43200
C5	273	2832	10961	20350	18536	6635	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	59600
C6	83	927	4581	12245	18839	15132	5005	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	56800
C7	42	599	3287	9903	18591	20576	12813	3405	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	69200
C8	7	173	1309	5349	13521	21186	20628	11667	2745	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	76600
C9	1	54	554	2567	8278	17043	22652	18913	9242	1929	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	81200
C10	3	54	469	2454	8868	21055	34780	39403	29174	12807	2434	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	152000
C11	1	34	320	1933	7983	22325	45367	64193	64783	42695	16702	2815	.	.	.	.	.	.	.	.	.	.	.	.	.	.	269000
C12	.	16	208	1307	5488	16855	37136	61101	71266	60071	33801	11794	1808	.	.	.	.	.	.	.	.	.	.	.	.	.	301000
C13	.	7	104	860	4089	13833	35071	65182	90261	89534	64433	30674	9185	1190	.	.	.	.	.	.	.	.	.	.	.	.	404000
C14	1	4	74	643	3237	12411	35917	76584	+++++	+++++	+++++	92018	40818	11013	1342	.	.	.	.	.	.	.	.	.	.	.	695000
C15	.	3	42	383	1998	8509	27430	66904	+++++	+++++	+++++	+++++	99274	40608	10210	1144	.	.	.	.	.	.	.	.	.	.	927000
C16	2	1	18	131	966	4716	17097	47526	+++++	+++++	+++++	+++++	+++++	+++++	40747	9691	1058	.	.	.	.	.	.	.	.	.	1170000
C17	2	.	8	92	655	3674	14677	45482	+++++	+++++	+++++	+++++	+++++	+++++	+++++	48659	10815	1112	.	.	.	.	.	.	.	.	1940000
C18	5	.	6	65	483	2699	12037	40524	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	56338	11998	1162	.	.	.	.	.	.	.	3060000
C19	7	.	1	33	288	1728	7992	29511	87021	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	55653	11047	1030	.	.	.	.	.	.	4160000
C20	2	1	1	15	131	886	4630	18722	60834	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	54641	10390	946	.	.	.	.	.	5510000
C21	14	.	1	5	62	481	2589	11387	40151	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	52231	9298	788	.	.	.	.	7200000
C22	.	.	.	.	4	38	165	698	2737	8877	23841	53778	+++++	+++++	+++++	+++++	+++++	+++++	67828	27586	7668	1364	117	.	.	.	1170000
C23	4	.	.	.	2	20	143	754	3402	11715	34645	84564	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	51985	14118	2383	179	.	2970000
C24	2	.	.	.	4	22	90	558	2596	10110	31919	85623	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	68298	17376	2799	214	.	5100000
C25	6	.	1	.	1	5	22	199	910	3747	13071	38100	94463	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	42419	10339	1607	139	4120000

In order to filter out false POPs, we first compute in advance for each possible pair a False-Positive Probability (FPP) (i.e. the probability that the two animals will share an allele at every locus compared, even if unrelated) based on which loci were scored successfully for *both* fish in the pair, and without looking at the actual genotypes that resulted. We then sort these FPP in ascending order, and find the cutoff such that the *total* FPP from all (sorted) pairs below the cutoff is below some pre-specified threshold T. Only those pairs whose FPP falls below the cutoff are subsequently checked for POPhood (whether or not they match at all relevant loci and therefore are a "true" parent-offsrpring-pair), the remainder being deemed too ambiguous, given the number and nature of loci involved in the comparison. Note that not testing POPhood of an ambiguous pair does not cause any bias in the proportion of included comparisons that yield POPs, because the FPP check is done *before* testing for POPhood, and is unrelated to whether the pair really is a POP or not. The threshold T is by definition equal to the total expected number of false POPs, so we choose it to be a small fraction of the number of true POPs, of which we have a shrewd idea of by this stage. For this report, we have set the threshold at 0.35, below 1% of the number of POPs actually found. Because false POPs lead to a proportional negative bias in abundance estimates, the upshot is that we have kept such bias to under 1%.

The resulting set of filtered comparisons is shown in Table 4. At least 11 loci must be compared to get an FPP above the cutoff, and less than 100 11-locus pairs squeeze in; these occur where the 11 happened to be amongst the most powerful[13] of the 25 loci used for the table. On average, the loci used have about a 0.65 chance of *not* sharing an allele by chance, and the table shows very clearly how (near-)binomial probabilities work; from right to left, the numbers in the columns decline rapidly, except for the leftmost column where true POPs appear.

Importantly, in the bottom-left-hand-corner, the Table shows "clear blue water" between the best-matching unrelated pairs (i.e. with fewest loci that do not share an allele) and the true POPs. The separation is less obvious in the rows above say C16, but by looking at how fast the numbers in each row decline from right to left through the F4-F3-F2 columns, it is clear that very few unrelated pairs would have made it into the F0 column. And of course this is what the FPP calculations suggest: given the filtering rule, we would only expect 0.35 spurious POPs in the F0 column. Given that expectation, it is certainly possible that one ($p = 0.25$) or maybe even two ($p = 0.05$) false POPs could have crept in, but very unlikely that false POPs make up an appreciable proportion of the total of 45.

It is also possible to compute an "expected" version of Table 4, assuming there are no true POPs (even thought we have identified 45). That is: for each comparison, taking into account which loci were used, we can compute the probability that there were 0, 1, 2, ... mismatching loci if the pair was truly unrelated. By summing the probability of, say, 1 mismatching loci over all

[13] I.e. genetically more diverse, and being least likely to share an allele by chance

Table 4: Number of *usable* pairwise comparisons, by #loci and #excluding loci. Comparisons are *not usable* if the adult was caught in or before the year of juvenile birth, and/or the false-positive probability was too high (see text). Columns F8-21 and rows C1-10 omitted for brevity.

.	F0	F1	F2	F3	F4	F5	F6	F7	>	F22	F23	F24	F25	TOTAL
C11	.	.	.	.	1	4	5	21	>	.	.	.	.	84
C12	.	.	5	42	340	1345	4019	9114	>	.	.	.	.	57,000
C13	.	1	16	151	887	3420	9900	20482	>	.	.	.	.	143,000
C14	1	4	61	587	2876	11277	32947	70962	>	.	.	.	.	652,000
C15	.	3	42	375	1962	8411	27165	66386	>	.	.	.	.	923,000
C16	2	1	18	131	966	4716	17097	47526	>	.	.	.	.	1,170,000
C17	2	.	8	92	655	3674	14677	45482	>	.	.	.	.	1,942,000
C18	5	.	6	65	483	2699	12037	40524	>	.	.	.	.	3,063,000
C19	7	.	1	33	288	1728	7992	29511	>	.	.	.	.	4,158,000
C20	2	1	1	15	131	886	4630	18722	>	.	.	.	.	5,512,000
C21	14	.	1	5	62	481	2589	11387	>	.	.	.	.	7,197,000
C22	.	.	.	.	4	38	165	698	>	117	.	.	.	1,170,000
C23	4	.	.	.	2	20	143	754	>	2383	179	.	.	2,966,000
C24	2	.	.	.	4	22	90	558	>	17376	2799	214	.	5,097,000
C25	6	.	1	.	1	5	22	199	>	42419	10339	1607	139	4,123,000
SUM	45													38,180,182

comparisons with, say, 11 loci, we can compute the expected value of the (C11, F1) element corresponding to Table 4. The left-hand columns of the result are shown in Table 5, after filtering out the same comparisons as in Table 4. By definition, the row-totals would be the same as in Table 4; the question is how close the column totals are, as shown in the bottom two rows of Table 5. And they are very close, except of course for the F0 column where we are seeing true POPs. This is good; the laws of probability seem to be working well today. The close correspondence between observed and expected totals for F1/F2/F3 suggests that the calculations leading to 0.35 expected false POPs are sound; of course, the *actual* number cannot be exactly 0.35, but it is most likely 0, and most unlikely to be more than 2.

Using a cutoff to exclude ambiguous comparisons does entail a bias-variance trade-off, because some true POPs may have been overlooked in the excluded comparisons, and any reduction in the overall number of POPs found will increase the uncertainty in our final estimates. However, given the threshold we used, it is only when the number of loci compared is 14 or less that substantial numbers of comparisons are excluded (from comparison of Figure 3 and Table 4), and overall only about 5% of comparisons are excluded. Thus we have managed to achieve less than a 1% bias while only incurring a $\sqrt{5} \approx 2\%$ increase in standard error compared to what we would have gotten from "perfect" genotyping (where every pairwise comparison is usable). This reflects very well on the tissue quality, the processing, and the selection of powerful, reliable loci.

Table 5: *Expected* number of comparisons with a given number of mismatching loci, given the loci actually used in each comparison, and assuming no true POPs. The TOT OBS row at the bottom is taken from Table 4.

	F0	F1	F2	F3
C11	.	.	0.02	0.17
C12	0.02	0.63	9.43	82.46
C13	0.04	1.27	17.91	149.85
C14	0.15	4.50	60.38	491.97
C15	0.08	2.56	36.35	315.94
C16	0.03	0.98	15.15	144.87
C17	0.02	0.55	9.16	94.78
C18	0.01	0.30	5.32	58.88
C19	.	0.15	2.75	32.39
C20	.	0.05	1.12	14.42
C21	.	0.02	0.47	6.48
C22	.	.	0.02	0.30
C23	.	.	0.02	0.25
C24	.	.	0.01	0.13
C25	.	.	.	0.03
TOT EXP	0.35	11	158	1392
TOT OBS	45	10	160	1496

Of the 45 POPs found, it is interesting that 9 included one locus where the two animals were scored as different homozygotes (one AA and the other BB). We had deliberately relaxed the exclusion rule to permit this situation, in case of "heritable nulls" (see Appendix, section **??**), and there was no ambiguity about the POP status of these pairs based on the remaining loci[14]. In all but one of the 9 cases the apparent mismatch occurred in one or other of the two loci which exhibited substantial excess homozygosity (D569 and D573; see Appendix 3), consistent with the "heritable null" possibility.

Note also that close-kin relationships at the level of uncles-and-nephews, while possibly as common as POPs in reality, are not going to lead to false POPs in this study. Between an uncle & nephew, only 50% of loci will share an allele by descent anyway, so with these loci the overall chance of sharing an allele is about $1/2*1+1/2*(1-0.65) = 0.68$ (compared to about 0.35 for an unrelated pair), and the chance of getting say 20 loci all sharing an allele through chance is about 0.0004— so there would need to be about 2000 uncle-nephew-level pairs to generate a single false POP.

[14]Including additional checks at the extra one or two loci which were not normally used in mass-screening for POPs

Table 6: Comparison of juveniles to themselves.

	F0	F1	F2	F3	F4	F5	F6	F7
C11	.	.	.	.	.	.	6	9
C12	.	.	8	45	329	1404	4611	10109
C13	1	.	7	63	399	1574	4935	10697
C14	.	1	36	257	1335	5386	15948	35522
C15	.	1	15	153	872	3307	10661	25493
C16	.	1	6	42	304	1465	5341	14986
C17	.	.	2	31	232	1236	4744	14436
C18	.	.	2	26	169	1010	4318	14160
C19	.	.	6	21	144	888	4136	14761
C20	.	.	1	14	85	603	3025	12153
C21	.	.	1	.	37	275	1644	7109
C22	.	.	.	.	.	22	97	524
C23	.	.	.	.	6	14	98	524
C24	.	.	.	1	2	8	69	403
C25	.	.	1	.	1	6	23	115

7.1.2 Cases where no POPs should be found

As an exercise, we can repeat Table 4 just comparing juveniles with themselves, where true POPs are impossible; see Table 6. The expected total in the F0 column is again 0.35; this time, the observed total is 1 (in C13/F0, so towards the lower end of the number of loci compared) which as noted earlier has about a 25% probability and gives no indication that the false-positive calculations are failing.

We can also compare all adults with all other adults (not shown). This time, POPs are actually possible, albeit likely rare because of the time required to reach maturity— see later discussion. There is indeed one possible POP (C18/F0; unlikely to be by chance, given 18 loci used), and it is plausible biologically. The female "parent" was aged 24 when caught in 2007, and the female "offspring" was 177cm (not aged, but any age from 12 up is plausible, given other length-at-age data) when caught in 2009; this gives plenty of scope for the "parent" to have been mature when the offspring was born.

7.1.3 Bounding false negatives?

What about accidentally excluding true POPs? That can only happen if there is genotyping error[15]. Large-scale errors involving multiple loci at once would be (and were) detected and fixed by our QC procedures described in the Appendix 3, so the concern here is about small-scale errors at a single locus and specimen. If such errors lead often to false-negative POPs, these should show up low down

[15]Or mutation, but with say ~50 POPs and ~20 loci each, and mutation rates thought to be about 10^{-4} per generation, mutation is unlikely to have happened amongst our POPs.

in the F1 column of Table 4, as near-POPs that apparently fail to match at one locus (false-negatives at multiple loci being correspondingly rarer). That is not what is seen; rows C17 down have only one entry in F1, compared with an expected total of 1.1 from Table 5.

Prior to producing Table 4, we independently re-scored[16] all the apparent true POPs in F0, all the F1s, and F2s in the rows from C17 down. The original version of Table 4 had 44 rather than 45 POPs; the re-scoring moved one pair from C15/F1 to C17/F0 (changing one existing score, and scoring 2 more loci originally deemed unscorable). The lower left-hand corner of the Table (apart from true POPs in F0) was still empty even without rescoring. Although rescoring changed only about 1 POP, it does give some indication of scoring error rates. Across the 1400 loci that were rescored, there were 8 individual changes, plus deleting one panel of loci for one fish; four of the changes were to delete a score altogether when a locus looked dubious, and the other four were to add a second allele to a "homozygote" (a definite error). Note that all 8-9 changes in the rescoring only unearthed one false-negative (corrected in Table 4), so the *effective* false-negative rate for POP purposes seems to be well under 0.5%. It would also be possible to produce per-locus estimates of scoring error rate based on the partial re-runs and re-used control fish in our QC procedures.

The most important line of evidence to suggest that false negatives from individual scoring errors are not a serious problem, though, remains the absence of entries in the lower left-hand corner of Table 4. Section 4 of appendix 3 presents a formal statistical approach to estimating false-negative rates by comparing Tables 4 and 5; the point estimate of the overall number of remaining false-negatives is in the range 1-2, and the upper 95% CI in the range 2-3. In any event, false negatives must be at most a small proportion of the 45 POPs.

7.1.4 Summary of genetic results

Extensive QC procedures were used to ensure consistent and reliable scoring throughout the project. In all, we conducted about 40,000,000 pairwise comparisons to look for POPs. A few pairs had to be excluded because they had too few scored loci to reliably screen out unrelated pseudo-POPs. However, because of the number and quality of loci used, we were able to choose a cut-off for exclusion that implies very little bias (i.e. unlikely to unearth false POPs) while incurring very little penalty in variance (i.e. using nearly all the comparisons). QC protocols were devised to catch large-scale mixups. With respect to small-scale (individual-level) scoring errors, the error rate is too low to cause a substantial proportion of true POPs to be overlooked. In all, we found 45 POPs in about 38,000,000 usable comparisons.

[16] "Re-scored" means: we re-examined all the peaks and came up with new scores, but did not re-do any of the chemistry.

7.2 Qualitative findings about the POPs

7.2.1 Sex, age and size of parents vs general adults

Of the 45 POPs, 20 were female and 25 male. All adults in POPs were aged; about 1/3 were aged under an Indonesian/Australian ageing program, and the remainder were aged specifically for this project after being identified through genotyping. On average, parents *at capture* are somewhat older (and bigger; not shown) than typical captured adults of the same sex. However, this comparison is not "fair" because the parents have had the opportunity to grow during the interval between juvenile birth and adult capture, which in this study is on average about $3\frac{1}{2}$ years.

Since juvenile age is known (3 in almost all cases), it is easy to back-calculate parental age when the offspring was born. The youngest successful spawners were aged 8, for both sexes. When back-calculated parental age is used instead of age of capture, the difference between parental and typical adult age distribution disappears for females, and actually reverses for males; but it is important to realize that this back-calculated comparison is also not "fair". Adults are subject to selectivity bias in favour of bigger/older fish, and the selectivity pattern on the parents would also have changed over the interval between giving birth and being caught. Back-calculated age distributions will be skewed towards younger/smaller fish, compared to what would have been found if the same set of parents had somehow been sampled in the year of offspring birth.

The upshot of this rather involved argument is:

- parents at capture are older/bigger than typical adults, because they have aged/grown since giving birth;
- back-calculated parental age distribution at offspring birth is similar to typical adult age, but...
- the back-calculated distribution is biassed towards smaller/younger fish, so...
- female parents would actually be bigger than typical adults if it was possible to sample them in the birth-year.
- It's not clear whether the same would be true for males.

These phenomena can only be fully disentangled with the aid of an integrated population model (see below).

7.2.2 Skip-spawning

From the small number of POPs identified in time for CCSBT 2011, there was no obvious indication of skip-spawning (See section 19.2.9, Appendix 6). However, the larger sample of POPs now available does show evidence of biennial spawning for younger fish. The test is to take each POP, and note how many years actually elapsed between juvenile birth and adult recapture, vs how many years *could*

Figure 4: Comparison of age of parents vs adults by sex and at capture vs at birth of offspring

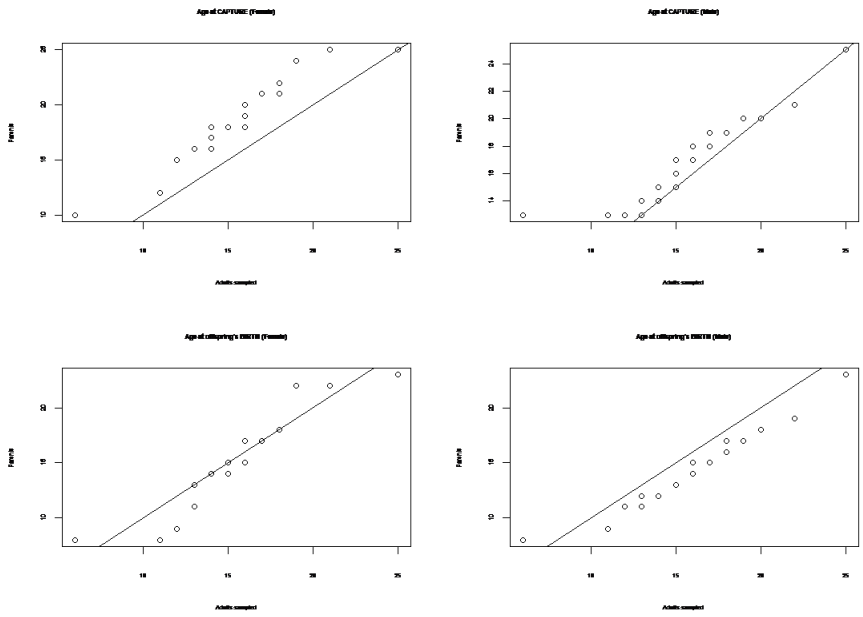


Table 7: Distribution of gap between Juvenile-Birth-Year and Adult-Capture-Year, for young & old parents. Dot means zero. Right-hand table is condensed to odd/even gaps.

Age	Gap (years) ->	1	2	3	4	5	6	7	Gap->	Even	Odd
8-12	Obs	1	6	.	2	.	4	.	Obs	12	1
	Exp	1.6	2.3	2.7	2.6	1.9	1.1	0.6	Exp	6.1	6.9
13-25	Obs	7	5	10	7	2	1	.	Obs	13	19
	Exp	4.3	6.8	7.4	6.2	4.5	1.8	1.0	Exp	14.9	17.1

have elapsed given the POP was eventually found. For example, if the juvenile in a POP was born in 2007, then only comparisons with 2008/2009/2010 adults would be meaningful, so the probability of matching to a 2008 adult is roughly[17] equal to the proportion of adults checked in 2008 relative to those checked in 2008+2009+2010. Table 7 shows the results, split by parental age at offspring's birth; for younger parents, almost all observed gaps are even-numbered, but not for older parents. The pattern is not sex-specific.

Any errors in ageing would obscure patterns such as seen here. Although the sample size is not huge, the difference for younger adults is significant at 1%.

Skip-spawning is not a particular problem for this close-kin study because the study covers many years and the even/odd effect should largely wash out; the general effect of smaller fish being less present is already allowed for in the estimation model, because average spawning-ground residence-time (including the probability of not being on the spawning grounds at all) gets estimated as a function of length and sex. However, in a more perfect world, probabilistic size/age-based skip spawning would be allowed for in the estimation of SSB.

7.2.3 Timing in spawning season

Parents of GAB juveniles have the same distribution of capture date within season as do "average adults" (Figure 5). Thus there is no evidence of temporal correlation in the dates of capture of parents, relative to other adults, that might lead the abundance estimates to be biased (eg we *might* have seen that parents of GAB juveniles always spawn early, and we might *not* have had equal coverage through the Indonesian fishing season). Breaking down by sex (not shown) does not reveal anything either.

7.2.4 Incidence of (half-)siblings among the POP juveniles

There are none. In other words, none of the POP adults match to more than one juvenile. That is a good thing, because if (half-)sibs are common among the *sampled* juveniles, then the pairwise comparisons become non-independent.

[17] Calculations are approximate: e.g. the "expected" rows do not account for growth or mortality, but should reflect any even/odd pattern OK.

Figure 5: QQ plot of day-of-year of capture of Parents (X) vs Adults-in-general (Y)

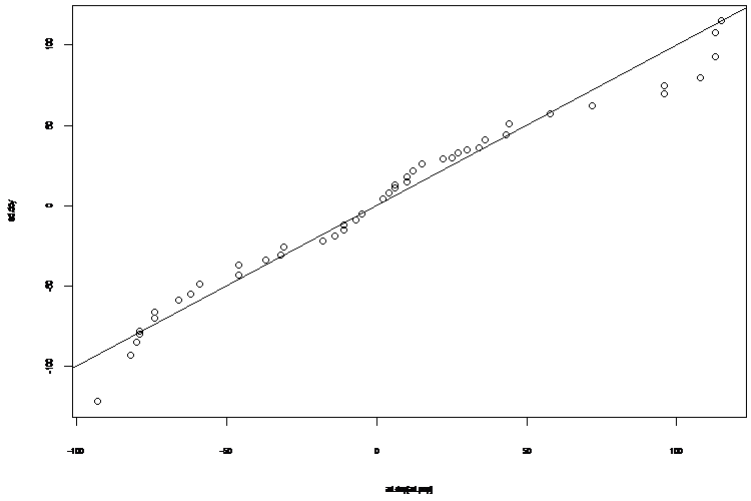


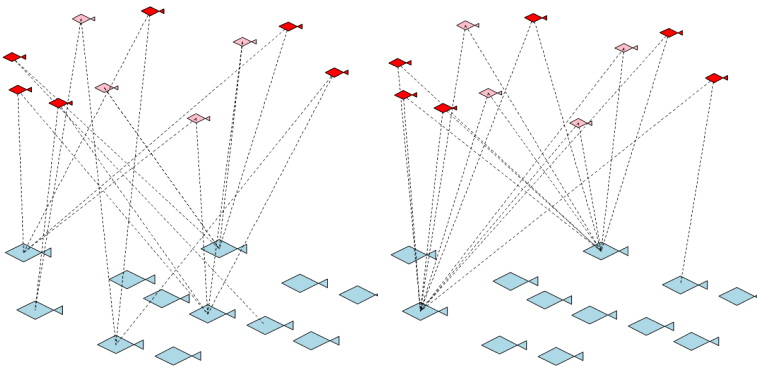
Figure 6 shows what might happen; if there are many (half-)sibs in the juvenile sample, then the number of links to parents remains the same so the abundance estimate is still unbiased (noting that an adult can "count" in more than one POP), but its variance would increase because the number of POPs actually found would depend critically on whether the "super-parents" were caught.

A preliminary check in 2010 (Appendix 6, 19.1) just among juveniles indicated that (half-)sibs could not be *very* common (a critical decision point for the project), and the 7 POPs found in 2010 contained no sibs or half-sibs. Having found none in this much larger set of POPs, we can maybe conclude that (half-)sibs are rare enough *among our juvenile samples* for their effects on variance to be ignored. This is not to say that (half-)sibs are at all rare among *all* 3-year-olds, but simply that our juvenile samples are a very small fraction of the total, and are well-enough-mixed to make sib-pairs rare.

7.3 SBT model estimation results

It will be apparent that an enormous number of different *versions* of the abundance estimation model could be run. A limited set of versions was run for the review by the CCSBT ESC in August 2012. The results presented then were from an almost-steady-state version of the model, with constant adult survival and constant recruitment from 2002 onwards but an age composition in 2002 that need not correspond to a steady-state prior to 2002 (See Bravington et al 2012, Appendix 7).

Figure 6: Cartoon depicting the impact that reproductive variability (high variability on the right and low on the left) would have on close-kin abundance estimate and CV. Small fish are juveniles, red ones are sampled. The number of matches (lines between adults and juveniles) is the same in both cases, but they originate from fewer adults in the high variability case. Hence, the number of estimated POPs (and therefore adult abundance) is the same in each case, but the precision of the estimate would be low in the high variability case (i.e. larger CV), as it would be disproportionately affected by how many of the "super-parents" were sampled.



These initial investigations indicated that:

- Mean L_∞ is appreciably larger for males than females. The evidence for any difference in k or t_0 is not overwhelming, but making these two sex-linked as well does not seem to overparametrise the model. CV of length-at-age appears to be similar for both sexes.
- Residence time appears to be lower for males of a given length than for females, so we do need a sex-specific intercept in this term. However, there is not enough data to estimate any sex difference in the *slope* of the relationship. Also, introducing extra flexibility in model form beyond the logistic (asymptotic) can give nonsensical predictions for very large fish. A good choice seems to be `~sex+length`.
- There is no information for estimating male daily reproductive output as a function of body length. We have assumed instead that male daily output is directly proportional to length (i.e. exponent of 1). There is no good reason for that particular choice, but fortunately the abundance and survival estimates seem not to be much affected by assumptions about male daily output in practice, even though it could matter in theory.
- Based on just one comparison in the initial investigations: changing the annual effective sample size for length/age data from 300 to 900 did not

have a substantial affect the abundance estimates much (i.e. by a few percent); but see below for further examination of impact of effective samples size.

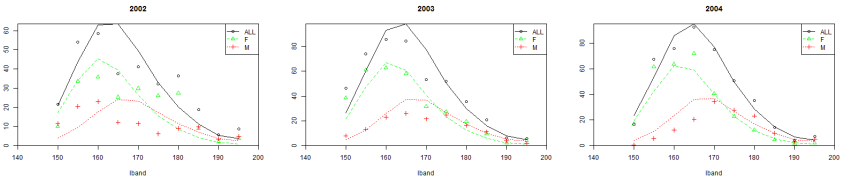
- Annual adult survival for the steady-state model was estimated at 0.73, fairly close to OM estimates. However, the estimated abundance of 10+ adults in 2004 is much higher than in the OM: 2.04M fish, with a biomass of 157kT. This happened to be fairly close to the simple "twice the comparisons divided by the POPs" estimator, but only by coincidence; the competing effects of dilution by incomers, growth, and selectivity are all strong, and merely happen to largely cancel each other out.

These results, were presented to the CCSBT EC in Tokyo, August 2012 (20) in a dedicated plenary session, followed by 2 technical sessions focussed on the Close-kin estimation model (as an independent model) and the preliminary work done on the incorporation of the close-kin results into the CCSBT OM (Hillary et al 2012). The outcome of this review was agreement that the data and number of POPs should be used in the assessment of the stock and a list of additional issues to be examined with respect to the structure of the close-kin estimation model and the ancillary input data (i.e. the size and age data from the spawning grounds). These included:

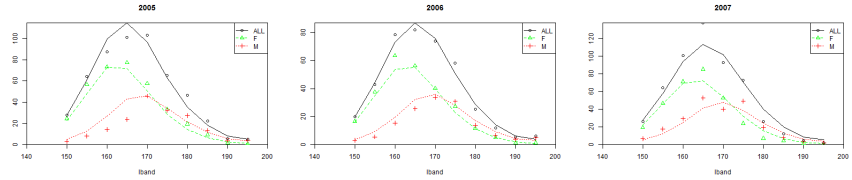
- relaxing the steady state assumptions to explore the implications of trends and variability in recruitment to the spawning population over time;
- the assumption of independence among juvenile samples, which may lead to over-dispersion in the number of POPs and therefore the estimated CV of the resulting abundance estimate.
- the influence of the assumed effective sample size of the length and age data;

These issues have been addressed in the reformulation of the estimation model to incorporate of random effects for recruitment (see section 18.3.2); elaboration of the statistical basis for independence among juveniles samples (see section 17); and estimation of effective sample size external to the abundance estimation model. The results for this revised model are presented below.

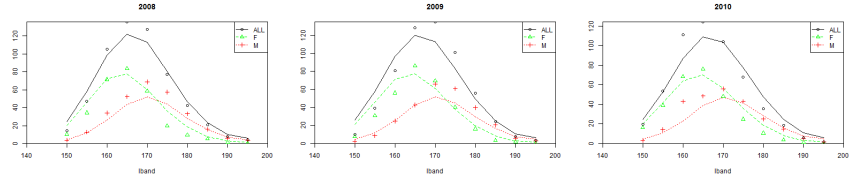
Diagnostic plots for the fit of the estimation model to the length data by year are presented in Figure 7 and by sex-ratio in Figure 8. These are presented for the length and sex data only, since the POP data are really too sparse for diagnostics. The length-frequency data, shown for few years only in Figure 7, are mostly not too bad despite the steady-state assumption, except for 2002 where the data seem completely different from other years. The fits to age-at-length are very good (not shown). However, there is a problem with the fits to sex ratio by length class (Figure 8): in the biggest length classes lower down the graph, where males tend to predominate thanks to their bigger asymptotic length, there is a strong decrease in proportion of females over the 2000s. This decrease is seen overall too (in the black dots), but is not apparent in the smaller



(a)



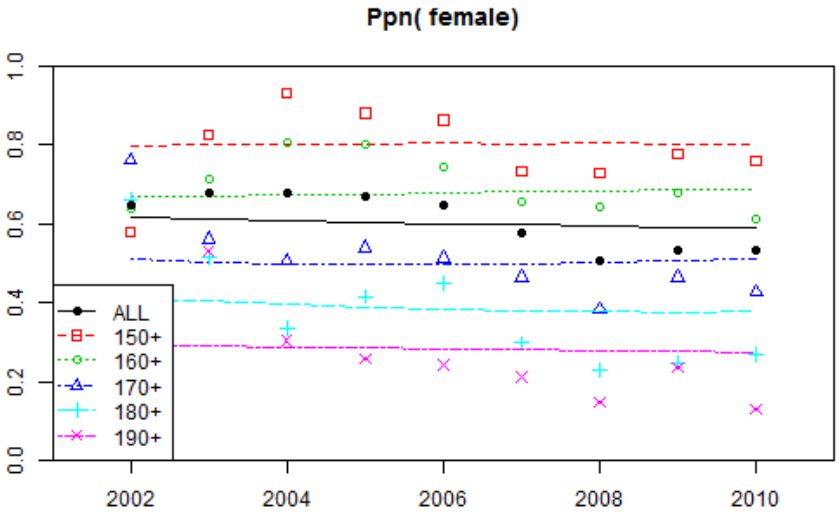
(b)



(c)

Figure 7: Fit of random effects abundance estimation model to length data from spawning grounds by year. Note Y-axis is rescaled sample sizes to reflect estimated effective sample size (see text)

Figure 8: Diagnostic fits to sex-ratio (Proportion Female) by length class and year.



lengths, where there is a rise followed by a dip. This difference in trend across length classes suggests that methodological changes in how sex is assessed are unlikely to be the cause. The underlying cause needs some further thought; it has nothing much to do with close-kin, and is a question for the OM as well as this mini-assessment.

7.3.1 Parameter estimates and uncertainty

Annual adult survival for the revised random effects model was estimated at 0.77 with 90% CI of 0.75-0.8). This is higher than the estimate from the preliminary investigations with the steady state model (0.73) but fairly close to CCSBT OM estimates. The estimated abundance of 10+ adults by year are given in table 8 and estimated 10+ biomass is provided in Table 9. These results are qualitatively similar to the preliminary result for a single year estimate (2.04M 10+ fish, with a biomass of 157kT in 2004) from the steady-state model (20) and much higher (1.87-1.21M fish, or 149,000 - 104,000t, over the 2002-2010 period) than in the most recent estimate from the CCSBT OM. There is a declining trend in the estimated 10+ biomass over the period (14%) which is on the margin of significance at 0.90. Estimates of annual recruitment to the spawning population are given in table 10. These are relatively stable around the average for the period with the exception of the lowest point int he series in 2008 (11). The nominal CVs for abundance of 10+ adults and 8+ "recruits" range from

Table 8: Estimated numbers of 10+yr-old SBT by year over the period covered by the project.

	2002	2003	2004	2005	2006	2007	2008	2009	2010
N (millions)	1.87	1.8	1.73	1.59	1.54	1.52	1.47	1.38	1.21
CV %	16.3	16	15.8	15.7	15.7	15.9	16.2	16.5	16.8

Table 9: Estimated 10+ yr-old biomass of SBT by year over the period covered by the project

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Biomass (kT)	149	145	141	132	128	127	123	116	104
CV %	15.9	15.6	15.4	15.3	15.4	15.5	15.8	16.1	16.3

15.7 - 16.8% and 19.7 - 28.5 (excluding the estimate for 2010), respectively. The theoretical minimum CV for abudance related parameters is 14.9% and is set by the number of observered POPs.

The estimated relationship between residence time (i.e. selectivity, in this model) and length is shown in Figure 9; the curve climbs steeply from 160cm for males and about 155cm for females, with males taking longer to "mature". The apparent asymptotic slowdown around 180cm may be a consequence of the functional form chosen (a logistic curve), and warrants further investigation.

By combining the estimated residence-time with the estimated growth curves (which have average L_∞ of 191cm for females, and 201cm for males) and the fecundity data, it is possible to infer the average female spawning contribution at age. The results are very different to what is assumed in the existing OM, i.e. that spawning contribution is proportional to biomass for ages 10 and up (Figure 10); the abundance estimation model suggests that older fish are comparatively much more effective spawners than younger fish. This highlights the structural difference between the close-kin estimation model and the OM and that the results need to be interpreted and compared with this in mind.

The trend in abundance by age group is provided in Figure 12. It can be seen that the decline in total abundance over the period evident in Table 8 is the result of declines in both young adults and the 25-yr-old plus group, while those age-classes at the peak of their adult lives (16-25 yr-olds) have reamined relatively stable. The combination of the estimates of abundance by age and spawning potential by age (10) allow us to compare the estimated spawning

Table 10: Estimated annual recruitment (numbers 8+ in Millions) and associated CVs

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Recruits	0.561	0.435	0.52	0.546	0.488	0.419	0.231	0.386	0.504
CV %	19.7	20.2	20.2	20.6	21.5	23	26.9	28.5	39.3

Figure 9: Residence time as a function of length by sex

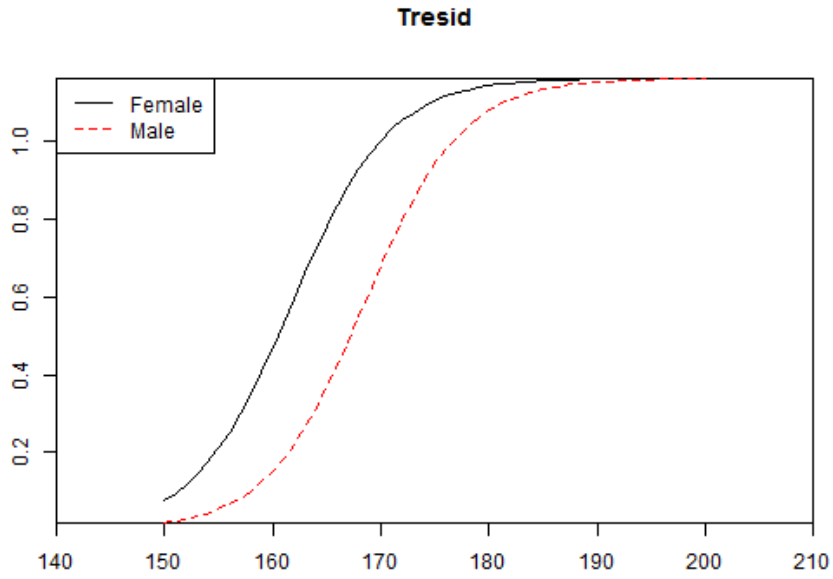


Figure 10: Relative spawning contribution as a function of female bodyweight. Average bodyweight at ages are indicated on closekin estimate (black line). Green line corresponds to current CCSBT OM assumption.

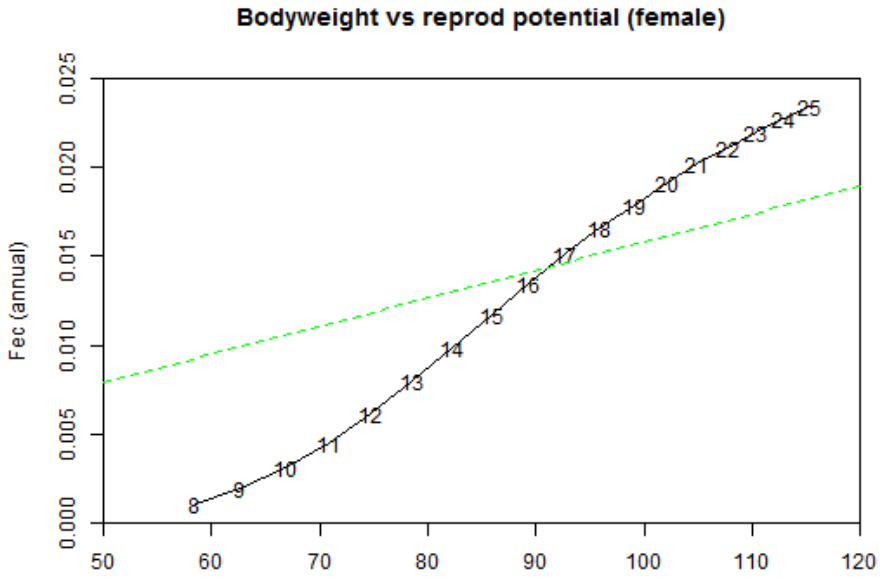


Figure 11: Estimated number of annual recruits to the spawning population by year from 2002-2010. Note the terminal estimates are inherently more uncertain due to the relatively low number of observations

Figure 12: Trends in numbers of SBT by age group (sexes combined)

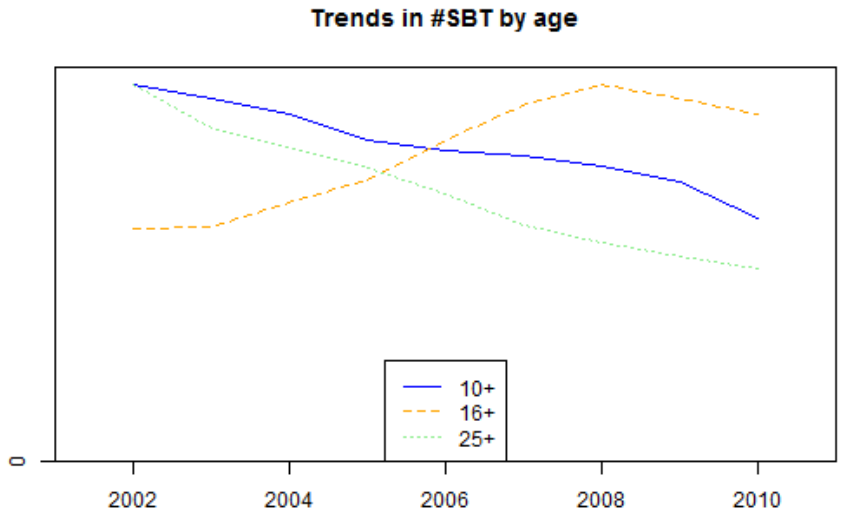
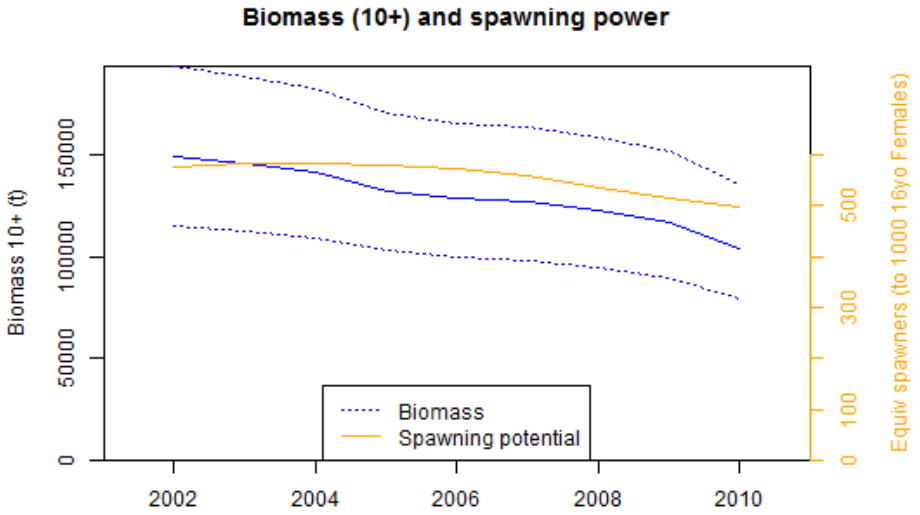


Figure 13: Estimated spawning biomass (10+ biomass as per assumption of current CCSBT OM) and "spawning potential" (as estimated from the close-kin model by year



potential as defined in the close-kin model with the definition of currently used in the CCSBT OM (spawning biomass for 10+). This comparison is shown in Figure 13. The Spawning Potential is a relative measure, in this case the reference point used is the spawning potential of 1000 16-yr-old females, which is close to the age at which 50% of maximum spawning potential is reached based on the close-kin model results. Figure 13 shows that both measures decline over the period covered by this project, although the estimate of spawning potential declines less by the end of the period. This is the result of the estimated disproportionate contribution of larger older females to the total reproductive potential.

8 Discussion

This project has successfully completed an enormous amount of genotyping with tight quality control. The data and estimation model appear to deliver an internally-consistent fishery-independent[18] estimate of adult abundance, with very respectable CV, as originally planned. The stand-alone estimates from the improved estimation model are clearly considerably higher than those from the most recent CCSBT OM; in the order of 3 times the point estimate from the

[18]Strictly: catch and CPUE independent

"base case" scenario, and on the edge of the upper confidence interval of the most optimistic scenario (Anon. 2011). While this may seem surprising, it should be emphasized that there is very little informative data in the OM with which to estimate absolute adult abundance (although other quantities, such as relative depletion, can be estimated more reliably), which of course is in itself was a the primary motivation for this study. The results from the quasi steady-state model were rigorously reviewed by the CCSBT ESC and the high value of the work to the assessment of SBT recognised. Investigation and incorporation of the major issues identified through this review, have improved the robustness of the estimates (i.e. relaxing of the steady state assumptions in the previous version therough inclusion of the variability in recruitment via random effects and external estimation of the effective sample size for the size and age data from the spawning ground) and a time series of estimates of spawning biomass, spawning potential and recruitment to the spawning population independent of the major sources of data included in the CCSBT OM.

As noted in section 7.3.1, the different notions of effective fecundity, or reproductive potential, in the two models (CCSBT OM and close-kin abundance estimation) means direct comparisons between estimates from the two is not straighforward; it is not strictly comparing "apples with apples". Hence, getting an adult abundance estimate that is very different to the CCSBT OM is by no means an indication of a serious problem with either the CK estimate or the main conclusions of the OM. Nevertheless, it is important to ask the obvious question: how wrong could these close kin estimates be? There are a limited number of issues to consider, given that we are not asking about small changes here— the point is to try to think of phenomena that could make a huge reduction to the estimate, of the order of 50%.

8.1 Is the number of POPs about right?

The genetic results strongly suggest that there are few if any false negatives or false positives, given the rigorous filtering criteria we have used, the proven effectiveness of our large-scale QC checks in detecting and resolving problems, and the absence of "near-misses" in the lower left-hand corner of Table 4. An independent implementation of the QC and POP identification software would be useful, and this conceivably might unearth a few further problems. However, while it is certainly possible that there are a small number of false negatives, or false positives in (or not in), our POPs, there is strong evidence that as proportion of the total it is not a substantial number and, hence, is unlikely to effect the substantive results of the project. The QC and POP identifcation procedures and software developed as part of this project represent a major output in their own right. Given the success of this application it would seem a a prudent investment to have these independently reviewed and codified as a basis for future applications both for SBT and for other suitable populations.

8.2 How precise is the estimate?

The nominal CV of the random effects estimate is ranges between 15.7 and 16.8%. This is largely driven by sampling variability in the number of POPs found (14.9%). That ~15% component seems robust, given there is no reason to expect substantial overdispersion in the number of POPs; (see 17). The other significant contributions to the CV are the effective sample sizes for the age and length data from the spawning ground and structural uncertainty in the model. These have been addressed in the revised version of the estimation model through the estimation of effective samples size of the length and age data external to the model and by implimenting recruitment variability via random effects. The additions have not substantially increased the CV for the abundance estimates relative to that from the steady stae model (Bravington et al., 2012). Some comfort can be taken from the fact that the CVs on the estimates of 8-yr-old recruit are consistantly higher than the estimates for 10+.

8.3 Is the abundance estimate about right, given the number of POPs?

If the number of POPs is about right, and if the adult sampling is simultaneous with juvenile birth and random, then the cartoon estimate can't go wrong— each juvenile really does have exactly two parents. Most of the other potential problems with close-kin— stock structure, or massive proportions of sibs/halfsibs— have been demonstrated not to apply to SBT. So the only other source of possible error is in the adult-assessment model. As mentioned at the end of section 7.3.1, the model is not completely finished and the remaining modifications will change the point estimates somewhat, but we do not expect those changes to be very large. So, aside from possible programming mistakes (this is still a very recent assessment, all coded by one person), there are two main points to consider:

1. The entire CK assessment, and the way in which the cartoon adjustments are implicitly calculated, rests on the assumption that selectivity is primarily driven by residence time— the longer a fish is on the spawning grounds, the more likely it is to be caught, all else being equal. The link between residence time and annual female reproductive output rests on the same assumption (more spawning opportunities). It is hard to see how these assumptions could actually be wrong, but the caution might be in the phrase "all else being equal". If there are other really major length-based effects on selectivity or on reproductive output (aside from female daily fecundity, for which we at least have some data), then bias could perhaps arise.

2. The only other way that an abundance estimate could be biassed, is if there is some type of heterogeneity between adults that is *not* just due to length and sex, and which results in some adults (i) being more likely to spawn offspring caught in the GAB, *and* (ii) more likely to be caught

themselves in the Indonesian fishery at least one year later. It is hard to imagine what might cause such heterogeneity.

Even if there do turn out to be errors in these estimates, they seem more likely to be programming errors, and therefore fixable, rather than being intrinsic problems with the data or its interpretation. The CK data fundamentally do seem to be extremely useful for SBT: they are bearing out their promise. There is obvious scope for continuing to collect and genotype in future, both to build up the time series and also (thanks to the retrospective qualities of the close-kin approach) to increase the number of POPs found from our already-genotyped juveniles from 2006-2010. The way this might fit into SBT management, and the links with other monitoring possibilities, is far more than can be explored in this study, but the potential value of further CK genotyping is clear.

8.4 Residence time, spawning behaviour and selectivity

Finally, we draw attention to the key role of residence time and spawning behaviour on the spawning grounds — or, to be accurate, how the average residence time depends on size— in getting to an actual abundance estimate, and a selectivity estimate, and an appropriate definition of spawning potential. Although there is just about enough data in the POPs and the age/length samples to infer the residence/size link indirectly, it would be immensely useful to have direct estimates from a small number of adult fish across different sizes, since this could both ground-truth the model and give a basis for estimating further length-dependent effects on selectivity, if that turned out to be necessary, and the relative time spent actively spawning vs resting on the spawning ground. Pop-up satellite tags could yield limited information quickly, but the best data would come from archival tags because they can record over several years, and are not as vulnerable to short-term tagging shock. The low fishing mortality on adults means that quite a few archival tags would be needed to ensure sufficient recaptures, and that we might have to wait a while to get the tags back, but the number of returned tags needed would not have to be at large to provide a very useful chinsight on, and input to, close-kin based abundance estimates in future. Such tagging ought not be a very expensive exercise in terms of the value of the fisheries, or indeed the cost of this close-kin project. It is important to note that the value of this information would not be restricted to estimation of abundance using close-kin. Similar assumptions and sensitivities apply to the CCSBT OM, either explicitly or implicitly.

9 Benefits and Adoption

The primary beneficiaries of this project are the SBT fishers, managers and policy makers. Additional beneficiaries include conservation NGOs with an interest in SBT, members of the CCSBT and the broader Australian public and the international SBT fleets.

As noted in the introduction and evidenced in Appendices 6 and 7, this project was designed from the outset to maximise the potential for adoption of the results in the assessment and management of SBT at the domestic and international level. The results of the project have been presented and reviewed by the ESC of the CCSBT and the relevant Australian Industry, management and policy bodies (ASBTIA, AFMA, SBTMAC, DAFF, Australian scientific delegation to CCSBT) as they have become available during the life of the project. The involvement and support of these bodies for the approach was central to the projects approval, progress and outcomes. As a result, the major results have already had direct input into the assessment of the stock and are likely to have substantial impacts in the future.

The estimates of absolute adult abundance from the close-kin estimate are ~3 times those from the base-case from the CCSBT OM. The independent estimate of adult mortality is similar to that estimated from the CCSBT OM. As noted in the discussion, the differences in the formulation of the OM and close-kin estimation models described above and, in particular, the definitions of the reproductive potential of the stock, mean that direct comparison between these two estimates is not strictly legitimate. Notwithstanding this, the close-kin estimate does indicate that the absolute abundance of spawning SBT is considerably higher than previously thought. The key question is how this relates to the relative depletion of the spawning stock and its productivity, which is what determines the level of long-term catch and rate of rebuilding. At present, this question can only be answered by incorporating the close-kin data in the CCSBT operating model.

9.1 Incorporation into CCSBT Operating Model

The CCSBT ESC has acknowledged the value of the close-kin data and recommended that further work be done on how best to incorporate the resulting data into the current assessment framework (i.e. the CCSBT OM). Hillary et al. (2012) provided an initial approach to directly incorporating the close-kin data (i.e. the POPs and their associated characteristics) into the CCSBT OM and a preliminary investigation of the implications. These preliminary results show that the addition of the close-kin data into the CCSBT OM, reduces the uncertainy in the trend in spawning biomass and reduces the level of depletion by approximately 40%. That is, the estimated depletion from the most recent assessment is 3-7% of unfished biomass (Anon., 2011); with the addition of the close-kin POP data this range changes to 6-11% (Hillary et al., 2012).

This work was also reviewed by the CCSBT ESC at their 2012 meeting and demonstrated that the close-kin data could be incorporated into the current

OM without major structural modifications. This work also demonstrated that the information from the close-kin data was not inconsistent with other data inputs, in particular the conventional tagging data. While this approach is yet to be adopted by the ESC, it has recommended that additional work be done in preparation for the next assessment of the stock, which is scheduled for 2014. Further explorations of the incorporation of the close-kin data will be considered by the CCSBT Operating Model and Management Procedure Working Group in Portland, Maine, in July 2013 and subsequently by the ESC at their annual meeting in Canberra, September 2013, where the decision on whether and how to incorporate the close-kin data will be made.

9.2 Implications for assessment of CCSBT

The implications of the assesment of CCSBT can be considered from two perspectives: short term and medium term. In the short-term, the close-kin data will be incorporated into the existing CCSBT OM, as outlined above, and the questions that the data and independent close-kin estimation model raise in terms of the current explicit and implicit assumptions in the formulation of the OM will be investigated and resolved, in one way or another, as part of the 2014 assessment process. As noted in the discussion, the definition of spawning potential and the interaction between natural mortality, spawning abundance and steepness of the stock recruitment relationship are likely to be key issues.

In the medium term, the results of this project demonstrate the feasibility of monitoring the spawning stock directly (as opposed to estimting trends via the OM) over time by maintaining the collection and periodic processing of samples from the spawning ground and Port Lincoln. This would provide a two-fold benefit: provide an additional fisheries independent data stream (POPs over time) as an input to the periodic updates of the OM; but more importantly, it would provide the basis of monitoring abundance of the spawning stock (and rebuilding, or not) that is independent of the fishery and independent of the known and unquantifiable biases (resulting from the unreported catches) in the longline CPUE index which dominates the trends in the current OM.

10 Further Development

This project has demonstrated it is possible to provide a robust and precise estimate of abundance and estimate of mortality of the spawning stock for a highly migratory and highly valued pelagic stock that is (almost) completely fishery independent. As such, it has exciting potential future applications to the monitoring and assessment of SBT, as well as other highly migratory, or "hard to observe" populations.

10.1 SBT

10.1.1 Close-kin as a monitoing series

Future developments for SBT have largely been covered in the Discussion and Flow of Benefits sections.

These include:

- the incorporation of the close-kin data into the current CCSBT OM;
- continued collection of samples to develop a time-series for monitoring the spawning stock directly, and;
- a design study to determine the optimum number and frequency of sampling of adults and juveniles, to ensure cost-efficiency of the approach.

10.1.2 Genetagging to estimate fishing mortality and recruitment strength

In addition to demonstrating the power of the close-kin approach, there are other indirect benefits and potential future developments.

A considerable proportion of the total budget of the preceding CSIRO proof of concept project and this project was allocated to the development and optimisation of the micro-satellite markers and their "validation" for SBT. This was central to high quality matching of parents and offspring. In addition to identifying POPs, these same markers can be used to match individuals to themselves (i.e. genetic mark-recapture). Hence, a substantial proportion of the upfront cost that would normally be associated with trailing or undertaking a genetagging project for SBT has already been met.

The value of mark-recapture estimates of fishing and natural mortality, growth and movement have already been demonstrated for SBT and other pelagic stocks. In the case of SBT, it would be feasible to estimate year class strength and fishing mortality for the surface and long-line sectors of the fishery and to do so in a way that did not rely on estimates of tag shedding or reporting rates. If cost-effectively implemented, the combination of close-kin abundance estimation (of the spawning stock) and gene-tagging (for fishing mortality and recruitment) has the potential to provide fisheries independent monitoring of each of the main components of the stock. It is highly likely, given the continuing reductions in the cost of large-scale genotyping, that this combination of approaches would be more cost-effective and sustainable, from a logistics perspective, than the current arrangements for the fishery.

10.2 Other Species

There is considerable potential for the application of the close-kin methodology to other species/populations that are otherwise difficult to monitor and/or observe. Careful consideration is required, however, of the populations life-cycle and stock stucture and appropriate sampling designs to determine whether

the approach is feasible and/or likely to be cost-effective. Currently, there are projects underway to estimate abundance of sawshark and white sharks using variations of this method.

11 Planned Outcomes

The planned outcomes from this project were to:

1. provide an independent check of the assessment model, which are entirely reliant on fishery-dependent data;
2. provide for the incorporation of the SSB estimates (from this project) into the assessment;
3. reduce the uncertainty in the current assessment, and;
4. provide an independent benchmark to measure rebuilding of the stock.

It was also an expectation that the work on improving the definition of "spawning stock biomass" would lead to a better understanding of stock productivity, the relative importance of different age classes to total reproductive capacity of the stock and likelihood of different rebuilding trajectories.

Each of these outcomes have been achieved, albeit with a necessary time and budget extension to the original project.

12 Conclusion

Cost-effective, accurate and reasonably precise methods for monioring and assessment of highly migratory species, such as tuna, has challenged fisheries scientists and managers for many decades. While not a panacea for all species, this project has demonstrated that the close-kin approach can provide cost effective and precise (CV of ~ 20% or less) estimates of the spawning abundance of SBT and, in the process, estimates of mortality and age specific spawning potential. The results indicate that the absolute spawning biomass is considerably higher (~3x) than the current estimate from the CCSBT OM. Direct comparisons between the two model estimates are not strictly legitimate, however, given the structural differences between the two. Preliminary results from including the close-kin data (POPs and their associated data) in the CCSBT OM (Hillary et al., 2012b) indicate that they are likely to substantially reduce the uncertainty in the trend of spawning biomass over the most recent decades and that the estimated level of depletion is likely to be reduced (from ~ 3-7% without the close-kin data to ~6-11% with the close-kin data included). The quality of these results clearly demonstrate the potential of the method for ongoing monitoring of the SBT spawning stock and the potential to extend the utility of the population genetic methods and protocols developed through this project to genetagging of the harvested components of the stock (i.e. 2-10 yr-olds). This

potential to provide fishery independent monitoring of the each of the each of the main components of the population would remove the uncertainty associated with longline CPUE as an index of abundance and improve the confidence of all stakeholders in the monitoring and assessment of this valuable stock.

13 References

Anonymous. 2006. CCSBT Report of the Eleventh Meeting of the Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna. 12-15 September, Tokyo, Japan.

Anonymous. (2006). CCSBT Report of the Thirteenth Annual Meeting of the Commission for the Conservation of Southern Bluefin Tuna. 10 – 13 October, Miyazaki, Japan.

Anonymous. (2006). CCSBT Report of the Special Meeting of the Commission for the Conservation of Southern Bluefin Tuna. 18-19 July, Canberra, Australia.

Anonymous. (2008). Report of the Thirteenth Meeting of the Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna, 5-12 September, Rotorua, New Zealand.

Anonymous. (2009). Report of the Fourteenth Meeting of the Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna, 5-11 September, Busan, Korea.

Anonymous. (2011a). Report of the Sixteenth meeting of the Extended Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna, 18-28 August, Bali, Indonesia.

Anonymous. (2011b). Report of the Eithteenth Annual Meeting of the Commission for the Conservation of Southern Bluefin Tuna, 10-13 October, Bali, Indonesia.

Anonymous. (2012). Specifications of the CCSBT Management Procedure. Appendix 7 of Report of the Seventeenth meeting of the Extended Scientific Committee of the Commission for the Conservation of Southern Bluefin Tuna, 27-31 August, Tokyo, Japan.

Bancroft (1949) available at: http://projecteuclid.org/euclid.aoms/1177729999

Davis T, Farley J, Bravington MV, Andamari R (2003). Size at first maturity and recruitment into egg production of southern bluefin tuna. FRDC Final Report project 1999/106.

Eveson P, Farley J, Bravington M (2012). The aerial survey index of abundance: updated analysis methods and results for the 2011/12 fishing season. Prepared for the 17th Meeting of the Scientific Committee, Commission for the Conservation of Southern Bluefin Tuna, 27-31 August 2012, Tokyo, Japan. CCSBT-ESC/1208/16.

Farley J, Andamari R, Proctor C (2007). Update on the length and age distribution of SBT in the Indonesian longline catch. Prepared for the CCSBT 8th Meeting of the Stock Assessment Group and the 12th Meeting of the

Extended Scientific Committee 4-8 September, and 10-14 September 2007, Hobart, Australia. CCSBT-ESC/0709/10.

Farley J , Eveson P, Clear, N (2012). An update on Australian otolith collection activities, direct ageing and length at age keys for the Australian surface fishery. Prepared for the 17th Meeting of the Scientific Committee, Commission for the Conservation of Southern Bluefin Tuna, 27-31 August 2012, Tokyo, Japan. CCSBT-ESC/1208/18.

L. Hascoet, V. Pascual, et al. Tapenade software. INRIA, project Tropics, www- sop.inria.fr/tropics, 2012.

Hillary R, Basson M, Davies C, Eveson P (2009). Further consideration of the potential for management procedures for SBT based on fishery independent indicators - short-term options using relative indices from the aerial survey and conventional tagging. Prepared for the Extended Scientific Committee for the Fourteenth Meeting of the Scientific Committee. CCSBT-ESC/0909/22.

Hillary R, Preece A, Davies C (2012). Developing a management procedure based recovery plan for Southern Bluefin Tuna. FRDC Final Report project 2011/034.

Maunder MN, Sibert, JR, Fonteneau, A, Hampton J, Kleiber P, Harley SJ, (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. Marine Science 63(8): 1373-1385.

Polacheck T (2006). Tuna longline catch rates in the Indian Ocean: Did industrial fishing result in a 90% rapid decline in the abundance of large predatory species? Marine Policy 30: 470-482.

Polacheck T, Eveson JP, Laslett GM (2006). Estimation of mortality rates from tagging data for pelagic fisheries: analysis and experimental design. FRDC Final Report project 2002/015.

Polacheck T, Davies C (2008). Considerations of implications of large unreported catches of Southern Bluefin Tuna for assessments of tropical tunas, and the need for independent verification of catch and effort statistics. CSIRO Marine and Atmospheric Research Paper 023, 2008.

Polacheck, T (2012). Politics and independent scientific advice in RFMO processes: A case study of crossing boundaries. Marine Policy 36: 132-141.

Proctor CH, Andamari R, Retnowati D, Herrera M, Poisson F, Fujiwara, S, Davis TLO (2006). The catch of SBT by the Indonesian longline fishery operating out of Benoa, Bali in 2005. Report for the 11th CCSBT Extended Scientific Meeting 4-11 September, Tokyo, Japan. CCSBT-ESC/0609/10.

Sibert J, Hampton J, Klieber P, Maunder M (2006). Fishery induced changes in biomass, size structure and tropic status of top-level predators in the Pacific Ocean. Science 314: 1773 – 1776.

Skaug HJ (2001) Allele-sharing methods for estimation of population size. Biometrics 57: 750–756

14 Appendix 1: Intellectual Property

15 Appendix 2: Staff

Dr Mark Bavington
- Dr Peter Grewe
- Dr Campbell Davies
- Ms Peta Hill
- Dr Rasanthi Gunasekara
- Mr Matt Lansdell
- Ms Naomi Clear
- Mr Thor Carter
- Mr Scott Cooper
- Ms Danielle Lalonde
- Mr Mark Green
- Mr Bruce Barker

16 Appendix 3: Genotyping and Quality Control

16.1 Terminology

This section is meant as a guide for a non-geneticist. It is not intended as an authoritative set of definitions from a genetic perspective, which are widely available in the genetics literature. These definitions include forward-cross-references in *italics*, and use **bold** to indicate additional definitions. In the text after this section, a few technical genetic terms have been used and marked with an asterisk, but deliberately not defined since their relevance will only be apparent to those who already understand them.

Locus: an identifiable place on the genome with characteristic start and end sequences of DNA, and a variable DNA sequence between them. The loci we used are **diploid,** so that each individual has two versions (**copies**), one copy being inherited from each parent. The sequences of the two copies might be different or might by chance be the same. We used **microsatellite** loci, whereby each sequence is characterized simply by its length or **size** (i.e. the number of *nucleotide bases* it contains), which will be a integer in the range say 80-600 depending on the locus and how it is to be purified away from the rest of the genome in any particular study.

Alleles: the set of possible sequences a locus can have, i.e. for microsatellites a set of integers. Alleles at different loci might happen to have the same length, but are in no sense comparable— it only makes sense to refer to an allele for a specific locus. The **allele frequency** for the locus is the frequency distribution of the different alleles across the population under study. A **highly variable** locus has a large number of different alleles and an allele frequency that is not dominated by just one or two common alleles. The probability that two unrelated animals will have an allele in common is lowest if the locus is highly variable, so such loci are preferred for close-kin work. A **null** allele is an allele that is present in the animal, but is not revealed by genotyping; possible causes include scoring error, and a mutation in or near the locus that causes the DNA amplification process to fail for that copy.

Genotype: which alleles an animal has. Usually means for all the loci together (sometimes called a **multilocus genotype** or **DNA fingerprint**), but can mean just the alleles carried at a single locus if specified.

Homozygote/heterozygote: An animal is said to be a **heterozygote** at some locus if the two copies are different alleles, or a **homozygote** if they are the same.

Scoring/genotyping/calling: deciding which alleles are present at a locus for a particular animal. This really involves many steps, but sometimes "scoring" just refers to the final step of adjudicating on the possible alleles proposed by the *GeneMapper* software. The protocol in our study is that,

if the genotype at a particular locus for a particular specimen is ambiguous, no score is recorded (rather than trying to make a subjective best-guess).

Scoring error: Recording the wrong genotype at one or more loci. Large-scale scoring errors affecting many fish and loci simultaneously can arise from inadvertently swapping or rotating entire plates of fish, or from miscalibration of the *sequencer* for a particular *run plate*. Small-scale scoring errors affecting individuals most commonly involve failure to detect a small second peak in a heterozygote, so that the locus is mistakenly scored as a homozygote instead. At least in this study, actual mislocation of peaks were very rare (based on a subset of the fish which were independently re-scored).

Amplification/PCR: the chemical process by which the DNA from certain desired loci only is selected and amplified for input to the sequencer.

(Nucleotide) base is one genetic "letter" (C/G/A/T), the molecular building-blocks which are linked together to form a DNA molecule. DNA occurs in two strands, and each base is paired with its complement on the other strand, so the term **base pair** is often used instead.

Tetranucleotide: The sequences within microsatellite loci are mostly repeats of some short subsequence of base pairs, such as GATA (four base pairs, so a tetranucleotide locus) or CA (a dinucleotide). Dinucleotide loci are more common in most genetic studies, but are more prone to scoring error. In this study we used only tetranucleotide loci.

Panel refers to a set of loci (usually 4-7) which can be analyzed simultaneously by the sequencer.

Plate is a group of 96 DNA samples (including a couple of controls— standard specimens included on every plate— and blanks) placed in wells numbered A1-H12 in an industry-standard format on a small rectangular tray ("**96-well microtitre plate**") ready to load into a sequencer. Each group of 96 fish is originally set up on a **template plate** from which are prepared several **run plates**, all with the same layout of specimens in the 96 wells, but with each run plate specific to a particular panel of loci.

Sequencers are the machines that physically do the genotyping. One run plate is **run** or **sequenced** at a time. For each locus on each specimen, the output is a graph with X-axis corresponding to allele length (as a continuous variable) and "signal intensity" on the Y-axis. Alleles are visible as peaks with a characteristic shape.

GeneMapper is software which identifies possible alleles from the sequencer's output. In most cases, GeneMapper will propose the correct peaks, but each sequencer graph and proposed scores is scrutinized by an experienced scorer who makes the final decision on which peaks truly represent alleles, and which peaks are artefacts.

Bins and binsets: Because of slight variations in run conditions, the locations of peaks reported by GeneMapper will vary fractionally between sequencer runs, even for the same sample. **Bins** are therefore used in GeneMapper to provide tolerance and to convert the continuous-valued peak locations into an integer-valued allele size. Each bin is a continuous-valued range such as $[137.2, 138.6]$, which should span the range of peak locations found for that allele across many runs. The **binset** for each locus is the collection of all its bins. The binset needs to be consistent throughout a study. In this project, we initially developed bins and binsets from genotyping the first 500 individuals, then revisited them after 5000 specimens had been genotyped.

16.2 QC for Consistency of Allele Size Calling

Examining the consistency of allele-size calling is fairly straightforward, and is mostly dealt with by use of an internal standard and use of an automated genotyping program developed by ABI-Life Technologies (supplier of the DNA sequencer used for fragment separation). To further minimise inter-run variation, all size fragmentations were run on only one DNA sequencer located at the Australian Genomic Research Facility (Adelaide node). This eliminated variation occasionally observed when the same samples are run at two facilities even on the same model of sequencer.

In addition, the ABI system uses an internal size standard added to each sample from which the size curve is extrapolated for estimating allele peak length relative to the standard curve. ABI states that variation using this system ensures +/- 0.5bp accuracy from run to run. Furthermore, the GeneMapper program analyses each individual size curve for peak quality and general fit to the theoretical ideal size curve. Any discrepancies detected by the software raise flags in the analysis window and can be scrutinized in further detail. We also examined each size curve analysis as well as the individual peaks that were used to generate the size curve for each individual in a run plate to ensure another level of QC in addition to that used by the GeneMapper software.

GeneMapper uses a standard set of allele size bins used to smooth out further subtle variation and ensured easy comparison among alleles from different individuals and provided another level of QC among plates. Bin sets are developed for each locus to permit automated genotyping using the GeneMapper software. Individual bins represent a value range centred on the median length value of each allele as ascertained following sizing of an initial set of individuals. Preliminary bin sets were developed following detailed analysis of about 500 fish. These sets were designed to encompass slight variations to permit detection of gross deviations from the norm greater than +/- 1.0 bp. After genotyping about 5000 fish, the bin sets were re-assessed for consistent allele calls, and a final consensus adjustment was determined. Bins permit assignment of an integer value to the continuous-valued allele length based on the GENESCAN size standard, and permit simple comparison of allele identities among individual genotypes. A gap of one to three base pairs between bins ensures that an objective decision

rule can be consistently applied to a genotype for inclusion of an allele into a designated integer bin. Alleles falling in the gap were rare and presumed to be a result of an insertion or deletion event on an individual's DNA. These were scored as "unknown genotype" but the real value could still be used for confirmation of parentage should it be required to confirm identity (not required with our samples to date).

The use of automated genotyping with a single set of GeneMapper bin-sets allowed us to detect if peaks were consistently falling outside of predetermined bins and would highlight a general problem with the running of a plate (eg. old buffer or polymer in the sequencer leading to general failure of proper electrophoresis and inconsistent separation). Runs where problems were found were re-run with new buffer and polymer; this rectified the problems in every case.

16.3 Avoidance of chimeras

Chimeric genotypes are (in this study) a composition of DNA from more than one fish, rather than (as in some other studies) DNA profiles resulting from multiple DNA in a well (two or more contaminated DNA leading to more than two alleles present for each locus). There are only two possible sources. First, a chimeric error will result from turning a run plate 180 degrees, whereby e.g. the A1 position became the H12 position. This error produces what looks like a legitimate DNA profile but made up of some loci from fish A1 mixed with the remainder of loci from H12 from the run plates that were not rotated. Second, if two run plates are swapped, the loci for those panels (but not for the other panels on the same fish) will be swapped. Clearly, these errors will lead to any POP members on the plate being overlooked, affecting 100-200 fish at a time, so it is important to catch them. Fortunately, once one is aware of these possibilities, it is fairly easy to write QC software using the check-plate results and/or the controls to detect and fix the problem. We did find both types of chimera in this study (rarely), but thanks to the QC protocols we were able to detect and fix them.

16.3.1 Further processing details for the first 5000 fish

For the first 5000 fish we developed a unique system to cope with the potential issues arising from PCR and fragment separation methods used at the outsourcing facility (AGRF). The first 5000 fish were run at AGRF as three single-plex (A, B, and C) and two multiplex (D and E) panels. At this point the multiplex PCR was clearly the most optimal solution and we included 7 additional loci that were incorporated into an optimised set of four multiplexed panels (H, I, J, and L). The A, B, and C panels were combined into the I and J panels while D and E were combined into H and L. To check for generation of chimeric genotypes we used the set of template plates that were the source of DNA for the D and E panels. Since D and E had a common locus scored for both plates we were able to ensure that there were no chimeric individuals there. We then ran the first column of each template plate for panel-I and for panel-J. This

checked the genotype calls of 8 individuals that should be identical if no mix up had occurred. We verified that all 8 genotypes for each locus was congruent across all tested plates indicating that no single-plex mixups had occurred. Since the template plates used were those used to set up D and E we were then assured that there were no chimeric fish generated in the first 5000 genotyped individuals.

16.3.2 Further processing details for the last 9000 fish

For the balance of the fish, a unique system to identify individual template plates was developed to ensure that the fish on the plate could be identified, and that it was not accidentally rotated prior to sequencing. The four panels had a common locus to check on plate to plate variation, and also to detect PCR contamination via negative water controls. Template plates were created in a specific routine fashion with four positions in each plate reserved for positive and negative controls. We used two positive control individuals on every plate with position A01 being control fish #1 (TC-2005, male) and G12 being control fish #2 (TC-2205, female). The positions of the negative water controls were used to uniquely identify each plate. For example, one plate would have water controls in position A02 and A07, while the next plate would have A02 and A09. Care was taken to ensure that the water was placed in one odd-numbered and one even-numbered well row due to the way the 48 capillary sequencer picked up the samples; every dip of the sequencer thereby had one positive and one negative control, so that each electrophoresis had internal controls to check run quality. The internal common locus control for each individual checked to see that each fish was scored with consistent fragment separation for each of the panels. By use of this system for the final 9000 fish, we were able to QC for chimeric individuals, check for PCR contaminants in the master mix, ensure that run conditions did not affect genotype scoring among the four panels, and also ensure that plates were not mislabelled or loaded into the sequencer incorrectly. Our QC caught a few errors but these were few and subsequently dealt with by a quick rerun of the PCR or fragment separation or both.

16.4 Rigorous estimation of false-negative (FN) rates

The question of interest is: what proportion of true POPs could have a scoring error that leads to the POP being overlooked? We can estimate this directly by comparing Table 3— observed numbers of (loci compared, loci failing to match)— with Table 4 (expected version of Table 3, assuming zero POPs and therefore zero FNs). If the expected-value calculations behind Table 4 are correct, and if there are numerous true POPs without FNs, then Table 3 should resemble Table 4 except for numerous entries in the F0 column— which is pretty much the case. If the Table 4 calculations were wrong for some reason[19], then

[19]The only theoretical reason we can see why the calculations in Table 4 might ever go wrong, is if genotypes at different loci within each fish are not independent, something which could arise from substantial cryptic stock structure, with different allele frequencies in the different

the upper-right-hand triangle of numbers in Table 3 would be stretched to the left compared to Table 4— which is not the case. Therefore, we can take the expected values in Table 4 as correct if there were no POPs, and use the differences between the tables to make inferences about the true number of POPs, and about how many FNs are in Table 3. We can do this because FNs will appear in Table 3 as an "echo" of the F0 column, predominantly in column F1, and somewhat weighted towards the lower rows because there is more chance of a scoring error when more loci are involved. Apart from chimeras and mass failures of PCR on a run plate, as described and ruled out in Appendix 1, there seems no reason why scoring errors should not be independent across loci on the same fish; hence, provided scoring errors are uncommon to begin with, FNs are most likely to be in the F1 column, less likely to be in F2, and rapidly less likely beyond that.

The numbers in Table 3 actually result from a second round of checking; we re-scored all the pairs in the F0 and F1 column, and in the lower rows of the F2 column. However, only a small percentage of the fish were re-scored during the second round, and the level of attention paid to these fish may not be typical of the rest of the sample. In this section, we have therefore analysed the data from the *preliminary* version of Table 3, before any selection of fish to re-score took place. This makes the analysis general, but also means that the results are pessimistic in terms of FN likely FNs compared to the final data, because the FN/near-FP status of many would have been cleaned up during re-scoring. The preliminary data, shown in 11, is very similar to Table 3, the main difference being that the C23 row starts (3,1) rather than (4,0); this is one case where a scoring error did cause a false-negative, though this was subsequently detected and fixed on re-scoring. The other differences did not affect POP status of any pairs.

16.5 Likelihood for estimating false-negative rate

Let θ be the probability that a pair of fish will be a POP (so θ is inversely related to abundance, etc), and let e be the probability that one shared locus in a POP will fail the parent-offspring compatibility test[20], either through mis-scoring or mutation. Assuming scoring errors at different loci are independent[21] and equally likely[22], then the probability of f loci failing in a POP where c loci are compared, is a simple Binomial probability. Also, for a non-POP pair where

stocks. That situation is *a priori* unlikely for SBT, and happily there is no suggestion of it in Table 3.

[20] The basic test is: do they share a visible allele? We used a more relaxed version, so that AA vs BB homozygotes are also deemed (potentially) compatible.

[21] Apart from chimeras, as described and ruled out in Appendix 1, and mass failures of PCR on a run plate which would be picked up by our other QC checks, there seems no reason why independence could fail.

[22] Strictly, the probability of a scoring error that leads to rejection of POPhood probably varies somewhat across loci, but there is not nearly enough data to estimate this; and since the set of loci that actually get used in a comparison is a random variable, and we are only concerned with one or two errors here, the approximation is statistically negligible.

Table 11: **Preliminary** number of *usable* pairwise comparisons, by #loci and #excluding loci, **before** re-scoring. First three columns only.

.	F0	F1	F2
C11	.	.	.
C12	.	.	5
C13	.	2	16
C14	1	4	61
C15	.	3	42
C16	1	1	18
C17	3	.	7
C18	5	.	7
C19	7	.	1
C20	2	1	1
C21	14	.	1
C22	.	.	.
C23	3	1	.
C24	2	.	.
C25	6	.	1
SUM	44		.

c loci are being compared, let p_{cf}^{NON} be the probability that f of the loci will fail the test. For any given pair, this actually depends on the particular loci involved, and is already calculated to form the basis for the expected values in Table 4. Any given pair with c loci compared is either a POP or not, and the probability p_{cf} that the pair will fail at f loci is therefore

$$p_{cf} = \theta \left(\begin{array}{c} c \\ f \end{array} \right) e^f \, (1-e)^{c-f} + (1-\theta) \, p_{cf}^{\mathrm{NON}}$$

Therefore, if n_c denotes the number of comparisons using c loci in Table 3, the expected value of cell (c, f) is $n_c p_{cf}$. Strictly, the distribution within each row is Multinomial, but in the first few columns the multinomial "size" is enormous (millions) and p_{cf} is small, so a Poisson approximation is perfectly adequate. If y_{cf} denotes the observed number of pairs in the (c, f) entry of Table 3, then the likelihood of the first few columns up to F failures is (up to a constant)

$$\prod_{c=11}^{25} \prod_{f=0}^{F} e^{-n_{cf} p_{cf}} \, (n_{cf} p_{cf})^{y_{cf}}$$

The term p_{cf} involves the parameters θ and e, which can be estimated via maximum likelihood.

The bulk of the information on false-negative rates is contained in the F1 column (and the F0 column, which is needed for estimating θ), with a little coming from the F2 column. To the right, the noise from the increasingly large numbers of almost-false-positives swamps any signal related to false-negatives with 2, 3, etc number of failures, which will be increasingly rare.

16.6 Confidence intervals on actual FNs

Although the Hessian from the above likelihood could be used in the standard way to derive a confidence interval for the *expected* number of FNs in a *replicate* of this study, that would be solving the wrong problem. Our interest lies in the *actual* number in *this* study; so, if FNs were very unlikely beyond the F1 column, then the number of FNs would be capped above by the total F1s seen, regardless of how many might be found if the study was repeated. This makes quite a difference in practice. A Bayesian argument is required to get the answer we need.

We need the probability distribution of the number of false-negatives $\#FN$ given the observed data, i.e. $\mathbb{P}[\#FN|y]$ where $\#FN$ is the total number of False Negatives and $y = (y_{cf} : c \in 11 \cdots 25,\ f \in 0 \cdots 1)$ is the observed numbers in the F0 and F1 and possibly F2 columns (F3 onward are irrelevant because the chances of 3 or more scoring errors is negligible). For simplicity of argument, say for now that we neglect the F2 column as well. Obviously, the maximum possible value of $\#FN$ is the observed number of F1s, in this case 12. Each of these F1 pairs is either a near-FP or an FN. The probability that an F1 pair with c loci compared is actually a FN rather than a near-FP, is

$$\frac{\mathbb{P}[\text{1 error in } c \text{ loci}] \times \mathbb{P}[\text{is POP}]}{\mathbb{P}[\text{1 error in } c \text{ loci}] \times \mathbb{P}[\text{is POP}] + \mathbb{P}[\text{match at } c-1 \text{ of } c \text{ loci}] \times \mathbb{P}[\text{is not POP}]}$$

One implication is that a (C12,F1) fish is much more likely to be a near-FP than a (C25,F1) is, because (i) the probability of a non-POP matching by chance at 11 of 12 loci is much higher than for 24 of 25, and (ii) the chance of a scoring error is about twice as high with 25 loci as with 12.

The FN-status of the pairs are independent[23], θ and e, so the total number of F1 pairs that are FNs is the sum of (in this case) 12 independent Bernoulli (0/1) random variables, with probabilities depending on the number of loci involved. There is an algorithm for calculating the Bernoulli-sum probability distribution, which is already used in the expected-FP calculations[24]. Hence, given a pair of values (θ^*, e^*), we can easily compute $\mathbb{P}[\#FN = x|y,\theta^*, e^*]$ for $x \in 0 \cdots 12$. What we actually need, though, is

$$\mathbb{P}[\#FN = x|y] = \int \mathbb{P}[\#FN = x|\theta, e, y]\, f(\theta, e|y)\, d(\theta, e)$$

which can be estimated by repeatedly drawing pairs $\left(\theta^{*j}, e^{*j}\right)$ from the posterior distribution of $(\theta, e|y)$ via importance-sampling, and then averaging the $\mathbb{P}\left[\#FN = x|y,\theta^{*j}, e^{*j}\right]$ across all the draws. This requires a prior for (θ, e), which we took to be independent uniform on $\log\theta$ and logit$e$, plus of course the likelihood from section 16.5. A fully-conditioned confidence interval on $\#FN|y$ can then be found simply by inverting the cumulative distribution of $\#FN|y$.

[23] I.E. the probability that a given F1 pair is actually FN or near-FP is unaffected by the FN-status of the other F1 pairs, given θ and e.

[24] K Butler, M Stephens (1993): The distribution of a sum of Binomial random variables. Tech Rep 467, Department of Statistics, Stanford University

16.7 Results of FN analysis

We ran the above algorithms first on just the F0 & F1 columns of Table 11, and then on the F0, F1, and F2 columns. In the first version, the Maximum Likelihood Estimate on #FNs was 1.95 and the 95% UCI was 2.46; in the second version, the numbers were 3.19 and 4.0. The difference is entirely driven by the (C25, F2) entry, discussed further below; without it, the two versions are almost identical. Both versions indicated a very low *expected* number of FNs in the F2 column or beyond (less than 10% of the number expected in F1), although the second version clearly identified an *observed* likely-FN at (C25, F2).

As noted above, these FN estimates are *prior to* rescoring the F0, F1, and F2 (from C16 down) columns. Rescoring certainly fixed one FN, at (C23, F1), so the appropriate estimates and limits for the number of FNs in our final dataset (after re-scoring) are no more than (MLE 0.95, UCI 1.46) or (MLE 2.19, UCI 3.0).

The nature of the mismatching loci for any pair provides additional information on whether an F1 or F2 pair is really a FN, as opposed to just being a lucky near-FP from an unrelated pair. This is because one type of mismatch arises from a comparatively common scoring error (overlooking one allele, so a fish is recorded as AA when it should be AB), whereas the other type (incorrect size for an allele) is extremely unlikely; this was apparent in the results from our routine QC rescoring exercises of individual fish. In particular, after carefully rescoring the (C25, F2) pair, the only way it could be a FN POP would be to have a mutation at one locus and a scoring error at a second— a very unlikely conjunction of events. However, this pair is also a very unlikely event under the only two other possible scenarios: an exceptionally-matched unrelated pair, or a well-matched uncle-nephew-pair (which must be much, much rarer than unrelated pairs). In the end, the only way to resolve the true status of the (C25, F2) pair will be to use more loci, which we plan to do as part of a different project. We cannot at present decide whether to treat (C25, F2) as a FN (in which case we should use the second version of the FN analysis, including the F2 column, to get a point estimate of about 2 FN), or not (in which case we should use the first version, with a point estimate of about 1 FN).

Thus, further detailed investigation of the rescored F1s and F2s might eventually shed some light on whether we should expect 0, 1, or 2 FNs in addition to our 45 POPs. However, whichever the answer, the analysis in this Appendix demonstrates that the proportion of FNs to true POPs must be small, and is certainly not going to affect the qualitative conclusions of this project.

17 Appendix 4: What might cause overdispersion in the POPs?

The CV of the "cartoon" abundance estimate is just the CV of the number of POPs found. We have treated this as "count data", so that its variance is equal to its mean. The question arises: under what circumstances might there be overdispersion in this count?

Overdispersion would arise when the 38,000,000 comparisons are substantially non-independent. It's easy to see why a high frequency of (half)sibs would do that: if every juve had one full-sib partner in the sample, then the results for one sibling completely predict the results for the other, and the information content would only be that of 19,000,000 independent comparisons. (Recall that each POP is counted, even if the same adult is involved in several POPs— so there's no bias, only a loss of precision.) Fortunately, (half)sibs do not seem to be common in our juvenile samples, and for clarity we therefore ignore the possibility of (half)sibs in the discussions below.

There are other phenomena that might at first be suspected of causing overdispersion, but careful thought is required. For example, the 38,000,000 SBT comparisons are based on "only" 13,000 fish, each being used in multiple comparisons. Does this somehow mean that the "effective sample size" is much smaller, i.e. that there is somehow serious non-independence amongst the 38,000,000 comparisons? No— but the reasoning is subtle. Ignoring sibs as per above, consider a comparison of two fish, juvenile J and adult A, in the "cartoon" version. With no further information except the population size N, the chance of a POP would be 2/N. Assume (as with SBT) that N is large, the sample is moderately large, and the number of POPs is small. Independence amounts to the following question: does knowing that (i) J is not in a POP with any of the *other* non-A adults, and (ii) A is not in a POP with any of the *other* non-J juveniles, help us to predict the outcome of the J-A comparison?

The information in (ii) is irrelevant (given that the other juveniles aren't halfsibs of J), because if N is large then the number of non-J offspring of any adult in the sample will almost always be zero anyway, so knowing that it really is zero for one particular adult is not informative. And as for (i): knowing that the other sampled adults aren't J's parents tells us almost nothing almost nothing about whether A will be J's parent[25]. Finally, comparisons that don't involve either J or A are obviously irrelevant. So, at least in the more than 98% of comparisons that don't involve a member of a true POP, knowing the result of all the other comparisons doesn't help us predict the outcome of this one— which is the definition of independence. [If the sampled fraction of fish was a substantial proportion of the total population size, and/or if a substantial proportion of the sampled fish turned up in POPs, and/or if there were many sibs in the samples, this argument would break down.]

Another phenomenon that might superficially seem like a source of overdis-

[25] "Almost" because this information does slightly reduce the potential pool of parents, from N to [N minus the adult sample size].

persion but actually isn't, is the non-random sampling of juveniles, e.g. shifts in sampling locations within the GAB between years. Non-random juvenile sampling has in fact been a deliberate aspect of the design all along, from the 2007 CCSBT paper onwards; for example, we don't sample any juveniles off South Africa. However, as noted in that paper, the only things that matter in order to keep the comparisons statistically independent, are that (i) there are few (half)sibs among the juvenile samples, and (ii) that the *adults* be sampled randomly (apart from selectivity and other effects that are specifically allowed for in the mini-assessment). Even then, all that "randomly" has to mean is: "a parent of one of the sampled juveniles is just as likely to be sampled X years after that juvenile's birth, as is another adult of the same sex, age, and size".

There is one other phenomenon which theoretically could be important for CK abundance estimates, not so much for overdispersion as for bias: an unholy trinity of cryptic stock structure, biassed sampling of adults, and biassed sampling of juveniles. A lengthy explanation was given in our 2007 CCSBT paper, and is copied below. The key point to add in 2012, is that we have now checked as suggested in 2007 for any *temporal* substructure on the spawning grounds (see 7.2.3), and found none; we have not checked *spatial* substructure, but as below this seems *a priori* unlikely.

[4.7 from CCSBT 2007 CK paper] Population structure

So far, it has been assumed that SBT form a single population with complete interbreeding. Although no previous study has found evidence of population structure, conventional population genetics applied to large populations is a notoriously blunt tool for that task. It turns out (see **[6.0.6]**) that the basic method is unbiased even when there is population sub-structure, providing that sampling is proportional to abundance across either the sub-populations of adults, or the sub-populations of juveniles. In our SBT project, juvenile samples come only from the GAB, so if there are substantial numbers of non-GAB juveniles out there somewhere, then juvenile sampling will obviously not be proportional. However, adult samples should cover the spawning season and spawning area, although not necessarily in strict proportion to adult SBT density. Hence, the basic estimator would exhibit population-structure bias if and only if three conditions all apply:

1. adults exhibit fidelity across years to particular parts of the spawning season and/or spawning grounds;

2. the timing or location of spawning affects a juvenile's chances of going to the GAB (rather than going elsewhere or dying young);

3. sampling coverage of the spawning grounds (in time and space) is substantially uneven, and correlated with the fidelity patterns in (1). (In other words, if adults showed timing-fidelity but not spatial-fidelity, whereas coverage was even across the spawning season but not across the spawning grounds, then the uneven spatial coverage would not matter.)

There is no direct information on condition 1. With respect to condition 2, much the greatest part of SBT spawning occurs within the North Australian Basin ([**?**]), and particularly towards the east and south of the basin beyond the Australian shelf, where the Indonesian through-flows in summer would tend to push the larvae together into the Leeuwin current. These conditions seem unlikely to induce a strong location-of-spawning effect on most juvenile's subsequent propensity to go to the GAB[26], although a timing-of-spawning effect is possible. With respect to condition 3, the Benoa-based operations that we are sampling coincide well with this main spawning area ([**?**], Figure 4.3.1; note that the fishing range has expanded southwards since then, as per [**?**]). Approximate timing-of-effort information could be probably be obtained from the sampling program; spatial information has proved harder to get, but the data obviously do exist somewhere at the company level, and some insights may be obtainable through, for example, the observer program ([**?**]) or the Fishery High School program ([**?**]).

Fortunately, there is enough information in the project data to check the first two conditions. If the seasonal/spatial distribution of identified parents of GAB juveniles is substantially different to the seasonal/spatial distribution of all adult samples, then that is a clear signal that the first two conditions do apply. Such evidence of population structure[27] would be of major qualitative importance to management, regardless of its impact on quantitative results.

If and only if the first two conditions do apply, then the third could be checked using timing (and perhaps location) information on Indonesian samples. And if all three conditions do apply, then it should be possible to adjust for the uneven adult sampling probabilities, again using sampling coverage information. That is very much a bridge to be crossed only if we come to it; but because the sampling coverage is at least fairly complete[28] even if not necessarily balanced, we would in principle be able to develop a correction if required.

[6.0.6 from CCSBT 2007 CK paper] Population substructure and sampling bias Suppose the entire adult population of N is made up of two sub-populations with proportions π and $1-\pi$, and that adults are sampled proportionally from their respective sub-population, so that the overall adult sample contains $m_A\pi$ fish from the first sub-population and $m_A(1-\pi)$ from the second. Juveniles, though, are not necessarily sampled in proportion to sub-population abundance; let m_{J1} and m_{J2} be the numbers sampled from each sub-population.

If the entire dataset is analysed without regard to sub-populations, then the expected number of POPs can be calculated by considering samples from each

[26] A small proportion of larvae are found to the north of the NAB and west of it. Different oceanographic conditions apply there, and those larvae could well end up somewhere different as juveniles. However, at least until 1981, this proportion was small.

[27] "Population structure" is probably the wrong phrase, because the behaviour does not have to be heritable; adult spawning preference need not be related to earlier juvenile GABness, even if offspring's GABness is driven by adult spawning preference.

[28] Again: over the great majority of the spawning area.

sub-population separately (since there will be no cross-POPs between juveniles from one sub-population and adults from the other):

$$\begin{aligned}\mathbb{E}[H] &= \frac{2m_{J1}(\pi m_A)}{\pi N} + \frac{2m_{J2}(1-\pi)m_A}{(1-\pi)N} \\ &= \frac{2m_{J1}m_A}{N} + \frac{2m_{J2}m_A}{N} \\ &= \frac{2m_J m_A}{N}\end{aligned}$$

just as in the case without sub-populations. In other words, the basic estimate is unbiased provided at least one life-stage is sampled in proportion to sub-population abundance. If both are sampled disproportionately, though, there will be bias.

18 Appendix 5: Specification of SBT Abundance Estimation Model

18.1 Population dynamics model

All population dynamics are handled separately by sex, so a "sex" subscript should be read as implicit throughout this Appendix except where explicitly mentioned. It is omitted for brevity.

Numbers-at-age in the adult population (from age $A_{\min}$up; see below) evolve from year to year according to the usual model:

$$N_{a+1,y+1} = N_{ay} s_{ay}$$

In principle, the survival rate s_{ay} could depend on age and year, but in practice a constant s is assumed and estimated[29]. It is structurally impossible to estimate separate survival rates for males and females, so this is one case where the sexes are not treated separately.

For the plus-group at age $A_+ = 25$, the equation is

$$N_{A,y+1} = N_{Ay} s_{Ay} + N_{A-1,y} s_{A-1,y}$$

The incoming recruitments $N_{A_{\min},y}$ are either constrained to follow some pre-specified functional form (such as an exponential trend over time) with parameters to be estimated, or to be independent random effects whose mean and variance are to be estimated. All the results in this report are for the latter formulation, which is quite flexible; it allows trends in recruitment, but only if the data are sufficiently supportive. The sex ratio within the recruits is an estimable parameter; it was assumed constant over time.

The adult age structure in the first year, $N_{A_{\min}:A_+,2002}$, is also assumed to consist of random effects around an exponential trend in age. In equilibrium, the slope of that initial age distribution would equal the survival rate, but assuming equilibrium in 2002 would be unreasonable for SBT, so the initial slope was left as a free parameter. The number of animal in the plus-group to begin with, $N_{A_+,2002}$, was also treated as a free parameter, with the mean age in the plus group (see 18.1.3 below) set equal to the mean in the age samples across the first three years[30].

A small proportion of aged adults are below $A_{\min} = 8$ (the youngest age observed for any successfully-spawning adult of either sex), and careful attention is required to truncate the length- and age-frequency data without causing bias; see 18.1.3.

[29] A variant with different survival for the plus-group was also tried, but there was no suggestion in the fit that the plus-group survival was any lower than for younger fish.

[30] Results were insensitive to the value used for the initial mean age inside the plus gropu, so it was not worth adding an extra estimable parameter since it could be conveniently estimated beforehand from the data.

18.1.1 Growth

Each animal is assumed to follow its individual von Bertalanffy growth trajectory towards its own L_∞, but to have the same k and t_0 as all the others of the same sex. Growth is assumed to be constant over time. The parameters needed to describe growth thus include a variance in individual L_∞, as well as the mean and the values of k and t_0 (all sex-specific). The distribution of length-at-age (ie of individual L_∞) needs to encompass some very large fish (eg 210cm) which would constitute extreme outliers if a Normal distribution was assumed; therefore, a heavier-tailed t_{12} distribution is used instead.

Because individual growth is assumed deterministic, it is straightforward for an adult fish of known age and length to infer its personal L_∞ and then back-project to infer its length in any previous year (e.g. to the birth-year of a juvenile that might be i). Not all adult fish are aged, so for an adult of known *length* but not age, it is necessary to average (in the correct Bayesian sense) over its possible ages and then back-project for each possible age.

18.1.2 Selectivity and residence time

Selectivity, i.e. the relative chance of being caught while on the spawning grounds, is assumed to be directly proportional to residence time spent on the grounds, which is assumed to depend on length (not age) and of course on sex. The mini-assessment assumes a logistic (S-shaped) relationship, parametrized by the body-length at which the residence time is 50% of the asymptote, and the slope of the relationship at that length. Separate relationships are fitted by sex.

18.1.3 Fecundity and annual reproductive output

Annual reproductive output is broken down into two parts: residence time, multiplied by daily output. Davis et al., 2003 conclude from histology data that (female) SBT on the spawning grounds alternate successive bouts of consecutive daily spawning, with bouts of consecutive daily resting. However, the average number of such bouts during any season cannot be estimated from the histology data; it requires estimates of overall residence time.

For females, fairly precise estimates of the latter can be obtained from the data in Davis *et al.*, 2003 and Farley et al., In Prep. The daily output is assumed equal to the proportion of days spawning, times the reduction in gonad weight associated with each spawning event; the proportion of days spawning is equal to the average duration of a spawning bout divided by the average duration of (spawning bout + resting bout). The parameters required to estimate all the parameters (as functions of length) can be estimated from three simple GLMs.

Although the estimates of daily output are treated as exact constants in the mini-assessment, they are of course just estimates, and are subject to some uncertainty. Their variance— which is not large— is known from the Farley et al In Prep GLMs. To allow for this source of uncertainty in the mini-assessment (i.e. to propagate it through to the final CV) without having to embed the

original GLMs, it suffices to augment the mini-assessment with a small number of "artificial random effects" of known variance.

There is no data on male daily output, and it is not obvious that it could be deduced even from histology data similar to what we have for females, given the breeding behaviour of similar large tuna species (no direct observations for SBT are known): . The mini-assessment assumes that male daily output is a function of length controlled an estimable parameter, but it turns out that this parameter is very imprecisely estimated and that overall results are largely insensitive to the value used. Since the point estimate (a very mild *negative* effect) is implausible, it has been assumed instead that there is no length-effect for males of daily output; this is consistent with the (weak) information provided indirectly on this parameter from the rest of the data.

18.2 Overall structure of log-likelihood

The data consist of length-frequency samples (by sex and year), age subsamples (by length, sex, and year), and the genotypes of all juvenile and adult fish. The overall log-likelihood can be decomposed as

$$\Lambda = \Lambda^{\ell s} + \Lambda^{a|\ell s} + \Lambda^{g}$$

where a,ℓ, s pertain to the age, length, and sex data, and g pertains to all the genotype data. Each term depends on the associated data and on the unknown parameters; estimation consists of The terms $\Lambda^{\ell s}$ and $\Lambda^{a|\ell s}$ are standard in stock assessments, constructed assuming Multinomial distributions based on the growth and abundance parameters. Because a preliminary analysis has been used to estimate overdispersion in the length and age data (see 17), and those data have then been downscaled to "equivalent independent sample size" (no downscaling required for age), the mini-assessment does not include any additional overdispersion parameters for length or age.

The term Λ^{g} is obviously not usual in stock assessment. The individual genotype data are summarized into a set of pairwise comparisons, (basically, each juvenile compared to each adult), each with a 0/1 outcome according as the juvenile and adult are, or are not, a POP; the outcome is assumed to be ascertained without error. For some pairs, there may not be enough mutually-scored loci to make pairwise comparisons reliable, in which case no comparison is made and the datum is treated as missing. Also, no comparison is made if the age of the juvenile and year-of-capture would imply that the adult was caught in the same year as the juvenile (because those comparisons are uninformative) or before the juvenile (impossible). Comparisons are assumed independent. We can conceptually write Λ^{g} as a sum-log of binomial probabilities

$$\Lambda^{g} = \sum_{j} \sum_{i \in \mathcal{C}_j} \log \left(\mathbb{P}\left[j \sim i | \text{data}_j, \text{data}_i \right]^{\mathbb{I}[j \sim i]} \left(1 - \mathbb{P}\left[j \sim i | \text{data}_j, \text{data}_i \right] \right)^{1 - \mathbb{I}[j \sim i]} \right) \tag{2}$$

where j is a juvenile and $\mathcal{C}_j$ is the set of adults that have an "included" comparison with j, i is one adult in $\mathcal{C}_j$, $\mathbb{I}[e]$ is the indicator function, i.e. 1 or 0 according as event e actually happened or not, and data_x means the data associated with fish x (i.e. for adults, year of capture, age if known, length, and sex; for juveniles, age, which is 3 for almost all cases).

The crucial point for computing the probabilities eqn (2), is that the probability of "this" comparison being a POP is equal to the fraction of total spawning by *all* adults of that sex in in year b that would have been contributed by "this" particular adult i. In the "cartoon version" of the CK approach, the fraction is just $1/N_{\text{adult, same sex as }i}$, but this needs substantial modification for real application to SBT.

For eqn (3), the size of adult i in the year when juvenile j was born would depend on r's age at capture, because a fish of length ℓ now could either be a fast-growing youngster or a slow-growing oldster, and the former would have been smaller at any given previous year than the latter. Only about 1/3 of the adult samples are aged (though all the adults involved in POPs have been aged, as a separate exercise to the normal age-sampling procedures). The data available for computing each probability in eqn (2) therefore sometimes include adult age and sometimes not. However, the genotyped sample sizes are so large that the age and length distribution within them is very close to that of the adult catch as a whole (e.g. the mean lengths are never more than 1cm different in any year or sex). The general age-at-length information is already picked up in the term $\Lambda^{a|\ell}$, so for the very large set of comparisons that involve an aged adult that is not part of a POP, the age is really not telling us more than the length; we might just as well use the probability of age given length for each non-POP adult, since across the very large set of such adults the age-given-length distribution will be very close to the actual age distribution. Therefore, we break Λ^g into two terms, where the first deals with all the comparisons without explicitly considering adult age, and the second looks at adult age amongst the identified POPs:

$$\Lambda^g = \Lambda^{g0} + \Lambda^{a|\text{POP}}$$

For Λ^{g0}, given the juvenile birth-year b, adult capture-year y, adult sex s, and adult length ℓ, we need to compute

$$\mathbb{P}[j \sim i | bys\ell c] \tag{3}$$

where c denotes the fact of capture. In eqn (2), c is formally irrelevant, but it does enter into the computational breakdown.

To get the numerator of eqn (3), i.e. the total reproductive output of r at $y-b$ years before it was caught, we need to sum over its possible ages at capture:

$$\mathbb{P}[j \sim i | bys\ell\,\{c\}] = \sum_a \mathbb{P}[j \sim i | abys\ell\,\{c\}] \times \mathbb{P}[a|\,\{b\}\,ys\ell c]$$

where the curly brace signifies that a conditioning variable happens to be irrelevant to that particular term, even though formally required by the laws of probability, and therefore can be omitted in subsequent lines. Because fecundity is assumed to depend on size not age, the term $\mathbb{P}[j \sim i|abys\ell]$ is equal to $\mathbb{P}[j \sim i|bs\ell'(\ell, a, a+b-y)]$ where $\ell'()$ is the back-projected length given length, age at capture, and when we are back-projecting to. Next, we work out the annual spawning output from a fish of that size (and sex). To get the **denominator** of eqn 3), i.e. the total spawning output across all fish that year, we work directly from the age structure in that year: sum over ages of the number of fish of that age, times the sum over lengths of the proportion at that length given age times the relative fecundity of that length.

There are about 38,000,000 pairwise comparisons, but computation can be speeded up considerably by grouping into all the possible combinations of adult length, sex, age, year of capture, and juvenile birth year (about 50,000 combinations).

18.3 Formal derivation of probabilities

This section gives a formal derivation of how to compute the probabilities described above. A key assumption throughout, is that a true parent of given length is no more or less likely to be caught this year, or to have died between now and the juvenile's birth-year, than any other adult of the same length and sex. Further, we assume that the "sampling rate" of adults[31] is small or fairly constant over time and length, so that we can *ignore* the small amount of information contained in the fact that an adult sampled now was *not* sampled in previous years. In fact, the number of adults in our annual samples is a very small fraction of the estimated total number of adults (under 1%), so this assumption is certainly reasonable.

First, some notation:

- j is juvenile, i is adult. f is any fish.
- t_f is the year-of-capture-and-inclusion-in-dataset of fish f. For fish not in the genotyped samples (whether still alive, or caught elsewhere, or caught in Indonesia but not genotyped), t_f is defined arbitrarily to be one year beyond the end of the study.
- y_{0j} is year of juvenile birth, y_i is year of adult capture, a_i and ℓ_i are age and length at capture. All calculations are separate by sex, but the sex-subscript is omitted for brevity. For the adult versions, year-subscripts may be added, and/or the per-fish subscript may be omitted if y, a or ℓ is being summed over.
- Assume deterministic growth for each individual, so that the length Δy years ago can be written as $\ell'(\ell, a, \Delta y)$ for a known function $\ell'(.)$.

[31] I.E. the chance of being in our sample in any year, as opposed to the mortality rate

- Sets are either in curly font, or have curly braces, except that braces are omitted around sets with one element where the context is clear.
- $\mathcal{M}_{jy}$ is the set of potential Mothers that are compared with j in year y, and $\mathcal{M}_j \stackrel{\triangle}{=} \cup_y \mathcal{M}_{jy}$.
- "$\sim$" denotes POPness of two individuals, and the setwise extension "$\mathcal{A} \sim \mathcal{B}$" means "elements in $\mathcal{A}$ that are POPs with one or more elements in $\mathcal{B}$", so that $\{\mathcal{M}_j \sim j\}$ is either the empty set, or one particular member of $\mathcal{M}_j$.
- A MOP is a Mother-Offspring Pair. It is easiest to deal separately with Mothers and Fathers; the formulae are the same but of course the numbers and the parameters will be different.

The probability of *not* finding a MOP for juvenile j is

$$\mathbb{P}\left[\{\mathcal{M}_j \sim j\} = \emptyset\right] \approx \prod_y \prod_{i \in \mathcal{M}_{jy}} \mathbb{P}\left[j \not\sim i | \ell_{iy}\left\{, y_{0j}, y\right\}\right] \tag{4}$$

and from now on all probabilities are implicitly conditional on juvenile birth-date, adult capture-date, and of course adult sex. The reason for using "$\approx$" rather than "=" in eqn (4) is that, when we test and reject a female adult, the pool of potential parents actually shrinks by one. The collection of tests is thus really "hypergeometric", whereas we are treating it as "multinomial" by assuming each test is independent. The justification is, again, that the number of potential parents is huge compared to the number tested.

It is most convenient to first compute eqn (4) for all juveniles, whether part of a MOP or not. Then, for the few cases that are in a MOP, the formula is modified like so:

$$\mathbb{P}\left[\{\mathcal{M}_i \sim i\} = \{j\}\right] \approx \mathbb{P}\left[\{\mathcal{M}_i \sim i\} = \emptyset\right] \frac{\mathbb{P}\left[j \sim i | \ell_{iy}\right]}{\mathbb{P}\left[j \not\sim i | \ell_{iy}\right]}$$

We will return to the expression $\mathbb{P}\left[j \sim i | \ell_{jy}\right]$shortly, for computing Λ^{g0}. For the other term $\Lambda^{a|\mathrm{POP}}$, the distribution of age amongst females in MOPs, we require

$$\begin{aligned}
\mathbb{P}\left[a_{iy} | c\ell_{jy}, j \sim i\right] &= \frac{\mathbb{P}\left[j \sim i | \{c\}\, \ell_{iy}, a_{iy}\right] \mathbb{P}\left[a_{iy} | c\ell_{iy}\right]}{\mathbb{P}\left[j \sim i | c\ell_{iy}\right]} \qquad (5) \\
&= \mathbb{P}\left[j \sim i | a_{iy}, \ell_{iy}, \ell'_{iy_{0j}}\left(\ell_{iy}, a_{iy}, y_{0j} - y_{ci}\right)\right] \frac{\mathbb{P}\left[a_{iy} | c\ell_{iy}\right]}{\mathbb{P}\left[j \sim i | c\ell_{iy}\right]} \\
&= \mathbb{P}\left[j \sim i | \ell'_{ji}, a'_{iy}\right] \times \frac{\mathbb{P}\left[a_{iy} | c\ell_{iy}\right]}{\mathbb{P}\left[j \sim i | c\ell_{iy}\right]} \\
&= \frac{\mathrm{fec}\left(\ell'_{ji}, a'_{iy}\right)}{\mathrm{totfec}_{y_{0j}} \mathbb{P}\left[\ell_{iy}\right]} \times \frac{\mathbb{P}\left[a_{iy} | c\ell_{iy}\right]}{\mathbb{P}\left[j \sim i | c\ell_{iy}\right]}
\end{aligned}$$

where $\ell'_{ji} \triangleq \ell'(\ell_{iy}, a_{iy}, y_{0j} - y)$ and $a'_{iy} = a_{iy} - (y_{0j} - y_{ci})$. To evaluate, it's easier to first compute

$$\mathbb{P}[a_{iy}|c\ell_{jy}, j \sim i] \propto \text{fec}\left(\ell'_{ji}, a'_{iy}\right) \times \mathbb{P}[a_{iy}|c\ell_{iy}]$$

and then just normalize to sum to one.

The reason a'_{iy} appears in fec() of eqn (5), is that the model enforces a minimum age $A_{\min}$ for successful spawning, regardless of size or presence-on-ground (see *Truncation...* below). The total fecundity for any year y^* and age a^* (and sex) is

$$\text{totfec}_{y^*} = \sum_{a^*} n_{a^*y^*} \sum_{\ell^*} \mathbb{P}[\ell^*|a^*] \,\text{fec}(\ell^*)$$
$$\mathbb{P}[A_{i^*y^*} = a] = \frac{n_{ay^*}}{\sum_{a^*} n_{a^*y^*}}$$

which is where the total population size comes in.

For Λ^{g0}, we also need

$$\mathbb{P}[j \sim i|c\ell_{iy}] = \sum_a \mathbb{P}[A_{iy} = a|c\ell_{iy}]\,\mathbb{P}[j \sim i|\{c\}\,\ell_{iy}, a] \tag{6}$$
$$= \sum_a \mathbb{P}[A_{iy} = a|c\ell_{iy}] \frac{\text{fec}\left(\ell'_{ji}, a'_{iy}\right)}{\text{totfec}_{y_{0j}}\mathbb{P}[\ell_{iy}]}$$

Conditioning explicitly on capture (and implicitly on sex throughout), and using "+/-" to denote "of adult/nonadult age" with $p_{+\ell y} \triangleq \mathbb{P}[+|\ell y]$ (see *Truncating...* below), we have for any genotypee:

$$\begin{aligned}\mathbb{P}[a_{iy}|c\ell_{iy}] &= \mathbb{P}[a_{iy}|c\ell_{iy}+] \times \mathbb{P}[+|c\ell_{iy}] + \mathbb{P}[a_{iy}|c\ell_{iy}-] \times \mathbb{P}[-|c\ell_{iy}] \\ &= \mathbb{P}[a_{iy}|c\ell_{iy}+] \times p_{+\ell y} + 0 \times (1 - p_{+\ell y}) \\ &= \\ &= p_{+\ell y} \frac{\mathbb{P}[c\ell_{iy}|a_{iy}\,\{+\}]\,\mathbb{P}[a_{iy}|+]}{\mathbb{P}[c\ell_{iy}|+]} \\ &= p_{+\ell y}\left(\frac{\mathbb{P}[c|\ell_{iy}\,\{a_{iy}\}]\,\mathbb{P}[\ell_{iy}|a_{iy}]\,\mathbb{P}[a_{iy}|+]}{\sum_a \text{numerator}}\right)\end{aligned}$$

18.3.1 Plus-group and back-projection

Suppose two female fish are now 190cm, but one is 25yo and the other is 40yo. The first is a faster grower than the second, so five years ago the first fish would have been *smaller* than the second was at the same time. Hence, the "correct" back-projection for an animal in the plus-group depends on the age distribution inside the plus-group, which by definition we are not attempting to track precisely. To deal with this, the model keeps track of mean age within the plus-group, as well as the total number of animals inside the plus-group. The change

in within-plus-group-mean-age from one year to the next is straightforward to calculate, given the numbers already in the plus-group, the incoming numbers, and the survival rate. For back-projection purposes, the model assumes that the age of all animals in the plus-group is equal to the mean. This is an approximation, but since growth is presumably slow within the plus-group (age 25+ here), the error in the approximation is unlikely to be large.

18.3.2 Estimation of random-effects variance

Incoming recruitments at age $A_{\min}$, plus the "recruitments" corresponding to the initial age structure in 2002, are modelled as random effects with an unknown variance ω. To estimate ω in a statistically consistent and approximately unbiased fashion, it is necessary to "integrate out" both the random effects and the numerous fixed effects (for which vague priors are assumed), so that the overall log-likelihood is reduced to a modified profile log-likelihood depending on just a single parameter:

$$\Lambda^*(\omega) = \log \int_\theta \exp\left(\Lambda(\theta;\omega)\right) d\theta \tag{7}$$

where θ contains all the fixed and random effects (but of course does not include ω). Eqn (7) is impossible to compute exactly, but Laplace approximation is an effective alternative; see e.g.

Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models: HJ Skaug, DA Fournier: Computational Statistics & Data Analysis 51 (2006) 699 – 709

Laplace approximation requires accurate evaluation of the derivatives of Λ with respect to θ, which is accomplished here using the Automatic Differentiation software TAPENADE (Hascoet et al., 2012).

18.3.3 Truncating the age & length distributions

For stable fitting, it has proved necessary not to include very small "adults", e.g. under 150cm. The problem is that there are a huge number of fish in the population at those sizes, but only a very small number in the adult catch; hence the residence time must be extremely low at lengths below 150cm, and the residence-time model can only accommodate that by "going wrong" for large lengths. The tail is trying to wag the dog, and we are better off ignoring this not-informative-but-nevertheless-influential part of the data. Likewise, it is desirable to avoid the statistical problem of having to estimate numbers-at-age at ages so young that most of the cohort won't have appeared in the adults yet, so we want to truncate by age as well as length. All identified parents were age 8+ and (inferred) length 150cm+ in the year-of-juvenile-birth, so those lower limits seem reasonable. However, it is still necessary to be careful about the "book-keeping", in particular the proportion of say 8yo fish that are excluded from analysis because they are too small, and the proportion of say 7yo fish

that are big enough to be included in the length-frequency samples but are (by assumption, but consistent with the data) non-contributors to overall population fecundity. (In the model, an 8yo fish of 150cm makes some small contribution to total population reproductive output, but a 7yo of the same size does not. This minimum-age criterion is the only way in which the model allows age, rather than size, to affect reproduction.)

Specifically, the model says this:

- Undersize fish, i.e. below $\ell_{\min} = 150$cm, are allowed to visit the spawning grounds and even to spawn successfully, in principle. Their contribution to total fecundity in any year is accounted for. However, they are *not* used in fitting length-or age-at-length data.

- Length-at-age is assumed to follow a t_{12} distribution, so the "mean undersize", i.e. the mean length of a fish of age *a given* that its size is below $\ell_{\min}$, is the mean of a truncated t_{12}distriubution, which can be easily computed. The mean-undersize is used to set the contribution of undersize fish to total fecundity, by assuming all too-small fish of that age are exactly at that mean-undersize, and applying the fecundity relationship there. In principle, this is mildly because the fecundity-length relationship is actually nonlinear, but the term is small enough that this should not matter.

- Fish younger than $A_{\min} = 8$ are **not allowed** to spawn successfully, regardless of how big they are. However, they are allowed to visit the spawning grounds, and will contribute to the observed length-freqs close to $\ell_{\min}$. Their numbers are not tracked in the population dynamics, and the inflation of $\mathbb{P}[\ell]$ from too-young fish is handled by "profiling out", as described below.

- Fish of age $A_{\max} = 25$ or above are put into a plus-group

- Fish of length $L_{\max} = 200$cm up are put into a "plus-group for length".

The main book-keeping difficulty is that we know that eg the 150cm length class will include some fish that were less than age 8, but the population dynamics sub model does not include terms for those ages, so we cannot predict how many there should be. Fortunately, since there is plenty of age-at-length data, we can avoid inconsistency as follows.

For any given sex (implicitly conditioning on this throughout) and year y, we want to estimate $\mathbb{P}[\ell|cy]$ (c for capture) using (i) quantities that are functions of parameters in the model (eg $\mathbb{P}[a|a \geqslant A_{\min}, y]$; note that we have to condition on $a \geqslant A_{\min}$, since the model doesn't keep track of younger fish) , and (ii) spare bits of data, as will become clear(er). It's crucial for this derivation that selectivity (probability of capture) does *not* depend on age, but only on length. Letting "+" stand for "$a \geqslant A_{\min}$ and "-" for the converse, and being flexible with

how they're used, and using the curly-brace convention above, we have

$$\begin{aligned}
\mathbb{P}\left[\ell|cy\right] &= \mathbb{P}\left[-\ell|cy\right] + \mathbb{P}\left[+\ell|cy\right] \\
&= \mathbb{P}\left[-|\ell cy\right]\mathbb{P}\left[\ell|cy\right] + \mathbb{P}\left[+|cy\right]\mathbb{P}\left[\ell|cy+\right] \\
&= \mathbb{P}\left[-|\ell cy\right]\mathbb{P}\left[\ell|cy\right] + \mathbb{P}\left[+|cy\right]\frac{\mathbb{P}\left[c|\ell\left\{y+\right\}\right]\mathbb{P}\left[\ell|y+\right]}{\mathbb{P}\left[c|y+\right]} \\
&= \mathbb{P}\left[-|\ell cy\right]\mathbb{P}\left[\ell|cy\right] + \frac{\mathbb{P}\left[+|cy\right]}{\mathbb{P}\left[c|y+\right]}\mathbb{P}\left[c|\ell\right]\sum_{a\in+}\mathbb{P}\left[\ell|a\left\{y+\right\}\right]\mathbb{P}\left[a|y+\right] \\
\implies \mathbb{P}\left[\ell|cy\right]\left(1-\mathbb{P}\left[-|\ell cy\right]\right) &= \kappa\mathbb{P}\left[c|\ell\right]\sum_{a\in+}\mathbb{P}\left[\ell|a\right]\mathbb{P}\left[a|y+\right] \\
\implies \mathbb{P}\left[\ell|cy\right] &= \kappa\frac{\mathbb{P}\left[c|\ell\right]}{1-\mathbb{P}\left[-|\ell cy\right]}\sum_{a\in+}\mathbb{P}\left[\ell|a\right]\mathbb{P}\left[a|y+\right]
\end{aligned}$$

where κ is defined below. Now:

- we can estimate $\mathbb{P}\left[-|\ell cy\right]$ directly from the proportion of age-at-length this year that are below $A_{\min}$. Technically, these terms are nuisance parameters, but are being "profiled out"— if they were included as estimable parameters, then their MLEs would end up being just as described. The values are small enough that the uncertainty associated with them is unimportant.
- $\mathbb{P}\left[c|\ell\right]$ is proportional to residence time;
- $\mathbb{P}\left[\ell|a\right]$ comes from the growth-curve;
- $\mathbb{P}\left[a|y+\right]$ comes from the population dynamics model;
- $\kappa \stackrel{\triangle}{=} \mathbb{P}\left[+|cy\right]/\mathbb{P}\left[c|y+\right]$ is a normalizing constant.

Since we are only interested in the cases $\ell \geqslant L_{\min}$, we can compute $\mathbb{P}\left[\ell|cy\right]$ from the above by setting $\kappa = 1$, and then normalize to sum to one over $\ell \geqslant L_{\min}$.

This would be complete, except that for some year/sex/length combinations, there are no age data at all, or all of the ages fall below $A_{\min}$. Such cases lead to impossible values for the correction term $\left(1-\mathbb{P}\left[-|\ell cy\right]\right)^{-1}$. Instead, therefore, we fit a GLM to $(\#<a_{\min}|\ell cys)\,|\,(\#\ell cys)$ with ℓys as main effects, and use the fitted probabilities as estimates of $\mathbb{P}\left[-|\ell cy\right]$. The vast majority are close to 0, a few around 0.05, and the biggest are around 0.2.

18.3.4 Tedium: what is mean undersize with t-distribution L|A?

There do not seem to be widely-known formulae for computing the mean of a truncated t-distribution (unlike for a Normal distribution). However, it turns out that there is a fairly simple formula when the degrees-of-freedom ν is an even-valued integer. We want the mean value when $t < x$, with $x < 0$ for truncation on the lower tail (as required for undersize fish). Starting with the formula for the PDF of a t_ν distribution, we have:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}}\frac{1}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$\mathbb{E}\left[t|t<x\right] = \int_{-\infty}^{x}\frac{t}{\left(1+t^2/\nu\right)^{\frac{\nu+1}{2}}}dt \Bigg/ \int_{-\infty}^{x}\frac{1}{\left(1+t^2/\nu\right)^{\frac{\nu+1}{2}}}dt$$

$$k \triangleq \frac{\nu+1}{2}$$

$$E_{xk} \triangleq \int_{-\infty}^{x} t\left(1+t^2/\nu\right)^{-k}dt$$

$$u \triangleq 1+t^2/\nu$$

$$du = (2/\nu)\,tdt$$

$$E_{xk} = \frac{\nu}{2}\int_{\infty}^{1+x^2/\nu} u^{-k}du$$

$$= -\frac{\nu}{2}\left[\frac{1}{1-k}u^{1-k}\right]_{1+x^2/\nu}^{\infty}$$

$$= -\frac{\nu}{2}\frac{1}{k-1}\left(1+x^2/\nu\right)^{1-k}$$

$$= -\frac{\nu}{\nu-1}\left(1+x^2/\nu\right)^{-\frac{\nu-1}{2}}$$

$$\Longrightarrow\ e_\nu(x) = \int_{-\infty}^{-|x|} tf(t)\,dt = -\frac{1}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\frac{\sqrt{\nu}}{\nu-1}\left(\frac{\nu}{\nu+x^2}\right)^{\frac{\nu-1}{2}}$$

Also we have

$$\int_{-\infty}^{x} f(t)\,dt = \frac{1}{2}I\left(x,\tfrac{\nu}{2},\tfrac{1}{2}\right)$$

where $I()$ is the normalized incomplete Beta function— note again that this assumes $x<0$.

WHEN ν is even, there is closed-form expression for $I()$. Bancroft 1949 eqn 15 gives

$$I_x(p+n,q) = \frac{1}{(p+n-1)^{(n)}}\sum_{r=0}^{n}(-1)^r\binom{n}{r}(p+q+n-1)^{(n-r)}(q+r-1)^{(r)}I_x(p,q+r)$$

so if we take $p=1$, $n=(\nu/2)-1$, $q=1/2$, we can reduce to a sum of terms of $I_x\left(1,\frac{1}{2}+r\right)$ multiplied by pre-computed coefficients. For computational purposes (and for fairly small ν), it is much more useful to apply this recursive

formulation:

$$
\begin{aligned}
n &:= \frac{\nu}{2} - 1 \\
p &= 1 \\
q &= \tfrac{1}{2} \\
c_0 &= \prod_{r=1}^{n} \frac{p+q+n-r}{p+n-r} \\
c_r &= -c_{r-1} \times \frac{q+r-1}{p+q-1+r} \times \frac{n+1-r}{r} \\
t_0 &= \sqrt{1-x} \\
t_r &= (1-x)\, t_{r-1} \\
I_x &= \sum_{r=0}^{n} c_r \left(1 - t_r\right)
\end{aligned}
$$

Also we need to compute the numerator for the truncated case. Provided $x < 0$, we have

$$
\begin{aligned}
t &:= \frac{\nu}{\nu + x^2} \\
u &:= t^n \sqrt{t} \\
d_0 &= 1/2 \\
d_r &= d_{r-1} \frac{r + \frac{1}{2}}{r} \\
e &= -d_n \frac{\sqrt{\nu}}{\nu - 1} u
\end{aligned}
$$

so that the overall answer is $2e/I$.

When $x > 0$, we know by symmetry that

$$
\begin{aligned}
0 &= \mathbb{E}\left[t\right] \\
&= \mathbb{E}\left[t | t < x\right] \mathbb{P}\left[t < x\right] + \mathbb{E}\left[t | t > x\right] \mathbb{P}\left[t > x\right] \\
\implies \mathbb{E}\left[t | t < x\right] \mathbb{P}\left[t < x\right] &= -\mathbb{E}\left[t | t > x\right] \mathbb{P}\left[t > x\right] \\
&= \mathbb{E}\left[t | t < -x\right] \mathbb{P}\left[t < -x\right] \\
\implies \mathbb{E}\left[t | t < x\right] &= \mathbb{E}\left[t | t < -x\right] \frac{\mathbb{P}\left[t < -x\right]}{\mathbb{P}\left[t < x\right]} \\
&= \frac{e_\nu\left(x\right)}{1 - \mathbb{P}\left[t < x\right]} \frac{1 - \mathbb{P}\left[t < -x\right]}{\mathbb{P}\left[t < x\right]} \\
&= \frac{e_\nu\left(x\right)}{\mathbb{P}\left[t < x\right]}
\end{aligned}
$$

19 Appendix 6: Reports to the Project Steering Committee

Steering Committee reports from 2010 to 2012 are reproduced below. These reports cover the substantive information and issues considered by the committee during the life of the project. The earlier reports (2008 and 2009) were largely project management focussed and have not been included for brevity and relevance to the final results of the project.

19.1 SBT close-kin abundance, Mark Bravington, Pete Grewe, Update May 2010

19.1.1 Checking for sibs/halfsibs in the juvenile sample

The idea which underpins close-kin estimates of adult abundance is:

$$\mathbb{E}[H] = \frac{2m_J m_A}{N_A} \tag{8}$$

where $\mathbb{E}[H]$ is the expected number of "hits" (i.e. identified parent-offspring *matches* between juveniles and adults), m_J and m_A are juvenile and adult sample sizes, and N_A is the adult abundance, which is to be estimated. With SBT, complications such as age structure and multi-year sampling make the real equation more complicated— it will actually be defined implicitly rather than explicitly, as a maximum-likelihood estimate (MLE)— but the basic statistical properties can be inferred from the above.

Once the juvenile and adult samples have been compared and we know the actual number of hits h, equation (8) leads to the simple abundance estimate:

$$\hat{N}_A = \frac{2m_J m_A}{h} \tag{9}$$

As per last year's document, equation (9) is approximately unbiased regardless of how many sibs and halfsibs there are in the juvenile sample, because h is the number of *matches* identified, not the number of *parents*. Each juvenile has exactly two potential matches, regardless of whether the adults that correspond to those matches are also parents of other juveniles in the sample.

However, even though a high incidence of sibship would not cause bias[32], it would affect variance. In the extreme case where all the juveniles are full sibs, there are only 2 "tagged" adults in the population, so there is a very high probability of finding neither (in which case $\hat{N}_A = \infty$) and a very low probability of finding one (in which case $h = m_J$ and $\hat{N}_A = 2m_A$). The variance$\mathbb{V}[H]$ of H, and thus the variance of $\hat{N}_A$, would be enormous. Less apocalyptically, suppose we found that 50 of the juvenile SBT sample were full-sibs from the same pair of super-parents; since we are only expecting ~ 100 hits overall, the estimate $\hat{N}_A$

[32] At least not in equation (8); as usual in statistics, there is a finite-sample-size bias in equation (9) which is not a great worry unless $\mathbb{E}[H]$ is very small or the sibship is extreme.

would change greatly if one of the super-parents was found in the adult sample, so the variance would still be high.

Some investigation of sibship incidence is therefore a prerequisite to a full-scale close-kin abundance estimation exercise. If sibship is very common, the sample size (which has been calculated on the assumption of negligible sibship, so that each adult-juvenile comparison is independent) will be too small to give a usefully precise abundance estimate. Although this could be fixed in principle just by collecting more samples, the cost and time required might make the whole project infeasible. Given what we know of the spawning biology and geography of SBT and related species, and the nature of our juvenile sampling (mostly 3-year-olds), we have always considered that a high incidence of sibship was *a priori* unlikely. However, the only way to be sure is to check. Before embarking on full-scale expensive genotyping of 5000+ adults and juveniles, therefore, we have sampled a limited subset of 480 juveniles from one year and one cohort (about 1/6 of our maximum available sample size for that cohort/year) to investigate the possible impact of sibship.

There were three possible outcomes: current approach looks fine; current approach needs modification; entire project is doomed. Happily, the current approach *does* look fine. Reaching this conclusion has required extensive, and highly technical, statistical development, which is described in sections 1.1 and 1.2 below. The results are shown in section 1.3.

19.1.2 Goal of sibship study

The key question for variance calculations is this: how many *unique* parents did our juvenile samples have? That quantity, P_J, is the number of "tags" that are sought in the adult sample. If the number of tags is very small, the variance of the abundance estimate will inevitably be high. The maximum possible value of P_J is $2m_J$ and, if we find that the estimate $\hat{P}_J$ is close to $2m_J$, then sibship is not a concern. If $\hat{P}_J$ turns out to be very low— say in the 100s, rather than the 1000s that we hope for— then the project as it currently stands is not feasible.

If $\hat{P}_J$ is substantially less than $2m_J$ but not too low, then the project might require some modification. There would be two options. First, we could stick with equation (1), make a more refined calculation of $\mathbb{V}[H]$, and reconsider the sample size & CV accordingly. The variance calculation would require not just $\hat{P}_J$, but also an estimate of the *distribution* of sibship, e.g. whether there are a moderate number of moderately-super-parents, or a small number of very-super-parents. Alternatively, we could recast equation (8) in terms of the number of *parents* (tags) found T, rather than the number of *matches*:

$$\mathbb{E}[T] = m_A \times \frac{P_J}{N_A} \tag{10}$$

$$\implies \hat{N}_A = \frac{m_A \hat{P}_J}{t} \tag{11}$$

because there are P_J tags in a population of size N_A, and m_A chances to find

those tags. The variance calculation now needs to take into account the uncertainty in $\hat{P}_J$, but that should be straightforward provided we can estimate $\hat{P}_J$ in the first place; and, unlike the first option, there is no need to infer anything about the distribution of sibship, which might be difficult. It may in fact turn out that equation (10) provides a better general framework for close-kin studies than equation (8). Such details could be resolved later, provided that the existing sample sizes are adequate.

To investigate the impact of sibship, we have therefore concentrated just on $\hat{P}_J$ and its variance. Over the last decade, many different programs have addressed aspects of sibship and parentage. [**?**] provide an up-to-date review, identifying 6 types of parentage/sibship analysis. Our situation is closest to type 6, sibship reconstruction, in which no parents are available and there are no known prior groups of sibs/halfsibs. However, there does not seem to be an existing sibship-reconstruction programs that focusses on our particular question of estimating P_J; normally, the focus is more on reconstructing individual family trees and ancestors. Our problem has several non-standard aspects:

- very large sample size (potentially 3000, if we were subsequently to estimate $\hat{P}_J$ for an entire juvenile cohort sample, rather than the current subsample of 500);
- results must be "extrapolable" from a subsample up to the larger sample;
- simple aggregated goal ($\hat{P}_J$), rather than complicated individual-level results;
- need variance as well as point estimate;
- no confidence in any prior assumptions about mating structure.

With these in mind, we developed a completely new method for estimating sibship, described below.

19.1.3 A method for counting parents via sibship

Although the number of parents depends on the incidence of sibship (including half-sibship), it is not just the number of sib-pairs, or the number of individuals *in* a sib-pair, but also their pattern that matters. Suppose we have a subsample of 50 fish, and there are 0 half-sibs and 6 sib-pairs to be found amongst our 50*25 unique comparisons. If these 6 pairs arise from 12 individuals each with 1 full-sib, then the number of unique parents is $(50 - 6) \times 2 = 88$ unique parents, with 6 "redundant" sibs contributing no additional parents. On the other hand, the 6 sib-pairs could arise from just 4 individuals who are all full-sibs. This would give $(50 - 3) \times 2 = 94$ unique parents.

The second point to make is that we are ultimately choosing loci in order to identify parent-offspring pairs, not to identify sibship. Even though sibs share 50% of genes on average, sibs are harder to identify reliably than parent-offspring pairs, because there is no guarantee of sharing at any one locus. Half-sibs are

much harder to identify, since they only share 25% of genes on average. Figure 14 shows how often the true sib status of a pair of individuals will be misclassified, from simulations using the same 11 loci we had for our subsample of 500 (see next section for some explanation of what is plotted). The numbers in green show percentage of "anomalous" pairs; for example, 11% of half-sib pairs have genotypes that are more compatible with being full-sibs. This degree of genetic "noise" can be handled statistically, but cannot be brushed under the carpet. For example, even if there are no true sibs or halfsibs in the subsample, a substantial proportion of pairwise comparisons will look most like half-sibs.

Figure 14: "Naive" misclassification rates with 11 SBT loci. The red lines show where log-likelihoods are equal.

Summarizing the data through pairwise likelihood ratios Large sample sizes pose some significant computational restrictions on sibship studies. With a per-cohort sample of 3000, there are about $3000 \times 1500 = 4.5 \times 10^6$ unique pairwise comparisons (i.e. consideration of possible sib-relationships between animals i & j). While higher-order comparisons could be considered, and would in principle carry more information about P_J, they would greatly increase complexity and dataset size (e.g. to $> 10^9$ with 3-way comparisons), so we have restricted analysis to pairwise comparisons.

Each pairwise comparison can be summarized by just 3 numbers, the likelihoods of the genotypes g_i and g_j of animals i and j:

$$\ell_{ijm} \stackrel{\triangle}{=} \log \mathbb{P}\left[g_i, g_j | m_{ij}\right]$$

Here, $m_{ij} \in \{0, 1, 2\}$ is the number of parents shared by i and j. For a given value of m_{ij}, the likelihood is calculated by summing over possible parental genomes consistent with that m_{ij}, assuming independent loci, no scoring error (see later), known allele frequencies, and that alleles are distributed "at random" in the parental generation; see e.g. [**?**] for details. In fact, only the likelihood *ratios* are informative about P_J, not the absolute values; each comparison can be summarized by just two numbers, and we may as well work with $\ell_{ijm} - \ell_{ij0}$.

Having used the genotype data first to calculate the population allele frequencies and then to compute all pairwise likelihood ratios, there is no further use for the genotypes themselves. This greatly speeds up calculations, since all the likelihood ratios need be found once only, requiring just a few seconds even for thousands of fish. Figure 14 shows the theoretical distribution of log-likelihood-ratio with our 11 SBT loci, each panel being for a different true (row) and tested (column) relationship. The red line in each mini-graph marks a log-likelihood-ratio of zero.

Counting parents using perfect relatedness data Given enough loci, the likelihood ratios would eventually tie down the sib status of i and j with effec-

tive certainty; in other words, we could compute the entire matrix M without error. The first question to ask is: if we *did* know M exactly, would that be enough to calculate P_J exactly? The answer turns out to be yes, provided that adults are not hermaphrodites (which tuna aren't). A proof based on graph-colouring is given in the Appendix, along with a hermaphrodite counter-example for interest's sake.

The graph-colouring approach does not lend itself well to statistical estimation when the number of loci is limited and M is consequently uncertain. Instead, we take a quite different approach to estimating P_J, by separately estimating the Unique Parental Contribution of each fish i and then adding them up. The UPC C_i is defined as

$$C_i = \frac{1}{1+D_i} + \frac{1}{1+S_i} \tag{12}$$

where D_i is the number of other fish that share a mother (Dam) with i, and S_i is the number that share a father (Sire). The true value of C_i must range between 0 and 2, being exactly 2 if the fish is an "only-fish". To see why $\sum_i C_i = P_J$, note that a female parent with $r > 0$ offspring will contribute to r of the C_i's, and for each of those the female term in C_i will be $1/r$; thus the sum of the female terms involving that parent will always be 1, regardless of r.

The basic idea of estimating C_i is to estimate the proportion of sibs and half-sibs to fish i, and also to use a subset of comparisons *between* likely half-sibs of i to work out the proportion of i's half-sibs that came through one parent *vs* the other[33]. Eventually this leads to an MLE$\hat{C}_i$, which (after adjustment) is approximately unbiased. These estimates can simply be summed across fish to get$\hat{P}_J$. The details of estimating each C_i are fairly nasty, however, and most of them are given next.

Estimating the Unique Parental Contribution from limited loci To estimate C_i given finite loci, first rewrite equation (12) as

$$C_i = \frac{1}{1+n\,(p_{DS}+p_D)} + \frac{1}{1+n\,(p_{DS}+p_S)}$$

where there are $n+1$ fish in the subsample (so there are n other fish besides #1), p_{DS} is the proportion of the n that are full-sibs of i, and p_D and p_S are the proportion that are maternal & paternal half-sibs of i respectively; these proportions should really have a subscript i, but it is omitted for brevity. Given the likelihood ratios ℓ_{ijm}, the estimation of p_{DS} and the combined proportion of half-sibs, $p_H \triangleq p_D + p_S$, is a standard exercise in ML estimation, from this log-likelihood:

[33] Of course, without mtDNA it is impossible to tell which sex of parent gave rise to which half-sibs, but it *is* possible to estimate how many come through one parent rather than the other. Since C_i is symmetric in males and females, this is all that's needed.

$$\log \mathbb{P}\left[\ell_{i,j\neq i}\right] = \sum_{j\neq i} \log\left(p_{DS}\exp \ell_{ij2} + p_H \exp \ell_{ij1} + (1 - p_{DS} - p_H)\right)$$

Now the question is how to partition p_H into its maternal and paternal components, *without* the help of mtDNA. It helps to first consider the perfect-information case, where M is known exactly. Start by picking any half-sib j of i, and assume it is related to i through the Dam not the Sire; set $n_D = 1$ and $n_S = 0$. Then compare it in turn to all other half-sibs $k \neq j$ of i. If M_{jk} =1 or 2, then set $n_D := n_D + 1$ because j and k must share their Dam with i[34]; if $M_{jk} = 0$, then set $n_S := n_S + 1$ because k must share its Sire with i. At the end of this process, set $p_S = n_S/n$ and $p_D = n_D/n$. Of course, i and j might actually be related through the Sire not the Dam; in that case p_S and p_D would be reversed, but C_i would be unchanged.

The basic idea in the above scheme is that the proportion of Dam-shares and Sire-shares can be worked out by checking how many half-sibs of i are half-sibs or full-sibs of each other; if all of them are, for example, then either $p_S = p_H$ and $p_D = 0$, or *vice versa*. The same principle applies with a finite number of loci and consequent uncertainty about M, but it is necessary to account for the possibility that the "reference half-sib" is not actually a half-sib of i. Some care is needed to avoid "using the data twice". The steps used are as follows:

1. Estimate p_{DS} and p_H by maximum likelihood, using a dataset $\left\{\ell^*_{i,j\neq i}\right\}$ consisting of all pairwise comparisons to i.
2. If $\hat{p}_H = 0$, form $\hat{C}_i = 2/\left(1 + n\hat{p}_{DS}\right)$ and stop.
3. If $\hat{p}_H > 0$, find the fish i^* with the highest posterior probability of being a half-sib to i
4. Choose a set of fish $\mathcal{K}_i$ that is very likely to include all i's other half-sibs[35] besides i^*, based on posterior probabilities.
5. Form an augmented dataset $\{\ell_{i,j\neq i}, \ell_{i^*,\mathcal{K}_i}\}$ and maximize its likelihood, taking into account possible all possible 3-way sib-patterns between i, i^*, and each $k \in \mathcal{K}_i$ but still using only pairwise likelihood ratios. The parameters of the augmented likelihood are p_{DS}, p_H, p_D, and three or four additional nuisance parameters to cover the possibility that i^* and/or $k \in \mathcal{K}_i$ are not in fact half-sibs of i. There are 14 different possible 3-way relationships, so the augmented likelihood is quite complicated. In principle, we should re-estimate p_{DS} and p_H, but the extra pairwise comparisons do not involve i and carry essentially no information about those two parameters; it suffices to leave them as they were from step 1.

[34] If $m_{ij} = m_{ik} = m_{jk} = 1$ and i and j share a Dam, then k can't share its Sire with i unless the adults are hermaphrodites; see Appendix.

[35] In practice, an upper limit of 10 is quite adequate, since C_i is insensitive to small variations in p_D vs p_S

6. Form $\hat{C}_i = (1 + n(\hat{p}_{DS} + \hat{p}_D))^{-1} + (1 + n(\hat{p}_{DS} + \hat{p}_H - \hat{p}_D))^{-1}$

The nuisance parameters and complexity of the augmented likelihood seems unavoidable in order to avoid selective use of data. In all, though, estimating C_i requires at most a 4-D maximization, and the use of $O(n)$ data (pairwise likelihood ratios). Consequently, the overall estimation of P_J requires $O(n^2)$ operations, and is quite quick: just 1-2 minutes for 500 fish.

Bias in estimating Unique Parental Contribution In the above scheme, $\hat{C}_i$ cannot exceed 2, provided the estimated proportions are constrained to lie between 0 and 1. For "only-fish" where the true $C_i = 2$— which we hope are by far the most common type— then with finite loci it sometimes happens that $\hat{C}_i < 2$ just by chance. Thus $\hat{C}_i \leqslant C_i$ for only-fish, and the unadjusted MLE is biassed. Although the bias would eventually disappear as the number of loci increased, it is quite substantial at least with the 11 loci we used in this preliminary study.

To correct this bias, we allow fish to have an estimated UPC of slightly more than 2, to counterbalance the cases where the UPC estimate is wrongly less than 2. A simple way to do this is first to calculate $p_{lo} \triangleq \mathbb{P}\left[\hat{C}_i < 2 | C_i = 2\right]$ and$e_{lo} \triangleq \mathbb{E}\left[\hat{C}_i | \hat{C}_i < 2, C_i = 2\right]$ by simulation, using the same sample size and allele frequencies as in the real data. We then chose a bias-correction ε to satisfy the equation

$$p_{lo} e_{lo} + (1 - p_{lo})(2 + \varepsilon) = 2$$

For the real data, whenever $\hat{C}_i = 2$ we *replace* the 2 by $2 + \varepsilon$. With our 11 loci and $n \approx 500$, it turns out that $\varepsilon \approx 0.08$, so the bias correction is small. By allowing fish to have slightly more than 2 estimated parents, it is possible to correct the "bias against only-fish", at the cost of a minor violation of commonsense. Note that the *estimated* total number of parents can now exceed $2m_J$.

Bias should be much less of a problem when $C_i < 2$, because parameter estimates will typically be away from the boundary (the case $p_{DS} + p_H = 1$ would be catastrophic, since the entire sample would be close relatives, but fortunately does not occur for SBT).

Extrapolating beyond the subsample Suppose we want to estimate the number of parents of a larger sample from which our subsample of size $n + 1$ has been drawn. If the number of unsubsampled fish is u, so that the total size of the sample is $u + n + 1$, then we can reason as follows. First, how would the unsubsampled fish affect the existing $\hat{C}$'s? Since an unsubsampled fish has an unknown genotype, there is no information on its relatedness to any of the fish in the subsample, so that the estimates of p_{DS} etc. will be unchanged. However, the expression for $\hat{C}_i$ involves the total numbers of relatives of i, not the proportions, and the UPC estimate will be changed to this:

$$\hat{C}_i^u = \frac{1}{1+(n+u)\left(\hat{p}_{i,DS}+\hat{p}_{i,D}\right)} + \frac{1}{1+(n+u)\left(\hat{p}_{i,DS}+\hat{p}_{i,H}-\hat{p}_{i,D}\right)}$$

Second, what $\hat{C}$'s would the unsubsampled fish themselves contribute? There is no reason to think that they would be any different to the $\hat{C}$'s in the subsample, assuming the latter is random. Therefore, a sample from the distribution of P_J^u, the total UPC from the unsubsampled fish, can be obtained by taking a multinomial draw from the $\hat{C}_i^u$'s for the subsample[36]. The effect on the point estimate $\hat{P}_J^u$ is just to scale it up by the average across possible draws, so that

$$\hat{P}_J^u = \frac{u+n+1}{n+1}\sum_{i=1}^{n+1}\hat{C}_i^u$$

Variance It is simplest to start with the no-extrapolation case. Each $\hat{C}_i$ is a MLE, and its variance could be approximated by standard methods, although the very common boundary-value case $\hat{C}_i = 2+\varepsilon$ needs attention. However, the $\hat{C}_i$'s collectively are not independent, so the variance of their sum is not the sum of their variances, and the covariance between any pair is not easy to calculate. A "fish-level" bootstrap or jack-knife of the entire procedure for estimating P_J would get round the non-independence, but would be very slow with an $n = 500$ subsample (hours), let alone with $n = 3000$ (days).

The "sandwich method" ([**?**], [**?**]), which is widely used in econometrics and social science, is an attractive alternative that is robust to complex non-independence structures such as ours. An alternative name which better describes its application here, is "infinitesimal jack-knife". We can regard the aggregated estimate $\hat{P}_J$ as a function of a vector of weights w applied to the genotype data from each fish; the point estimate is obtained when all the weights are equal to 1. Writing $\hat{P}_J(w)$ for the resulting weight-dependent estimate, the infinitesimal jack-knife result is that, in a very general sense,

$$\mathbb{V}\left[\hat{P}_J\right] = \frac{d\hat{P}_J(w)}{dw}.\frac{d\hat{P}_J(w)}{dw}^{\top}$$

where the derivatives are evaluated at $w = 1$; they are the sensitivity of $\hat{P}_J(w)$ to the weights.

In our case, it is not obvious what might be meant by the "weight" attached to data from a fish. One way is to imagine that each fish might be genotyped at more or fewer loci than was actually the case; the weight could be proportional to the number of loci used. Each pairwise-log-likelihood-ratio is the sum of log-likelihood-ratios over independent loci, so using more or fewer loci would scale the observed likelihood ratio proportionately to the number of loci. If the

[36] This is an application of the "plug-in principle", as used to justify boostrapping

original point estimate is written as a function of all the pairwise-log-likelihoods $\{\ell_{ij}\}$, we might therefore define the weighted estimate as

$$\hat{P}_J\left(w;\{\ell_{ij}\}\right) \triangleq \hat{P}_J\left(1;\{w_i w_j \ell_{ij}\}\right)$$

i.e. applying the existing estimation procedure to a modified set of pairwise log-liklihoods.

Remarkably, the sensitivity vector $d\hat{P}_J\left(w\right)/dw$ can in principle be calculated by Automatic Differentiation (e.g. [**?**]) in almost the same time required to estimate $\hat{P}_J$ in the first place, regardless of n. The sandwich-method variance is thus far faster than bootstrapping or standard non-infinitesimal jack-knifing.

One other refinement is needed to deal with the boundary case $\hat{C}_i = 2 + \varepsilon$. The weighted estimate $\hat{P}_J\left(w\right)$ needs to change continuously as a function of w for the infinitesimal-jackknife theorem to apply; but if ε is fixed, then small changes to the weights will not move boundary-valued estimates off the boundary, and will not affect those $\hat{C}_i$ at all. We can handle this by replacing the fixed adjustment ε with a data-dependent adjustment that depends on the derivative of the log-likelihood at the boundary; details are omitted.

The variance calculation can be adapted easily when extrapolation is required, breaking down as follows:

$$\mathbb{V}\left[\hat{P}_J^u\right] = \mathbb{V}_U\left[\sum_{i=1}^{n+1} U_i \hat{C}_i^u\right] + \left(\frac{u+n+1}{n+1}\right)^2 \mathbb{V}\left[\sum_{i=1}^{n+1} \hat{C}_i^u\right]$$

where U is a Multinomial draw of size u from the set $\{1,\cdots,n+1\}$; the left-hand variance is easily computed from the moments of a multinomial distribution. The term on the right is computed via the infinitesimal jack-knife.

Scoring errors The above takes no account of the possibility of scoring error, which could be problematic. For example, null alleles leading to spurious homozygotes are liable to bias $\hat{P}_J$ downwards, because any two individuals that are (apparent) homozygotes for the same rare allele will appear much more likely to be sibs. The right place to handle scoring error would be when calculating the log-likelihood-ratios. The problem with handling "generic scoring error" is how to model it statistically, and in particular how to model its heritability. Also, in the absence of known parentage or sib groups, it is not easy to estimate the extent of general scoring error from within the sample.

It would be possible, though, to develop self-contained and logically consistent models for one type of scoring error: null alleles arising from a mutation in the primer sequences. Such mutations should be heritable, so an apparent homozygote can be modelled as an unobservable mixture of alleles subject to standard inheritance. It would not be hard to extend our methods to this case; perhaps some attention needs to be paid to dropouts, which could be "double nulls", although loci with a substantial proportion of dropouts are unlikely to have made it this far in a close-kin study.

Null alleles arising from short-allele-dominance can be accommodated in a parentage context (see following document), but are much harder to handle in a sibship context because an explicit model for the "dominance" is required. If a locus exhibits evidence of short-allele-dominance, it is best avoided when parent-counting.

19.1.4 Results of parent-counting

Using the 11 loci listed in last year's document, and a sample of 480 juveniles from the same cohort and year, the point estimate of P_J without extrapolation— i.e. compared to a maximum possible 960 distinct parents— is 895. Even if taken at face value, this is close enough to 960 that there is no need for concern about sample size feasibility in the project as a whole. However, two of the loci (D225 and D235) showed significant excess homozygosity, presumably through null alleles (see other document), and spurious homozygotes are liable to inflate the estimated proportion of sibs/halfsibs. With those loci removed, the point estimate jumps to 960 (actually, to 1035 which *exceeds* 960; but this is just a chance result arising from the bias correction).

Given this encouraging result, I have not yet carried through the (complicated) variance calculations, nor extrapolated the results to a sample of size 3000; there is no reason to think that sibship is common enough to threaten the viability of the project, which is all that needs to be settled at this stage. The question of variance, and of whether to consider working in terms of equation (10) rather than equation (8), will be examined once we have genotyped more juveniles from the cohort using a larger set of loci (and e.g. excluding loci with evidence of null alleles) .

Simulations suggest that the procedure should estimate P_J without much bias with these 11 loci and $n = 500$. However, bias (tendency to over-estimate P_J) can occur even with n as high as 100, at least when P_J is small. Limited experiments suggest that the bias arises because the juveniles' allele frequencies as estimates of the allele frequencies in the *contributing parents*, and this in turn biasses the calculation of log-likelihood-ratios. When P_J is very small, the allele frequencies can vary markedly between the parents and offspring; we are basically studying a subset of the population with a very small effective population size.

19.1.5 Appendix: Proof of parental countability

This algorithm applies when the relatedness (sib, halfsib, unrelated) of any pair of fish is known exactly. It considers each fish in turn, and increments the count of hitherto-unidentified parents by either 0, 1, or 2 accordingly. The key step is (3c), which is explained more fully after the algorithm

1. Loop through the fish, discarding any fish that is a full-sib of a fish already examined. Full-sibs have identical kin structures, so the second and subsequent full-sibs will contribute no new parents. After this step, there are no full-sib relationships to consider.

2. Colour all remaining fish white; they will progressively turn black as the algorithm proceeds. Initialize $P = 0$.

3. While there are any white fish left, pick one and check its half-sibness against all black fish only:

   (a) If it has no black half-sibs, set $P := P + 2$ because both of its parents are "new"

   (b) If it has one black half-sib, set $P := P + 1$ because its other parent is "new"

   (c) If it has more than one black half-sib, check whether all its black half-sibs are half-sibs of each other. If so, they must all come through the same parent, so the other parent is "new"; set $P := P + 1$. If not, both the mother and the father have already been found, so do not change P.

4. Colour that fish black, and return to step 3.

To see why (3c) works, suppose we have eliminated redundant full-sibs, and that animal i has two half-sibs j and k. If i and j are maternal half-sibs but i and k are paternal half-sibs, then j and k cannot share either a mother or a father; if they shared one, they would have to share both, and would be full-sibs to each other and to i. Thus, either j and k are non-sibs, in which case both parents of i are already "spoken for"; or, if they are half-sibs, they must share the same parent with i.[37]

When adults are hermaphroditic, matters are more complicated. Figure 15 shows two possible half-sib triads arising from different numbers of parents: 4 parents on the left, 3 on the right. However, the right-hand arrangement requires at least one parent to be a mother *and* a father (to different offspring).

Figure 15: Two ways for three animals to be half-sibs.

[37] There must be a better way to explain this...

19.2 SBT Close-Kin Abundance Estimation: update, Mark Bravington, Pete Grewe, Campbell Davies, December 2011

19.2.1 Introduction

The project is now entering its final phase. The "draft final report"— which for those unfamiliar with this particular Australian funding process, *is* the final report prior to being sent out for reviewers' comments and any consequent revision— is due 1/1/2012, though we will need a few weeks more as explained below. This will probably be the last Steering Committee meeting in the project, although there may be follow-up with individuals on particular aspects.

Last year, we had processed about 5000 genetic samples, analysed about 4000, and identified 7 POPs. Based on the rate of POP-finding, it was clear that by the time we'd completed our original planned sample size of 7000 fish, we would have a lot fewer POPs than our notional expectation when we began the project, and certainly not enough to make usefully precise abundance estimates. We were able to secure funding for doubling the sample size to 14000 fish, dipping into the "back catalogue" of collected-but-unprocessed fish in order to make up the numbers.

Ideally, to get the most "POPs per $", we would sample equal numbers of Indonesian (adult) and Port Lincoln (juvenile) fish. However, the number of additional Indonesian samples in the back catalogue was quite limited, and regulatory and changes and delays with export permits have precluded getting extra adult samples (though they have still been collected in Indonesia, just not exported). We have therefore had to concentrate more on juveniles to make up the numbers to 14000. Almost all juveniles were 3-year-olds[38], chosen deliberately from the length-frequency distributions to minimize the chance of including a big 2-yr-old or small 4-yr-old.

Most fish have now been scored at 26 loci, up from 21 last year; the 5000 last year have only been scored at the 21.

The genetic work is almost complete. 100% of our target of 14000 fish have now had their DNA extracted, PCRed, and run thru the sequencer at AGRF[39]. 97% have been scored at least provisionally (i.e. their genotype data are sitting on a computer), and 80% have been fully QCed post-scoring. All adults are completely and utterly finished, and the remaining gaps are juveniles, spread across all five years of sampling. All genotyping and checking will be complete by Xmas 2011. Results in this report are based on the 80%.

It has been rather a busy year.

[38] IE: a 3-yr-old from the 2007 season would have been spawned between Nov 2003 and Apr 2004.

[39] AGRF=Australian Genome Research Facility; we use the Adelaide branch.

	IN 05/06	In 06/07	In 07/08	In 08/09	In 09/10	Total	To do
Everything	216	1944	1776	736	1172	5844	0
	PL 2006	PL 2007	PL 2008	PL 2009	PL 2010	Total	To do
Extracted	1708	1792	1680	1536	1536	8252	0
AGRF'd	1708	1792	1680	1536	1536	8252	0
Scored	1708	1792	1488	1344	1536	7868	384
Full QC	748	1504	1104	1152	1056	5564	2688

Table 12: Status of genetic data, 9/12/2011

19.2.2 "Executive" summary

To cut to the chase: we have found 36 POPs, with no genetic ambiguities and no nasty surprises. Since the *juvenile* sample size will increase by about 50% once the remainder have been checked and incorporated into the database, we will presumably end up with 50-60 POPs. The immense task of QCing the genetics has taken longer than expected and leaves us about 6 weeks behind schedule, but on course to finish in Feb 2012.

One key element of abundance estimation is having otolith-based age estimates for all identified parents, i.e. adults in POPs. Only about 1/3 of adult otoliths are normally read for CCSBT purposes, so we have to make special arrangements for most of our POPs. Results for the so-far-known parents are expected today (9/12/2011). As soon as the remaining juveniles have been assimilated, the additional parents will be identified and their otoliths will be read by mid-Jan 2012.

This document gives some background on our QC and POP-finding procedures, describes initial results of looking for patterns in the POPs found (e.g. how big are parents, compared to typical adults?) and outlines the statistical mark-recapture model that we are planning to use to estimate abundance. The main task left is to actually estimate the abundance, which will require coding the model(s), fitting them, and exploring variations in parametrization.

19.2.3 QC and POP-finding

The goal of our QC and POP-finding process is to make sure that we have:

1. found as many POPs as we can be reasonably confident of;
2. not missed any POPs by mistake.

First, there is a whole suite of basic genetic QC (not fully described below). Once each plate/panel of 96 fish gets past that first layer of genetic checks, most of the remaining things that can go wrong will affect large numbers of fish at once, and we have developed a suite of checks for such large-scale errors. Finally, once we have weeded out and fixed all the large-scale errors, we needed to settle on good criteria for whether to bother actually making a POP comparison between a given pair of fish, and if so then for deciding whether the pair really is a POP—

bearing in mind that there may still be errors in the scores for individual fish at some loci.

19.2.4 The genotyping process

The basic genetic QC process is *not* described in detail here (it follows well-established lab protocols, and PG will provide details). However, it is worth outlining something about the whole genotying process, to give some idea of what can go wrong. Note that the terminology here is MVB's own, not geneticist-speak.

Each original sample of flesh from Indonesia or Port Lincoln has a tiny subsample of tissue removed, before being re-frozen for long-term storage. DNA is then extracted from the subsample and stored in liquid form in a vial. Vials are organized onto "plates" of 96 fish at a time. Plates will include a few Control fish and empty (water-only) vials. Each plate is then "PCRed" several times in Hobart[40], using a different batch of "primers", a different primer being used for each locus; each PCRing results in a different daughter "plate-panel" with DNA from 4-6 of the loci, ready to be sent to AGRF for "analysis of fragment separation". AGRF runs several QC checks of its own, and if all goes well then its 3730 sequencing machine sends back one set of computer files per plate-panel. That fileset contains graphs, one graph per locus and vial. Each locus will contain two alleles, characterized as the lengths of DNA fragments in integer units of "base pairs" (bp). Experienced staff ("readers") then use a program called GeneMapper to manually "score" each locus and vial in the fileset, i.e. to identify two (or one) peaks in the graph whose locations correspond to the lengths of the two alleles (only one peak if both alleles are the same), or to note that there's a problem and that the locus/vial is unscorable. Only certain locations for peaks are permitted to be chosen, unless the reader explicitly overrides; the collection of allowed peak locations for a locus is called a "bin-set", and has to be determined from experience of scoring that locus many times. The program GeneMapper automates most of the scoring, and the reader usually only has to confirm Genemapper's suggestion. Once a vial/locus has been scored in Genemapper, the storage of results is highly automated and not suceptible to further error. The GeneMapper scores then need to be reassembled across panels from the same plate, to form the entire genotype of a fish. Sometimes plate/panels are re-processed or re-scored, and some fish will occur on more than one plate for checking purposes. These multiple versions of genotypes need to be cross-checked, reconciled, and linked to the original records concerning collection of the fish (date, place, length, otolith age if known, etc.).

Many problems can be detected at or before the GeneMapper stage, and usually lead to a plate-panel being either rerun by AGRF, or completely regenerated in Hobart from the master plate.

Descriptions of most of the loci that we've used can be found in previous documents. We added more loci this year just to be on the safe side, and they

[40] For the first 5000 samples, as in last year's report, PCRs were done individually per locus by AGRF Adelaide.

have fairly similar properties to the existing set (e.g. all are tetra-nucleotide microsatellites, not di- or tri-). Most loci show some very slight excess of homozygotes, corresponding to true null alleles or failure to see both peaks in a heterozygote; in all but two cases, the excess is below 3 percentage points. With 14000 cases, the statistical power to detect even a minuscule excess of homozygotes is very high, so the surprise is perhaps that some loci had no excess at all. Two loci had extremely wide spreads of alleles (over 200bp) and would therefore be expected to suffer long-allele dropout, whereby a true heterozygote might appear as a homozygote. However, the very long alleles were not common at those loci.

19.2.5 Detecting large-scale genotyping errors

Because several mechanical processes, many people, and two laboratories 1000km apart are involved, a variety of "administrative errors" can in principle go wrong with the genotyping process. The two worst, in the sense of badly disrupting POP-finding, are:

- "chimaeras", whereby two sets of fish have had their genotypes inadvertently mixed up, and
- changes in scoring protocols and bin-sets for the same locus over time (i.e. between different plate-panels).

Note that any type of error is liable to lead to true POPs being falsely rejected, rather than to false-positive POPs sneaking in, because random changes to unrelated fish do not make it any more likely that they will appear related. (About the only exception is duplication of a sample between the adult and juvenile datasets; this very unlikely event is easy to check for, and didn't happen.)

Chimaeras can occur when the plate-panels from different original plates were inadvertently switched, so that some of the loci from one plate were ascribed to the other plate, and vice versa. We safeguarded against this by including an extra "barcode" locus scored on *all* plate-panels (normally, the set of loci on a panel are specific to that panel)[41]. It is then easy to check that the locus has scored consistently across all plate-panels from the same plate. Such swaps did in fact happen on two occasions (out of hundreds of plate-panels), including one occurrence in last year's 5000 fish, which wasn't picked up then since we did not have all the QC in place. Another type of chimaera occurs if an entire plate-panel is accidentally rotated thru 180 degrees, so that first & last vials etc are swapped for some loci. Again, the barcode locus allowed us to detect and correct the (one or two) occasions that this occurred.

One of the main headaches has been maintaining the consistency of bin-sets and scoring, especially since many new valid alleles (peak locations) came to light only *after* already scoring thousands of cases for a locus. Retrospective checks on allele and homozygote frequencies across the timespan of the project

[41] The safeguard is more elaborate for the original 5000 fish where we didn't use a barcode on all panels, but the result is the same.

did identify some changes, which in the worst four cases have led to re-scoring a locus for several thousand fish. The majority of the problems found date back to the first parts of the project, when the processing and scoring protocols were still settling down; there have been very few anomalies found among the most recent 7000 fish.

A variety of other checks were applied, including homozygote frequency by plate and locus, and allele frequency consistency between batches of fish[42]. A few problems were found, necessitating in some cases the re-scoring of existing AGRF filesets of some loci, and in others to the re-processing of some plates.

It's also possible that a few fish were mislabelled, and in particular swapped, especially during Indonesian sampling where more variables are recorded, and there is more to go wrong. The consequence might be that a few lengths and ages and sexes are wrongly recorded. If this happened to a fish in a POP, then it would slightly disrupt the *patterns* in the POPs but not the *number* of POPs, which is the key determinant of abundance estimate; thus any such errors would only have small impacts.

19.2.6 Individual-scale genotyping errors

Apart from large-scale "administrative errors" described in the next section and arising from the multi-step nature of the process above, there are three main classes of genotyping error specific to an individual vial/locus that are of particular relevance to POP-finding. The differing implications of these errors for POP-finding are discussed in the next subsection.

- Type-1 scoring error: identification of spurious peaks, leading to recorded heterozygotes that should have been homozygotes, or possibly even to the *wrong* heterozygote score being recorded. These errors should be very rare if the DNA quality is good and the readers are experienced and careful; genuine peaks have a characteristic appearance, our samples are collected and stored in very good conditions, and the loci we use are selected specifically with reliability of peak-identification in mind. "If in doubt, don't score the locus" is the practice we have established[43].

- Type-2 scoring errors, where the reader fails to detect a genuine (albeit often inconspicuous) peak, so that e.g. a true AB heterozygote is scored as AA. Scoring errors (type I or II) can often be picked up by re-examination in GeneMapper, but since there are about 280,000 locus scores in the database, any decisions to re-examine scores needs to be precisely targetted!

- *Process* errors in which the peak for B doesn't appear at all, and the reader correctly (on the evidence available) scores AA even though the

[42] The 14000 fish were organized into six batches for processing, and there were some protocol changes at the end of each batch.

[43] This policy would not necessarily be appropriate in other projects. For example, the rule might lead to disproportionately more true homozygotes being discarded, which would bias certain types of inference in "classical population genetics".

truth is *AB*. This can happen through "long-allele dropout", or through "true nulls" (sensu MVB): variations in the flanking sequence around the allele which cause the primer to fail to bind, and the PCR step to fail, and the allele to be overlooked by AGRF's machine. Long-allele dropout means that some particularly long alleles will occur in the *scores* much less often than in reality, and true-null alleles will not appear at all (though partial binding and sporadically successful PCRs are also possible). True nulls are heritable, and very relevant to POP-finding. It is not really possible to assess process errors by re-scoring.

Type-2 scoring errors and process errors both lead to a statistical excess of apparent homozygotes across all cases, but it is not possible to ascribe a cause; in theory, long-allele dropout is statistically detectable, but I did not find *statistical* evidence for it last year despite clear indications from the scoring process.

19.2.7 Finding POPs

Every juvenile inherits, at each locus, one allele from one parent and one from the other. Unless there is a mutuation (extremely rare), a POP must therefore have at least allele in common at every locus. Whether this is actually the case in the *scored* genotypes is another matter, because of the possibility of genotyping errors. The "default" approach to finding POPs would be to compare every adult and every juvenile, checking each locus to see if an allele is shared, and to deem the pair a POP if and only if no non-matching loci are found. Of course, an unrelated pair can certainly have an allele in common at any given locus just by chance, but if enough loci are checked then there is a very low chance (the false-positive probability) of this happening at all the loci. However, the default approach would run into severe problems, certainly with false positives (because many pairs can only be compared at a few loci) and potentially with false negatives (because of genotyping errors). A more practical approach needs to take account of the following issues. The ordering might seem a little strange, but makes sense with hindsight!

- [re false-negatives] Should we weaken the per-locus matching criterion to be robust to some or all of the genotyping errors in the previous section?
- [re false-positives] Given the per-locus criterion used, what should be the "entry criterion" for deciding whether a pair of fish contain enough mutual information to assess POPness reliably?
- [re false-negatives] Should we weaken the overall rule "all loci must match for the POP to count", so that failures at one or more loci are tolerated?

A further aspect, of course, is that it is possible to re-examine borderline cases of "looks like a POP but has a bit of a mismatch at one or two loci", provided these are few in number. The following sections say more about these issues, and the practical results are given afterwards. To cut to the chase again, we found that:

- the only genotyping error common enough to be worth weakening the per-locus criterion for, was true-nulls (a process error);
- the entry criterion could be adjusted quite easily to get high efficiency (i.e. using a high proportion of pairs) with low total expected false-positives;
- after re-examining the (very few) borderline cases, a couple of genotyping errors were fixed, and there was then no need to relax the overall rule.

Per-locus matching criterion The excess of homozygotes found at most loci means we have to allow for the possibility of true nulls (though the cause of the excess may be in other types of scoring error). If a parent contains one true-null allele and one "normal" allele A, and its offspring inherits the true-null plus a different allele B from the other parent, then the parent will score as AA and the offspring as BB. A "null-friendly" matching criterion[44] therefore treats AA-BB as a match. Clearly, there will be some genuine-AA and genuine-BB pairs compared, and these will slip through the net, just as any pairwise comparison might share an allele by chance. But since the homozygote excesses are pretty small, this relaxation only marginally increases the false-positive probability. It turns out that this kind of genotyping error is *relatively* common in our samples— rare, but certainly far commoner than any of the other kinds— and that the relaxation of the criterion is well worth doing.

A more drastic relaxation would be to assume that Type-2 scoring errors (missing one peak in a heterozygote) are common. In this case, an AB parent could be scored as AA, while its offspring passing on its B allele to an offspring who inherits a C from his other parent; if the B is seen this time, the offspring will score as BC. A rule that allows a match if *either* fish is an apparent homozygote would be robust to this kind of error. However, this weakening really does increase the false-positive probability, and would substantially reduce the number of valid comparisons we could make. It might be better to take the chance that the occasional true POP will fail at one locus thanks to a Type-2 error, and just to check borderline cases and perhaps relax the overall rule.

An intermediate approach is to make a LAD-friendly criterion (details last year), at least for those loci that might require it. However, this is considerably more fiddly and substantially complicates the computation of false-positive probabilities. It didn't prove necessary.

Criterion for bothering to compare Most fish have been examined at 25 loci. If two unrelated fish are successfully scored at all these loci, then the false-positive probability— i.e. of their having an allele in common at every locus, and thus seeming to be a POP just by chance— is $\sim 10^{-14}$. Since we are "only" making about $46,000,000 = 4.6 \times 10^7$ comparisons, false-positives would not even be worth discussing if all fish were scored at all loci. However, many fish end up missing scores at a few loci, because of processing problems or deliberate decisions not to score ambiguous-looking genotypes. This means that

[44] Called "flanking-robust" last year.

many pairwise comparisons are based on few enough loci that, even though their individual false-positive probabilities are still very small, the *overall* number of false-positives expected would become substantial if all such "weak" comparisons were included. Since the number of *true* positives is sure to be low (no more than say 100), it is vital to only consider "strong enough" pairwise comparisons in order to avoid contamination with false-positives.

Last year's POP-finding process used a fairly crude criterion, based on either completely including or completely excluding a fish depending on how many loci it was scored at. This year we have used a more refined criterion: first work out the false-positive probability for *each* pairwise comparison based on which loci are scored for both members of the pair, and then choose only those comparisons whose FP prob is below a threshold, chosen to ensure that the sum of the chosen comparisons' FP probs ends up equal to a pre-set limit on the expected number of false-positive POPs (say, 0.5 or 1 POP). This slightly complicates subsequent analysis, because not all fish are compared to all others. However, the complication is worthwhile because the FP-per-pair approach is mathematically optimal: it gets the maximum number of pairwise comparisons (and thus the maximum number of expected true POPs, and thus the minimum CV) out of the data, without excessive contamination from false-positives.

Because our sample sizes are extremely large, we have very good estimates of allele frequencies (including an allowance for frequency of true-nulls, which is one way to explain homozygote excess) and the false-positive probability calculations are very accurate.

Relaxing the overall rule This is an approach of last resort, only to be used if there are some almost-POPs which simply *must* be POPs because they match at so many loci that it could not occur by chance, but nevertheless still mismatch at one or maybe two. The borderline cases can be checked by hand, and the *nature* of the remaining mismatch is important in making a decision: two really clear heterozygotes with no allele in common surely has to indicate a non-POP, whatever the rest of the loci have to say, whereas an AA-BC pair may show some hint of a second peak in the AA fish.

19.2.8 POP-finding results

The simplest summary of these is to show all pairs compared[45], cross-classified into a table of (# loci compared) * (# loci that mismatch). Just to show the pattern, it is easiest to start with what happens comparing *all* juveniles to *all* other juveniles, ie no entry criterion— clearly there cannot be any true POPs involved to muddy the waters! If there are no genotyping errors (apart from true-nulls, which are allowed for in the per-locus rule), then POPs would appear in the LH column of Table 13; genotyping errors could cause some to spill over into the adjacent columns. Clearly, when few loci are involved there will be

[45] Pairs where the adult was caught before or during the season of the juvenile's birth are excluded. The "during" case is logically possible, but hard to analyse.

many unrelated pairs with no mismatching loci, so the entry criterion needs to be tight enough to exclude these.

.	F0	F1	F2	F3	F4	F5	F6	F7	...to F25
C0	.	.	.	.	.	.	.	.	
C1	7415	13026	.	.	.	.	.	.	
C2	2299	13967	12834	.	.	.	.	.	
...									
C8	14	165	1247	4710	10870	15523	13902	6959	
C9	4	86	706	3137	9450	19126	24445	19683	
C10	9	67	654	3786	13799	34468	58319	67284	
C11	.	48	512	3681	16496	50406	109191	165470	
C12	.	8	138	900	4270	15128	38037	69807	
C13	.	4	40	326	1473	5054	12323	22937	
C14	.	3	41	280	1419	5628	15600	34387	
C15	.	2	23	177	1089	4042	12685	29516	
C16	.	2	6	43	281	1460	5167	14376	
C17	.	.	2	15	168	1032	3708	11449	
C18	.	.	2	25	196	1259	5243	17522	
C19	.	.	13	31	234	1367	6239	22368	
C20	.	.	2	18	124	886	4603	18342	
C21	.	.	2	2	63	400	2351	10027	
C22	.	.	.	.	.	23	77	384	
C23	.	.	.	.	2	15	90	425	
C24	.	.	.	.	2	9	59	341	
C25	.	.	.	.	2	7	14	113	

Table 13: Juvenile-juvenile comparisons, tabulated by #loci compared (rows) and not sharing an allele (columns). Dot means 0. There are no POPs! Middle rows & RH cols removed.

The results for adult-juvenile comparisons are shown in Table 14, this time with an entry criterion that would allow about 1% average contamination by false-positive POPs overall. The pattern is similar except for the bottom-left corner, where POPs are evident. Inspection of the borderline cases reveals *no* cases of true POPs with more than an *AA-BB* genotyping error (which does not count as a mismatch, since we have weakened the per-locus criterion); whereas e.g. the 16*1 pair has a glaring *AB-CD* mismatch at the offending locus, and is surely not a true POP.

An earlier version of this table led us to check on a number of borderline cases. It turned out that one fish in a near-POP was mis-scored at one locus as *AA* that should have been *AB,* which did then match the corresponding locus in the other fish; they are now one of the 23*0 pair. Another pair had an ambiguous locus in both fish that should not have been scored in either; it is now the 16*0 pair. A couple of other borderline cases also had missed-heterozygote scoring errors, but these did not affect their status— they were still clearly non-POPs.

All re-scoring was done "blind" without knowing what the putative partner of each fish was, and thus without knowing which alleles to "hunt for".

.	F0	F1	F2	F3	F4	F5	F6	F7	...to F25
C11	.	.	.	.	.	.	1	.	
C12	.	3	32	239	1405	5548	15552	32316	
C13	.	2	15	105	581	2193	6085	12265	
C14	.	4	47	419	2148	8421	24503	52431	
C15	.	8	42	447	2292	9756	30900	73794	
C16	1	1	10	95	682	3437	12453	34554	
C17	1	.	5	70	440	2564	10094	31461	
C18	4	.	5	41	360	2041	8905	30219	
C19	7	.	.	27	219	1412	6635	24313	
C20	2	.	1	12	114	728	3926	15806	
C21	13	.	.	5	52	393	2068	9195	
C22	.	.	.	.	3	25	99	445	
C23	3	.	.	.	2	7	89	486	
C24	.	.	.	.	2	13	55	333	
C25	5	.	1	.	1	4	13	117	

Table 14: Adult-juvenile comparisons, tabulated by #loci compared (rows) and not sharing an allele, i.e. inconsistent with POPhood (columns). Dot means 0. Null-friendly criterion; no other mismatches tolerated; entry criterion set so that expected number of false-positives=0.35. No comparison with <11 loci passes the entry criterion. Columns on right deleted.

Genotyping error rates inferred from the POPs Of the 36 POPs, 25 match using the strictest $AA \neq BB$ criterion at all loci, 9 require an $AA \sim BB$ relaxation at a single locus, and 1 requires it at two loci. That amounts to 9 null-like errors in about 770 *pairs* of locus comparisons (it's not clear whether the error, insofar as it is an error, applies to one or both fish), ie somewhere around the 1-2% mark— but still worth relaxing the match criterion for, since it avoids making a goodly number of POPs into slightly-false-negatives. Only one AB was miscalled as an AA among the true POPs— a rate of 1 in about 32*22*2=1500 scores. No AB-as-CD errors were observed.

There are two points to emphasize in all this:

- we have scored enough loci to set an entry criterion strict enough to effectively exclude false-positives while still retaining about 95% of possible comparisons;
- the best of the near-false-positives are far enough away from the true POPs that the number of borderline cases to check is small (again because we have a lot of loci)
- genotyping errors are not all equal, and aside from the "true-null or looka-like" are quite rare in this study.

A similar cross-check just between adults does reveal one adult-adult match (no mismatching loci), between one female of 183cm aged 24 and another of 177cm (unknown age) caught two years later. It is fully consistent with growth & maturity data to suppose that the bigger fish spawned at say age 10 a female offspring who 16 years later had grown to 177cm. Unfortunately, it would (will!) require a *lot* more data before we accumulate a useful number of adult-adult POPs. The number of adult-adult POPs is sure to be far lower than the number of adult-juvenile POPs, because at least 10 years of mortality and incoming recruitment has greatly diluted the "grandparent generation". To put it another way: if we had otolith ages for all adults, then we should only be considering adult-adult comparisons where one fish is at least about 10 years older than the other. This would drastically reduce the number of valid adult-adult comparisons vs adult-juvenile comparisons.

How many POPs are there? 36 so far, and presumably 50-60 once all the juveniles have been assimilated. The more difficult question is: how many comparisons were considered to get the 36? This depends on the entry criterion of "false-pos probability no more than X", which in turn depends on the number of expected false positives tolerated. The results shown are for 0.35 expected false-positives, i.e. about a 1% (negative) bias in the abundance estimate. This left about 27,000,000 out of a maximum[46] possible 29,000,000 comparisons, so it is quite efficient. With an expected 0.35 false positives, there is some chance (about 0.3) that one of the existing POPs is actually false, though more than one is unlikely. Choosing the entry criterion always entails a bias-variance tradeoff, and too tight a criterion means losing some true POPs as well as any false ones. For example, the comparison leading to the 16*0 POP (which looks solid but of course *could* be false) disappears if the criterion is substantially tightened. The average number of POPs-per-comparison is unaffected by the choice of criterion, except through random noise, but fewer POPs overall means higher CV. We do not have an overabundance of POPs in this project, and a 1% bias seems a price well worth paying for keeping the "actual sample size" high.

When doing "back-of-the-envelope" comparisons between the number of POPs and the number of comparisons, it is important to note that many of the adults in our Indonesian samples are small fish that might make little contribution to the spawning potential (if indeed they were on the spawning grounds when caught; see earlier footnote), and in particular might not have been mature *at all* in the year of birth of many of the juveniles. Although the current SBT stock assessment is far from the last word on SBT maturity, its assumption of knife-edge maturity at age 10 corresponds to an average length of ~155cm (undifferentiated by sex), and a substantial proportion of our adults are below that "threshold". The proportion would be even greater if allowance was made for growth between juvenile birth and adult capture. Section 4 describes a systematic framework for addressing this and other complications that affect close-kin

[46] After excluding about 2,500,000 pairs where the adult was caught prior to or in the same season as the juvenile was born.

abundance estimation for SBT.

Figure 16: Length frequencies of the adult samples

19.2.9 Patterns in POPs

Are there any (half-)siblings among the POP juveniles? No.

In other words, none of the POP adults match to more than one juvenile. That is a good thing, because if (half-)sibs are common among the *sampled* juveniles, then comparisons become non-independent. The adult abundance estimate still wouldn't be biassed (see below and previous explanations), but its variance would increase, potentially to the point of rendering the estimate useless. A preliminary check just among juveniles in 2010[47] indicated that (half-)sibs could not be *very* common (a critical decision point for the project), and the 7 POPs found in 2010 contained no sibs or half-sibs. Having found none in this much larger set of POPs, we can maybe conclude that (half-)sibs are rare enough *among our juvenile samples* for their effects on variance to be ignored; see section 4 for more details. This is not to say that (half-)sibs are at all rare among *all* 3-year-olds, but simply that our juvenile samples are a very small fraction of the total, and are well-enough-mixed to rarify the sib-pairs. As an academic exercise, it will at some point be interesting to re-run the juvenile-only sib check with the greatly expanded set of loci now available.

Size/age The parents (i.e. adults that are in POPs) are somewhat bigger than average among all adults in our samples (LH pair of graphs). Much of this is due to adults below 160cm; the smallest parents found were 159cm (female) and 161cm (male) (ages not yet known for these)[48]. When adults below this size are excluded, it is still true that female parents are bigger than female adults, but not so for males (RH pair of graphs).

Figure 17: Q-Q plots of parent (Y) vs adult (X) length; males & females; RH column only for adults above min parental size

These graphs should not be over-interpreted in terms of big fish being more fecund; the "right" comparison would be between parental size *in the year of offspring birth*, whereas the graphs show parental size in the year of *capture*. Capture is often several years later, so the parents have had extra time to grow. This cannot be resolved until the parental otolith ages are available,

[47] See the "Steering committee update" from May 2010

[48] In 2005/6 and 2006/7, some Indonesian boats operated south of the spawning grounds, and their catch was "contaminated" by some non-spawning adults. We have not yet checked whether any of our adult samples were affected.

because individual SBT clearly have very different asymptotic sizes (see age-length distributions below), and the size of a fish N years ago depends strongly on its age as well as its current length.

Figure 18: Lengths at age in adult samples (1350 otoliths read)

It is notable that the offspring of the 159cm female parent was born 6 years before the parent was caught. The typical age of a 159cm female appears to be about 10-11 from the age-length graph, but that is clearly incompatible with being mature 6 years previously. However, it's quite possible to have a 20-yr-old female of 159cm (the smallest females of age 20 are a little below 159cm), in which case she would have been 14 at time of birth— long-since mature— and her length would have been around just over 150cm (looking at the smallest females of age 14, and assuming fish "follow the quantile" of growth)— again plausible for maturity. Little more of substance can be said until the parental otoliths have been read; but the relationship between size, age, and maturity for SBT is glossed over in the current stock assessment, and one of the byproducts of this study should be a firmer handle on "real" spawning stock biomass, and effective fecundity at age.

Is skip-spawning common? Probably not.

If SBT spawn only every other year, say, then all POPs would have an odd number of years between offspring birth and parental capture. There is no indication of this in Table 15.

	1	2	3	4	5	6
M	4	9	5	3	1	1
F	3	1	3	4	0	2
Total	7	10	8	7	1	3

Table 15: Number of POPs by gap-in-years between offspring birth & parental capture. 0-year gaps not checked yet. Should really include the "null distribution" of year-gaps across *all* comparisons, but the conclusion about skip-spawning won't change.

One might nevertheless wonder whether skip-spawning is prevalent amongst younger/smaller fish, because Table 15 would hide this if older/bigger fish dominate the production of offspring. In fact, a graph of adult size vs gap shows no pattern at all, whereas one might expect smaller average gaps for bigger fish. Further checks await the advent of the otolith ages, but the evidence so far points away from skip-spawning.

Note that skip-spawning, even if it does turn out to be present, poses no particular difficulty for close-kin abundance estimation[49]. The point is rather

[49]Provided the study lasts long enough to comfortably cover the skip period, which is certainly the case here; see CCSBT-SC/0709/18 from 2007.

that it needs to be allowed for in estimation; knowing that young fish bred only every second year would, for example, affect the probability-of-capture equations in section 4 in a fairly straightforward way.

Is there any indication of "stock separation by time"? No.

Figure 19: Date of capture for parents (above) and parent-sized adults (below). NB small sample sizes on top!

To get an abundance estimate from close-kin, it's necessary to assume that adults are sampled randomly; or, if non-randomly, then at least in a way that can be compensated for through a measured covariate (e.g. if catchability is linked to age or size; see section 4). One possible outcome of this study– and one devoutly not to be wished for– was that all the POP parents would have been caught in a limited part of the Indonesian fishing season. This would point to a consistent behaviour whereby juveniles in the GAB would come from one part of the adult population, and the rest of the adult population would be generating a different pool of juveniles (which might or might not survive). If this adult behaviour was heritable, it would constitute stock separation— albeit by time rather than space (spawning ground). Heritable or not, the implications for SBT management would have been enormous.

So it is good news that the distribution of parental catch date matches that of adults in general. Since the Indonesian fishing season encompasses the entire spawning season and spawning ground[50], it seems reasonable to assume random sampling of adults (except as affected by age, length, and sex).

19.2.10 Sex ratio of parents

The 36 POP parents included 23 males and 13 females, or 64% male. This is quite different to the sex ratio in our genotyped *adults*, of 44% male. However, parents are bigger than average among adults, and the sex ratio in the adults changes rapidly towards males at larger sizes (e.g. 75% male above 180cm), so some of difference between the parent and adult sex ratio is due to size effects. The sex ratio in adults "corrected" to match the length distribution in parents is 53%, not significantly different to the actual parent sex ratio of 64% (p=0.13) although there is still a mild preponderance of males.

Because of the sexual dimorphism in SBT adult growth, and quite likely in fecundity-size-relationships, it is most sensible to estimate abundance separately by sex. As discussed next, some parameters may reasonably be assumed equal for males and females, but abundance is not necessarily among them.

[50] *Almost* the entire spawning ground; we miss out on sampling the small proportion of catches taken by the Cilicap-based fleet.

19.2.11 How to REALLY estimate adult abundance

The cartoon version of close-kin abundance estimation is shown below. Each juvenile has two links into the adults (its parents); genotyping a juvenile "tags" the two adults it is linked to; genotyping one adult at random has a chance 2/ (# adults) of matching one of those tags. In the cartoon, it is absolutely true that an unbiased estimate[51] of the number of adults is given by 2∗(#juves sampled)∗ (#adults sampled) / (# POPs found).

Figure 20: Simplest case of close-kin abundance estimation. Each juvenile has two parents, though adults have different numbers of offspring.

The reality for SBT is not all that much worse, but there are four linked complications that do need to be addressed when developing a proper abundance estimate:

1. **What is "an adult"?**
2. **Sampling probability** of an adult is **related to fecundity**
3. **Time delay** between sampling a juvenile and sampling the adults that might be its parents;
4. **Time series** of juvenile samples, rather than a single year's sample.

It is worth noting that none of these is a surprise. All were discussed in the original planning document CCSBT-SC/0709/18, although the time-series aspect gets only a perfunctory mention since the original plan was for a shorter project. As may become clear, there are some benefits to have the time-series of juveniles born in different years, despite the associated increase in complexity.

The potential problem with time delay is illustrated by the small number of adult-adult compared to adult-juvenile POPs found. As time elapses since the birth of a juvenile cohort, the "cohort" of real potential parents starts dying off and is diluted by incoming "impossible" parents. The unadjusted cartoon estimate of abundance then comes out too high. The solution is to eliminate the impossible parents by accounting for parental age, restricting comparisons to the same cohort of adults regardless of elapsed time. This means using an additional age-based "entry criterion" for comparisons and, since we don't have otolith ages for most adult SBT, some probabilistic length-based chicanery will be required. Fortunately, the age-length data for adult SBT is extensive enough to do this reasonably well.

The issue of "what is an adult?" is related. Unless the maturity ogive with age is truly knife-edged (as assumed in the stock assessment), there will be a small proportion of very young fish that contribute just a few offspring. We will have little information on the abundance of these youngest adults, precisely

[51] *Almost* unbiased; technically, it's the solution of an "unbiased estimating equation", which is not bad.

because they occur in few POPs. However, they are sure to be numerically abundant because they have had less time to die. Consequently, if we set the threshold for adulthood too low, our abundance estimate will be dominated by age classes that we can't estimate precisely. Setting the threshold too high, on the other hand, means doing a poor job of estimating "the" spawning stock biomass— itself a quantity whose appropriate definition this project should help with.

The link between sampling probability and fecundity arises because both are affected by (most likely proportional to) residence time on the spawning grounds. Fecundity in SBT females (not males) was quite well-studied in the 1990s, though not recently. SBT are batch spawners who take a few days to recharge the ovaries between spawning events; the gap between events is independent of fish size, but the number of eggs released per batch is size-dependent. Females will spawn multiple times during a season, but there is no direct evidence on how often, or on how this is related to size. However, the spawning grounds off Indonesia are not good habitat for *adult* SBT (albeit ideal for larvae), and the adults leaving the grounds are thin and bedraggled compared to those arriving, so it is reasonable to assume that fish stay as long as they can bear to, and that bigger fish can endure the warm waters for longer. Bigger females contribute more eggs both through generating more per spawning event, and through having more events by staying longer. In staying longer, though, they are of course more likely to be caught, and thus to be "recaptured". They are also more likely to be "tagged", in the close-kin cartoon sense, through generating more offspring. From a mark-recapture perspective, they are "trap-happy": more likely to be marked *and* recaptured. This phenomenon always causes some trouble in mark-recapture, and its consequences are explored below in the algebra.

A model framework The aim of the exercise is to obtain a "fishery-independent" estimate of adult abundance without using the two usual mainstays of standard fish stock assessments (total catch, and catch per unit effort), because of long-standing issues with these datasets for for SBT. However, we do still need to use Indonesian age-composition (or length-composition) data, as well as the ages/lengths of the identified parents.

The clearest way to present the model, is to imagine that we have an enormous number of POPs available over a great many years, and then to write down the equations governing the number and patterns expected in them in full generality. For actual purposes of estimation, we will have a limited number of POPs, and it will be necessary to cut some corners to keep the number of parameters-to-be-estimated in sensible proportion to the data: e.g., by assuming functional forms for some relationships so they can be described in terms of just one parameter, by assuming certain things are steady-state or constant over the limited time period of the study, and/or by assuming certain quantities are independent of age. But writing down the full approach makes it easier to consider which assumptions might and might not be tenable.

For simplicity, this entire derivation is entirely age- (not length-) based, ignores any restriction to certain age classes of adult, and assumes that age is available for *all* adults sampled, not just the parents. Complete allowance for length-based effects would be very difficult, but a statistical compensation for only having length for most adults is not hard. We also ignore skip-spawning, which is really just a presentational detail that could be easily handled quite easily. The derivation is also presented only in terms of female adults; males are discussed afterwards. The idea of this derivation is to explain the concepts and "where the information comes from"; the actual way that parameter estimates is obtained is quite different, by embedding the probabilities in a maximum-likelihood framework.

Consider a juvenile born in year 0. We will check it for POPhood against each all sampled adult females of a variety of ages caught in year 1, year 2, etc. The probability that a comparison with a female aged A in year Y will show her to be the mother, is

$$\mathbb{P}\left[\text{real mother was aged } A\text{-}Y \text{ at birth}\right] \times \mathbb{P}\left[\text{this female is the real mother}|\text{r.m.w.a.}A-Y\text{a.b.}\right]$$
$$= \frac{N_{0,A-Y}\phi_{A-Y}}{\sum_{a>0} N_{0a}\phi_a} \times \frac{1}{N_{0,A-Y}} \tag{13}$$

where ϕ_a is the relative fecundity of females aged a. The term on the left is the probability that the offspring's mother comes from the "right" cohort; the term on the right is about this particular female then being the one from that cohort to spawn this particular offspring. Because of the age-fecundity link, we cannot simply "add up and cancel" these equations across A to get back to the female-only version of the cartoon formula, $\mathbb{E}\left[\#\text{maternal POPs}\right] = 1/\sum_{a>0} N_{0,a}$; we could only do that if ϕ was independent of age. However, we also know the age composition of female adults (not parents) caught in year 0. If q_a is catchability at age a, then the probability of a female adult in the year-0 sample (not the parents) being age A^* is

$$\mathbb{P}\left[\text{sample age} = A^*\right] = \frac{N_{0,A^*}q_{A^*}}{\sum_{a>0} N_{0,a}q_a} \tag{14}$$

Now make the assumptions that catchability is proportional to residence time, and that fecundity is proportional to residence time (i.e. number of spawing events) multiplied by eggs per spawning event, which has already been estimated as a function of age from fecundity work[52]. In other words:

$$q_a \propto r_a$$
$$\phi_a \propto r_a b_a$$

where b_a is a *known* parameter denoting age-specific relative batch fecundity, i.e. number of eggs per spawning event. Substituting this into the previous

[52]Some work— and assumptions— may be required to update the results of the 1990s fecundity study, given the possibility of growth changes since then.

equations, we get:

$$\mathbb{P}\left[\text{r.m.w.a.}A-Y\text{a.b.}\right]=\frac{N_{0,A-Y}r_{A-Y}b_{A-Y}}{\sum_{a>0}N_{0a}r_a b_a}\times\frac{1}{N_{0,A-Y}} \tag{15}$$

$$\mathbb{P}\left[\text{sample age}=A^*\right]=\frac{N_{0,A^*}r_{A^*}}{\sum_{a>0}N_{0,a}r_a}$$

Thus, the sampled age composition tells us relative$N_{0a}r_a$ as a function of a; if we then scale this up by $b_a$and substitute this into the parental age composition and the *numbers* of POPs given the number of comparisons, we can directly estimate N_a.

To formalize this, note that the likelihood consists of two parts, one coming from the age-compositions of the female adult samples (a multinomial distribution) and the other from the juvenile "recapture histories". To obtain the latter, consider all the possible outcomes from checking juvenile i born in year 0 against females of different ages in year Y, assuming that i's mother was not already found in any previous year. One outcome is "not found"; another is "found to be a female of age 1", another is "f.t.b.a.f.o.a.2", etc. The total number of outcomes is of course 1, and the likelihood contribution for that year is a multinomial probability with "size=1" and the probability for the outcome "checked females aged A include$i$'s mother " being

$$(\#\text{FadComp}_{iYA})\times\frac{\phi_{A-Y}}{\sum_{a>0}N_{0a}\phi_a}$$

where $\#\text{FadComp}_{iYA}$ is number of female adults aged Y in year A that are compared with this particular juvenile. The probability of the mother *not* being found at all that year is one minus the sum of the other probabilities. Note that $\#\text{FadComp}_{iYA}$ will differ between juveniles born in the same year because they will have different missing loci, and thus different sets of female adults with whom they pass the "entry criterion" for POP-checking.

To form the entire likelihood for that juvenile, we need

$$\begin{gathered}\mathbb{P}\left[\text{outcome in year 1}\right]\\ \times\mathbb{P}\left[\text{outcome in year 2}|\text{mum not found in year 1}\right]\\ \times\cdots\\ \times\mathbb{P}\left[\text{outcome in year }Y|\text{mum not found earlier}\right]\end{gathered}$$

where year Y is either the year its mother *was* found, or the final year of the study otherwise. (Once the mother is found, there is no information from comparisons with subsequent years.) Technically, the probabilities after the first year are all conditional, but the conditioning merely amounts to saying "we already checked a very small fraction of the females from each cohort, and the mother wasn't among them" so it makes very little difference and can be ignored in practice; the overall probability of finding either parent across all years of the study is only about $36/5600\approx 0.6\%$.

The question then arises of how to combine the likelihoods arising from the different juveniles. It is by far easiest computationally to treat them as independent. However, if there are sibs or half-sibs in the juvenile samples, then their comparisons are not independent. Since there is evidence that (half-)sibship is low (0 in 36) amongst the juveniles, it may be reasonable to just ignore this.

It turns out (not shown here) that we can in principle use the time-series nature of juvenile birth dates to help diagnose whether the residence assumptions are accurate. The age-profile of identified parents from juveniles born in year 1 should differ predictably from that in year 0 because of the change in *relative* fecundity at age [53]. The data should eventually show whether this prediction is borne out, but it might require more than the 6-year span of juvenile birthdates in this study, and/or merely more POPs.

It also seems to turn out (not shown here either, and again in principle) that a straightforward direct estimate of adult mortality z can be made *post hoc* (unless it is built into the parametrization of the model. The idea is to compare an estimate of the number of adults aged $\geqslant A+1$ based on juveniles born in year 1, with the estimated number of adults aged $\geqslant A$ based on juveniles born in year 0. Whether there is any aliassing with the treatment of new adults from incoming cohorts, though, requires further exploration. Note that, because we are avoiding CPUE and thus not making any assumptions about the amount of "relative effort" required each year to collect our adult samples, it is not possible to construct time-at-liberty estimates of z based on the difference between "tag date" (juvenile birth-year) and "recapture date" (adult capture-year).

What about males? We have no data on "batch fecundity" for males. However, it seems reasonable to assume that males of given size can endure Indonesian conditions (i.e. the daily weight loss) as long as females of the same size, so that the r_{Fa}estimated from females can be converted into the male equivalent r_{Ma} after allowing for size differences. Thus, for females we use b_{Fa}as an input to estimate r_{Fa}; for males, we use r_{Ma} as an input to estimate b_{Ma}. In both cases, the desideratum is actually N, but it is necessary to deal with r and b en route.

An adult-oriented alternative Another approach to estimation is to consider in turn each adult, and to formulate the probability distribution of the number of its offspring found by birth-year, given the sex, age, and perhaps length of the adult when caught. The multinomial contribution from adult age distribution is unchanged, but the capture-history contribution is quite different. The idea here is that, for a female adult j caught in year Y aged A, we can

[53]It would also be affected by differential-across-age mortality in the adults, but it is probably OK to assume this is not the case or not severe. For what it's worth, the stock assessment mortality estimates show no age effect among adults until the fish reach their late 20s. Although we deliberately avoid using stock assessment *data* in this analysis, apart from age composition, it might be OK to appeal to qualitative results from the assessment.

write the expected number of offspring found in any previous birth-date-year T as

$$\mathbb{E}\left[\#\text{off}_{jT}\right] = \left(\#\text{juComp}_{jT}\right) \times \frac{\phi_{A-(Y-T)}}{\sum_{a>0} N_{Ta}\phi_a} \tag{16}$$

where $\#\text{juComp}_{jT}$ is the number of juveniles born in year T that get compared with our female j. Note that we also need to consider adults that are not part of POPs, and for many of these the age will be unknown, so that an "integration" over likely age given length would be required.

The adult-centric approach is mathematically very similar to the juvenile-centric version, but perhaps harder to explain because the elegant simplicity of "each juvenile had two parents" has been lost. However, it does have two attractive features:

- it is much easier to allow for length effects on fecundity (although this is not an immediately appealing prospect for the statistician);
- it *is* reasonable to assume that "offspring capture histories" are truly independent between adults, even if juveniles contain sibs or half-sibs; whereas the independence of "parent capture histories" across different juveniles is suspect in principle, even though not contra-indicated by our results so far.

The flipside of the second assumption is that equation (16) only describes mean values— it does not constitute a probability distribution, and some distributional assumption must be made in order to develop a likelihood. The choice will affect the CV of the final answer, though not (much) the point estimate. The obvious assumption is Binomial with size $\#\text{juComp}_{jT}$, which might as well be approximated as Poisson since the expected values are so small, but there is an implicit assumption of random reproductive contribution (in the juvenile samples, not necessarily in the population). If we knew the extent of beyond-random reproductive variability, e.g. from a sib-based juvenile-only parent-counting exercise like that shown last year, then we could pick a suitably over-dispersed alternative to the Binomial/Poisson distribution. But for the moment, since we have no sibs or half-sibs among the juvenile POPs, there is no obvious basis to assume anything other than Binomial/Poisson.

Fitting the model To fit either the juvenile- or adult-oriented model, we need to parametrize most or all of the relationships. Instead of trying to estimate each N_{YA}, for example, we assume a parametric relationship such as $N_{YA} = N_{00}e^{-\alpha_1 Y}e^{-\alpha_2 A}$, reducing many N_{YA}s to 3 parameters N_{00}, α_1, and α_2. A number of other details will require further attention: for example, dealing with length-age issues, setting up b_{Fa}, how to handle incoming recruitment, and the detailed impact of trying to ignore or somehow "lump" young/small adults. Because the timespan covered is much shorter than what stock assessments have to deal with (decades), it *may* be acceptable to make some steady-state

assumptions over the period considered (e.g. mortality ~= recruitment), or similar. Until we actually begin estimation, it is not clear how much "the data will stand", in terms of how many parameters can be estimated.

Although the residence-time/fecundity complication does make the task of estimation substantially more formidable, it *may* not make that much difference in practice. Equations (4.1) and (4.3) still involve a "1/N" term, so the number of POPs is still inverse to the abundance estimate. All that "fecundity" does, is to change the relative weights of the expected numbers of POPs from the different parental age classes when estimating an "aggregate N". The closer the real maturity ogive is to knife-edge, the closer the task will be to the cartoon (since we could be sure we were checking the right "cohort" of potential parents in each year), and the smaller will be any increase in CV arising from these complications. And if we are willing to estimate some kind of age-weighted N— for example, deliberately weighted by age-specific fecundity— then there may not be sensitivity to residence-time-related assumptions or much cost in CV at all.

Ultimately, though not as part of this project, it will be desirable to integrate the close-kin estimates into a full stock assessment, which should help out with the weakest part of the close-kin model (handling the youngest/smallest adults, with their low per capitata fecundity); and the converse is certainly true, that the close-kin estimates should help out considerably with the rest of the stock assessment! But for now it is important to do the close-kin estimation quite separately from the stock assessment, to provide a check on the structural assumptions of stock assessment by avoiding most of its "data issues".

Assumptions Fitting the above model will mean reducing the number of free parameters, as just explained, and this inevitably will entail *some* assumptions: for example, perhaps assuming roughly constant mortality rates over the 6-year-period of juvenile birthdates. On the whole, though, the assumption load is rather light. The main complication is the need to *assume* the nature of the relationship betweens residence time, fecundity, and catchability, and then to estimate the parameters of the relationship between residence time and age, rather than having direct evidence about these things. Luckily, the existing fecundity work does give us a way in, albeit rather indirectly, and these assumptions do seem biologically reasonable. But they *are* still assumptions. This is a good moment to point out that we ***could*** obtain residence-time-at-size-and-age data directly over the next few years, by archival-tagging a goodly number of adult or near-adult SBT and examining archived tags returned over the following years from the Indonesian fishery (presumably via our sampling program in Indonesia). The archival tags would provide information on residence time as a function of size, and also on time-at-depth, which relates to catchability. Existing archival tag deployments have almost all been on small SBT, and few taggees will survive to adulthood with functioning tags. The Indonesian connection provides a much better prospect of getting tag returns than some of the other SBT fisheries worldwide, and the fishing and natural mortality rates on

adults (each estimated at around 0.1 per year) are adequate to suggest a clear result could be obtained from moderate numbers of archival tags within a few years, assuming enough tags can be deployed on big-enough fish. Given the potential of close-kin to provide a rather cheap and tamper-proof monitoring tool for the future of SBT, the investment in a ground-truthing archival tag study would be worth considering.

19.3 SBT Close-Kin abundance: Final Steering Committee Brief, Mark Bravington, Pete Grewe, Campbell Davies, May 2012

This paper should be the final Steering Committee update for this project. It summarizes the genetic results (for background detail, see earlier reports), describes the process of turning the genetics into an abundance estimate— which has become a self-contained adult-abundance stock assessment— and presents preliminary abundance estimates. While the original project plan did not cover developing an actual assessment, it became necessary because of the extended period over which the samples were collected, and the consequent need to allow for e.g. adult mortality. The assessment combines the close-kin data with age and length samples from the Indonesian fishery, but does not use any total-catch or CPUE data. Using a formal statistical framework for assessment makes the most of the information in the data, and is conceptually consistent with, but independent of, the OM[54]. The assessment framework will facilitate consideration of the results by the CCSBT Scientific Committee, and the incorporation of the results into the OM.

Our CK abundance estimates of SSB or adult numbers in the early 2000s, based on close-to-steady-state (i.e. constant abundance over time) versions of the assessment, are considerably bigger than the estimates in the OM. For any one year, a wide range of estimates can be obtained from CK using non-steady-state versions, but the time-averaged abundance during the 2000s seems more consistent across runs. Although CV calculations are not yet complete, the CV on time-averaged abundance seems likely to be respectable, say around 25%; note that this still gives considerable latitude in comparisons with the OM since the two estimates are largely statistically independent. CK's estimated annual survival is similar to the OM's, but the relationship between female bodyweight and female fecundity (i.e. annual reproductive output) is rather different to the direct proportionality *assumed* in the OM, with older fish relatively more important according to CK. That is important when comparing the CK and OM abundances or SSBs: the meaning of "SSB" is in a sense quite different between the two models, and merely enforcing a different bodyweight/fecundity relationship in the OM might substantially change its estimates of SSB and adult-abundance, without directly using the POPs at all.

In all, the project seems to have worked well. The genotyping has been thorough and has consequently allowed us to make very efficient use of the large numbers of samples collected. There seems to be little room for doubt in the number of POPs found, and enough have been found to make a defensible independent estimate of adult abundance. The focus of this document is on what we have found to date, but it is also worth considering how we could build on the existing foundations in future. Now that we have streamlined the genotyping and built up a large "back catalogue" of genotyped fish, genotyping will become relatively cheaper (i.e. fewer $ per POP), because we have more

[54]OM = "Operating Model"; basically, the existing stock assessment used by CCSBT

Table 16: Final tally of fish genotyped successfully. For Indonesia, "year 2006" means the spawning season from November 2005 to April 2006; this is consistent with the definition of "SBT birthdays".

	2006	2007	2008	2009	2010	Total
Indonesia	214	1457	1526	1394	1164	5755
Port Lincoln	1523	1707	1448	1338	1432	7448

"existing" genotyped fish to compare each "new" fish with than when we started. Points to think about for the future include:

- necessity of building all the genetic QC into a formal database setting, and of independently checking the CK assessment code;
- possibility of using more of our stored-but-not-genotyped juvenile samples from 2006-2010 to improve the existing estimates, particularly in non-steady-state versions of the model;
- desirability of continuing data collection and genotyping in future years, at a level that can deliver a time-series of estimates and a cost-effective monitoring tool;
- long-term importance of obtaining direct data on spawning-ground residence times as a function of length from archival tags on big SBT[55], since residence time is crucial to both the current OM and the CK estimate.

19.3.1 Genetic results: finding POPs

The gruelling process of genotyping and error-checking has been described in previous documents. The breakdown of the 13203 fish successfully genotyped is shown in Table 16; a few hundred more were genotyped, but excluded in the end for assorted quality-control reasons described previously. Almost all the Port Lincoln juveniles were age 3 in the year of sampling (based on clear separation of modes in the length frequency), except for a few in 2006 that were age 4. Although the optimal scheme for a given budget would have been to genotype equal numbers of juveniles and adults (since this is likely to yield the greatest number of POPs for a fixed amount of genotyping effort), regulatory changes and delays with Indonesian export permits meant that we had to shift the balance somewhat towards juveniles.

The goal of the genotyping is to identify as many true Parent-Offspring Pairs as we can be sure of. The things to check, therefore, are whether many of our identified POPs are likely to have been "invented", and whether many other ones are likely to have been missed. Barring errors, a POP must have at least one allele in common at every locus, so if a pair is unrelated we will eventually be able to rule it out as a POP by finding a locus that does not share an allele,

[55] Many SBT have been archival-tagged already, but almost all were age 2-4 only.

Figure 21: All comparisons, broken down by #loci compared and #loci inconsistent with POPhood. Dot means zero, plusses mean too big to fit.

provided that we look at enough loci. We have scored 25 loci[56] overall, but not all loci are scored for every fish, so some pairwise comparisons involve many fewer loci. If too few loci are used in a comparison between unrelated fish, there is a substantial probability that all the loci will share an allele just by chance. We therefore need to do some filtering, to exclude comparisons that are too likely to give a false positive. Figure[57] 21 shows what happens if we *don't* do any filtering. True POPs— plus false POPs, which just happen by chance to share an allele at every locus compared— are in the leftmost column "F0", i.e. with zero loci compared that do not share an allele. False POPs are obvious in the top-left of the table, where very few loci are being compared.

Note that the Table includes a small proportion of (i) impossible and (ii) useless comparisons, where the adult was (i) caught in a year before the juvenile was born, or (ii) caught in the same year. Type (ii) comparisons are biologically possible, but it's not helpful to include same-year comparisons in abundance estimation, because in the year of its capture an adult cannot achieve its normal annual reproductive output. All such comparisons have been removed in subsequent summaries and results.

As in the Xmas 2011 update, in order to filter out false POPs we first compute in advance for each possible pair a False-Positive Probability (i.e. the probability that the two animals will share an allele at every locus compared, even if unrelated) based on which loci were scored successfully for *both* fish in the pair, and without looking at the actual genotypes that resulted. We then sort these FPP in ascending order, and find the cutoff such that the *total* FPP from all (sorted) pairs below the cutoff is below some pre-specified threshold T. Only those pairs whose FPP falls below the cutoff are subsequently checked for POPhood, the remainder being deemed too ambiguous. Note that not testing POPhood of an ambiguous pair does not cause any bias, because the FPP check is done *before* testing for POPhood, and is unrelated to whether the pair really is a POP or not. The threshold T is by definition equal to the total expected number of false POPs, so we choose it to be a small fraction of the number of true POPs, of which we have a shrewd idea of by this stage. For this report, we kept the threshold at the Xmas 2011 level of 0.35, which was 1% of the number of POPs then found, and is now less than 1% since we have subsequently genotyped more fish and found more POPs. Because false POPs lead to a proportional negative bias in abundance estimates, the upshot is that we have kept such bias to under 1%.

The resulting set of filtered comparisons is shown in Table 17. Given the cutoff used, at least 10 loci must be compared to get an acceptable FPP, and

[56] Plus another two that showed occasional anomalies, and were therefore omitted from routine pairwise comparisons, but were used in checking ambiguous possible-POPS.

[57] which would have been a table if the software had worked properly.

Table 17: Number of *usable* pairwise comparisons, by #loci and #excluding loci. Comparisons are *not usable* if the adult was caught in or before the year of juvenile birth, and/or the false-positive probability was too high (see text). Columns 8-21 omitted for brevity.

.	F0	F1	F2	F3	F4	F5	F6	F7	>	F22	F23	F24	F25	TOTAL
C11	.	.	.	.	1	4	5	21	>	.	.	.	.	84
C12	.	.	5	42	340	1345	4019	9114	>	.	.	.	.	57,000
C13	.	1	16	151	887	3420	9900	20482	>	.	.	.	.	143,000
C14	1	4	61	587	2876	11277	32947	70962	>	.	.	.	.	652,000
C15	.	3	42	375	1962	8411	27165	66386	>	.	.	.	.	923,000
C16	2	1	18	131	966	4716	17097	47526	>	.	.	.	.	1,170,000
C17	2	.	8	92	655	3674	14677	45482	>	.	.	.	.	1,942,000
C18	5	.	6	65	483	2699	12037	40524	>	.	.	.	.	3,063,000
C19	7	.	1	33	288	1728	7992	29511	>	.	.	.	.	4,158,000
C20	2	1	1	15	131	886	4630	18722	>	.	.	.	.	5,512,000
C21	14	.	1	5	62	481	2589	11387	>	.	.	.	.	7,197,000
C22	.	.	.	.	4	38	165	698	>	117	.	.	.	1,170,000
C23	4	.	.	.	2	20	143	754	>	2383	179	.	.	2,966,000
C24	2	.	.	.	4	22	90	558	>	17376	2799	214	.	5,097,000
C25	6	.	1	.	1	5	22	199	>	42419	10339	1607	139	4,123,000
SUM	45													38,180,182

only a few 11-locus pairs squeeze in; these occur where the 11 happened to be amongst the most powerful[58] of the 25 loci used for the table. On average, the loci used have about a 0.65 chance of *not* sharing an allele by chance, and the table shows very clearly how (near-)binomial probabilities work; from right to left, the numbers in the columns decline rapidly, except for the leftmost column where true POPs appear. In the bottom-left-hand-corner, the Table shows "clear blue water" between the best-matching unrelated pairs (i.e. with fewest loci that do not share an allele) and the true POPs. The separation is less obvious in the rows above say C16, but by looking at how fast the numbers in each row decline from right to left through the F4-F3-F2 columns, it is clear that very few unrelated pairs would have made it into the F0 column. And of course this is what the FPP calculations suggest: given the filtering rule, we would only expect 0.35 spurious POPs in the F0 column. Given that expectation, it's possible that one ($p = 0.25$) or maybe even two ($p = 0.05$) false POPs could have crept in, but it's very unlikely that false POPs make up an appreciable proportion of the total of 45.

Using a cutoff to exclude ambiguous comparisons does entail a bias-variance trade-off, because some true POPs may have been overlooked in the excluded comparisons, and any reduction in the overall number of POPs found will increase the uncertainty in our final estimates. However, given the threshold we used, it is only when the number of loci compared is 14 or less that substantial

[58] I.e. genetically more diverse, and being least likely to share an allele by chance

numbers of comparisons are excluded (from comparison of Figure 21 and Table 17), and overall only about 5% of comparisons are excluded. Thus we have managed to achieve less than a 1% bias while only incurring a $\sqrt{5} \approx 2\%$ increase in standard error compared to what we would have gotten from "perfect" genotyping (where every pairwise comparison is usable). This reflects very well on the tissue quality, the processsing, and the selection of powerful, reliable loci.

What about accidentally excluding true POPs? That can only happen if there is genotyping error[59]. If false-negatives are common, they should show up low down in the F1 column of Table 17, as near-POPs that apparently fail to match at one locus (failures at more than one locus are correspondingly less likely). None are to be seen. We have re-checked all the scores for pairs in Column F1 of Table 17, and they all look like genuine "lucky misses". During the checking of the F0 and F1 columns, the genotypes of a few individual fish-loci were adjusted in Table 17, but in all no more than one or two fish changed columns. Given these results, there is no reason to build in any safety-net rule for genotyping error, such as the "ignore single-locus exclusions and treat as POPs", which some genetic studies are forced to invoke in order to deal with genotyping error. That is a jolly good thing, because such rules would drastically increase the FPPs and would require us to be a lot more brutal with the cutoff to keep false-positives under control; we would have to exclude a much larger number of comparisons, which would reduce the eventual number of POPs found, which would reduce the precision of our abundance estimates and our ability to choose appropriate models for doing that estimation. The lessons for future close-kin studies are: do the genotyping really carefully, and don't skimp on the number of loci[60].

Note also that uncle-type relationships, while presumably at least as common as POPs, are not going to mess up the table. Between an uncle & nephew, only 50% of loci will share an allele by descent anyway, so with these loci the overall chance of sharing an allele is about $1/2 * 1 + 1/2 * (1 - 0.65) = 0.68$ (compared to about 0.35 for an unrelated pair), and the chance of getting say 20 loci all sharing an allele through chance is about 0.0004— so there would need to be about 2000 avuncular relationships to generate a single false POP.

Some small checks As an exercise, we can repeat Table 17 just comparing juveniles with themselves, where POPs are of course impossible[61]. When I did this pre-Xmas 2011 on the smaller sample, there were no entries in the F0 column after applying the filter, as one would hope. An uncooperative computer currently prevents me including the results for the complete sample, but from inspection there is now 1 spurious "POP" in the F0 column (row C18). Given an expected value of 0.35, it is not very surprising that one false POP sneaked through ($p = 0.3$ of at least one, as before), but the main point to make is that

[59] Or mutation, but with say ~50 POPs and ~20 loci each, and mutation rates thought to be about 10^{-4} per generation, mutation is unlikely to have happened amongst our POPs.

[60] Bearing in mind also that if more samples are collected, then the central splodge in the Tables will spill over further, towards the F0 column.

[61] Provided we exclude comparisons between a fish and itself.

Table 18: Comparison of adults to themselves. Note that all numbers are doubled, because A gets compared with B and B with A.

.	F0	F1	F2	F3	F4	F5	F6	F7	>
C11	.	.	.	.	.	.	.	.	>
C12	.	.	4	18	74	250	602	974	>
C13	.	2	14	182	932	3776	9994	19430	>
C14	.	.	64	590	2850	11244	31856	68376	>
C15	.	.	42	348	2230	9238	30900	73962	>
C16	.	4	34	188	1276	5620	20492	56768	>
C17	.	.	8	116	874	4132	16898	53060	>
C18	2	.	4	64	496	3118	13398	45348	>
C19	.	.	4	42	306	1616	7364	27712	>
C20	.	.	2	10	108	592	2998	12324	>
C21	.	.	.	6	38	312	1780	7820	>
C22	.	.	.	.	.	18	104	502	>
C23	.	.	.	.	4	22	134	616	>
C24	.	.	.	2	2	22	80	426	>
C25	.	.	.	.	.	4	30	154	>

there was only one.

We can also compare all adults with all other adults (Table 18). This time, POPs are actually possible, albeit likely rare because of the time required to reach maturity— see later discussion. There is one possible POP, and it is plausible biologically. The female "parent" was aged 24 when caught in 2007, and the female "offspring" was 177cm (not aged) when caught in 2009; this gives plenty of scope for the "parent" to have been mature when the offspring was born.

Summary of genetic results We ran about 40,000,000 pairwise comparisons to look for POPs. Some pairs had to be excluded because too few loci were used to reliably screen out unrelated pseudo-POPs. However, because of the number and quality of loci used, we were able to choose a cutoff that implies very little bias (i.e. unlikely to unearth false POPs) while incurring very little penalty in variance (i.e. using nearly all the comparisons). There was no evidence that genotyping error rates were high enough to cause true POPs to be overlooked. In all, we found 45 POPs in 38,000,000 comparisons.

19.3.2 Qualitative results

This section presents some simple summaries of the POPs found, and addresses a few points that were raised as possible concerns at earlier stages of the project.

Sex, age and size of parents vs general adults Of the 45 POPs, 20 were female and 25 male. All adults in POPs have now been aged; about 1/3 were

Figure 22: QQ plots of parental age vs adult age, by sex . Points right/below the line mean parents are bigger/older.

aged anyway under a long-term Australian/Indonesian ageing program, and the remainder were aged specifically for this project. Parents are typically somewhat older and bigger than average adults[62], especially females (Figure 22, top row; just age). This is not too surprising since parents are usually caught several years after giving birth, so have had extra time to grow (and to age, of course). Backdating to the offspring date-of-birth, subsequently-identified male parents were actually slightly *younger* when they become parents than typical males. For female parents at offspring-birth, there is no striking difference from typical females, but maybe a suggestion that the slope is below the equality-line; this would be expected if larger females have greater reproductive contribution even after allowing for selectivity.

Surprisingly, one of the female parents must have spawned successfully at just 6yo; she was caught 6yrs later at age 12. At capture, she wasn't particularly big for her age, just 159cm, so her back-projected size (see Model) at age 6 would have been about 135cm. That is remarkably small; out of about 10,000 fish measured on the on the spawning grounds during the 2000s, only 4 were smaller. Some of us suspected there had been a processing mixup and that the age & size data came from another fish; however, it proved possible to get enough DNA from the otolith to confirm that this particular fish was indeed the parent (though it is still just possible that her age was incorrectly recorded, and this is being checked). This one outlier has quite an effect on the QQ plot, but fortunately not very much on the abundance estimate.

The graphs shown here will eventually be done for length as well as age, but haven't been yet because of the extra complication of having to back-calculate lengths to birth date (see Model section).

Skip-spawning? With the expanded sample size, there is now some evidence of biennial spawning for younger fish. The test is to take each POP, and note how many years actually elapsed between juvenile birth and adult recapture, vs how many years *could* have elapsed given the POP was eventually found. For example, if the juvenile in a POP was born in 2007, then only comparisons with 2008/2009/2010 adults would be meaningful, so the probability of matching to a 2008 adult is roughly[63] equal to the proportion of adults checked in 2008 relative to those checked in 2008+2009+2010. Table 19 shows the results, divided by parental age (at juvenile birth); for younger parents, almost all observed gaps are even-numbered, but not for older parents. The pattern is similar if broken down by sex.

[62] I.e: average amongst those caught, which are not typical of the population because of selectivity.

[63] Calculations are approximate: e.g. the "expected" rows do not account for growth or mortality, but should reflect any even/odd pattern OK.

Table 19: Distribution of gap between JuBy and AdCapY, for young & old parents. Dot means zero.

Age	Gap->	1	2	3	4	5	6	7
6-12	Obs	1	6	.	2	.	4	.
	Exp	1.6	2.3	2.7	2.6	1.9	1.1	0.6
13-25	Obs	7	5	10	7	2	1	.
	Exp	4.3	6.8	7.4	6.2	4.5	1.8	1.0

Table 20: Gap distro by even or odd years

Age	Gap->	Even	Odd
6-12	Obs	12	1
	Exp	6.1	6.9
13-25	Obs	13	19
	Exp	14.9	17.1

Condensing to just even-or-odd makes it obvious:

Any errors in ageing would obscure patterns such as seen here. Although the sample size is not huge, the difference for younger adults is significant at 1%.

Skip-spawning is not a particular problem for this close-kin study because the study covers many years and the even/odd effect should largely wash out; jumping ahead, the general effect of smaller fish being less present is already allowed for, because average spawning-ground residence-time (including the probability of not being on the spawning grounds at all) gets estimated as a function of length and sex. However, in a perfect world probabilistic skip spawning would be allowed for in the "full estimation model". This is theoretically feasible, but is something for the future.

Timing in spawning season? Parents of GAB juveniles have the same distribution of capture date within season as do "average adults" (Figure 23). Thus there is no evidence of "temporal stock structure" in a way that might lead the abundance estimates to be biassed (eg we *might* have seen that parents of GAB juveniles always spawn early, and we might *not* have had equal coverage through the Indonesian fishing season). Breaking down by sex does not reveal anything either.

Are there any (half-)siblings among the POP juveniles? No. (This is a repeat of the Xmas 2011 section, and is still true with the larger sample size.)

Figure 23: QQ plot of day-of-year of capture of Parents (X) vs Adults-in-general (Y)

In other words, none of the POP adults match to more than one juvenile. That is a good thing, because if (half-)sibs are common among the *sampled* juveniles, then comparisons become non-independent. The adult abundance estimate still wouldn't be biassed (see below and previous explanations), but its variance would increase, potentially to the point of rendering the estimate useless. A preliminary check just among juveniles in 2010[64] indicated that (half-)sibs could not be *very* common (a critical decision point for the project), and the 7 POPs found in 2010 contained no sibs or half-sibs. Having found none in this much larger set of POPs, we can maybe conclude that (half-)sibs are rare enough *among our juvenile samples* for their effects on variance to be ignored; see section 4 for more details. This is not to say that (half-)sibs are at all rare among *all* 3-year-olds, but simply that our juvenile samples are a very small fraction of the total, and are well-enough-mixed to rarify the sib-pairs. As an academic exercise, it will at some point be interesting to re-run the juvenile-only sib check with the greatly expanded set of loci now available.

19.3.3 Abundance estimation

The cartoon version of close-kin abundance estimation for an SBT-like species goes like this. Each comparison between a juvenile and an adult has probability $2/N$ of yielding a POP, where N is adult population size, so if we do C comparisons we expect to get $2C/N$ POPs. Thus if we divide the *actual* number of POPs found into $2C$, we have an estimate of N. In this case, from Table 17 we have $C \approx 38,000,000$ and $\#POPs = 45$, so the "number you first thought of" is $2 \times 38,000,000/45 \approx 1,700,000$. However, this is *not* the right way to estimate abundance for real, because the cartoon version ignores time, in particular the lag between juvenile birth and adult capture.

As a preliminary to the real situation, consider a more sophisticated cartoon. If we had knife-edge maturity at known age A for SBT, and were able to age all adults, then we could for any juvenile ensure that all comparisons are made only to adults that are in the cohort-group of its potential parents, i.e. that were aged A or more at its birth. In that case, the cartoon calculations would in fact remain valid (provided the mortality rate of parents is the same as that of other adults of the same age).

The reality for SBT, though, is that maturity is not knife-edge (e.g. 6yo spawners exist, but are rare) and we do not have age data for most of the adults. We therefore have to make comparisons between each juvenile and *all* adults caught in each year after its birth, while recognizing that some of those adults could not have been its parents. The magnitude of the effect depends on the duration of the average gap between juvenile birth and adult capture (about 4 years in this study), and on the rate of turnover of the adult population. After say 4 years from the juvenile date-of-birth, then 4 years of mortality have taken their toll on the "parental cohort-group", counterbalanced by 5 years of incoming recruitment. The pool of real potential parents have been diluted

[64] See the "Steering committee update" from May 2010

by incoming recruitment of non-potential parents. If the adult population is roughly in equilibrium over that period, then mortality balances recruitment, so that if the survival rate is s then only s^4 of the comparisons are really "valid" after 4 years— a big difference, if say $s = 0.8$. The magnitude of the effect depends on the timespan of the study, and on the age of juveniles when caught because that extends the effective timespan. The effect is mitigated because fish in the original parent-pool also grow over the four years, and bigger fish are more likely to be caught per capita than smaller fish (see section 19.3.3). However, the mitigation cannot be complete; after say 15 years, almost all of the adults will be new recruits rather than the original parental pool, so almost none of the comparisons will be "valid", and this is why we see at most POP-pair in a purely adult-adult comparison (18). In all, the gap between juvenile birth and adult capture artificially inflates the number of comparisons; the nominal C is higher than the "ideal C", causing positive bias in the cartoon abundance estimate.

There is a second important effect which the cartoon overlooks, and which works in the opposite direction. Suppose adults could be sampled non-lethally in the spawning season, so that we could compare a juvenile against adults caught in its birth-year. Because selectivity is skewed towards bigger fish (see section 19.3.3 again), and bigger fish are more likely to be parents, the probability of a randomly-chosen adult from the *catch* being the parent is greater than $2/N$. This effect is related to selectivity, not to survival, and unlike the previous effect is conceptually independent of the duration of the study; however, the two effects are linked in practice, because growth means that selectivity changes between juvenile birth and typical adult recapture. But there is no reason to assume that either effect is small or that the two should cancel out, since one depends on study duration and the other does not..

One further issue arises from of the extended timespan of this study, which spans juvenile birth-years from 2002 and adult capture-years to 2010, as well as the initial age structure of the adults in 2002 which was determined by even earlier events. The 1990s and 2000s have been eventful decades for SBT, and it may be such that steady-state assumptions are simply not viable.

A proper close-kin abundance estimate for SBT therefore has to deal with survival, selectivity, fecundity, and growth, and perhaps with changes in abundance over time. The requisite data come from the length and age-at-length samples from Indonesia, plus fecundity studies explained below. While not "fishery-independent", length and age data are not subject to the same problems as CPUE or total catch. It also makes sense to split the analysis by sex: the cartoon applies equally well if applied to males and females separately, where the chance of a POP comparing to a male adult is $1/N_{\text{male}}$ not $2/\left(N_{\text{male}} + N_{\text{female}}\right)$, and C is split into $\left(C_{\text{male}}, C_{\text{female}}\right)$. Rather than labour away at some *ad hoc* redrawing of the cartoon to allow for turnover and growth, it seemed best to develop a "proper" adults-assessment model: one that has a coherent framework for all the effects, and that can directly estimate parameters using POP, length, and age data (no catch or CPUE). Such an adult-assessment model is also a good starting point for subsequent incorpo-

ration of the close-kin data— as well as a more nuanced approach to the adult age and length data— into the CCSBT OM. The next sections describe relevant aspects of SBT spawning biology and of the Indonesian length and age data, before launching directly into a description of the adult-assessment model.

Residence time, selectivity, and fecundity Indonesia is really no place for an adult SBT, an animal which is built superbly for much cooler temperate waters. Adults arrive on the spawning grounds fat, and leave thin. Of course, the longer they can stay on the grounds, the more chances to spawn they will have, so it seems reasonable to suppose that they will put up with Indonesian conditions for as long as their bodies let them. The key for disentangling the effects of fecundity, survival, and selectivity, is average **residence time** on the spawning grounds, as a function of length. A cursory glance at length distributions from Indonesia shows that few fish under 150cm, and none under 130cm, are caught on the spawning grounds, so there is obviously some link to length. As per the skip-spawning discussion, "average residence time" already factors in the probability that a fish won't be there at all in any given year. Our model specifically assumes that, given length and sex:

- Selectivity $\propto$ residence time
- Annual reproductive output $\propto$ residence time $\times$ daily reproductive output

Except as specifically noted later, we assume that length and sex are the driving influences behind the behaviour of adult SBT, rather than age.

Of course, there could be other "second-order" phenomena which slightly change the above relationships (e.g. different depth distributions by size, and thus different exposure to hooks; different egg *quality* with parental size; etc etc) but these seem likely to be small beer compared to the dominant effect of residence time. For the rest of this document, it may be helpful to think about selectivity and residence time as directly equivalent.

We have no direct data on residence time as a function of length (**yet**), so this needs to be inferred indirectly from data as described later. Independent data on residence time and depth distribution as a function of length, from archival tags placed on big fish, would be extremely useful: primarily in tightening up the existing model, and secondarily in assessing whether the effects that we hope are "second-order" really are.

Fecundity analyses: daily reproductive output The canonical reference for SBT (female) spawning biology and fecundity is a study from the early 2000s by Davis et al.[65]. In summary, female SBT while on the spawning grounds have an on-off cycle, consisting of several days of consecutive daily spawning (one spawning event per 24 hours), followed by several days of rest while more eggs are built up. This on-off cycle may be repeated several times. As soon as the

[65] T. Davis, J. Farley, M. Bravington, R. Andamari (2003): *Size at first maturity and recruitment into egg production of southern bluefin tuna* FRDC project 1999:106

final spawning cycle is complete, they leave. The mass of eggs released per daily spawning event can be estimated from the change in gonad weight between just-about-to-spawn and just-after-spawning fish; it scales as length$^{2.47}$. The average length of each part of the cycle (and thus the proportion of days on the spawning grounds when spawning actually occurs) can also be estimated as a function of body length using histological data, because the first day of a spawning sequence can be distinguished from the other days, and similarly for a resting sequence. However, the number of cycles per season is completely unknown, and is obviously set by the residence time.

To summarize, the factors involved in daily reproductive output are:

- reduction in gonad weight per spawning event
- duration of consecutive spawning day sequences
- duration of consecutive resting day sequences

A reasonable amount of data is available for all three of these, and the relationship to length can be estimated from fitting three GLMs. For now, we have treated the parameter estimates as exact in the rest of the assessment.

We have no comparable data for males.

Indonesian length, sex, and age data A goodly proportion of the Indonesian SBT catch is sampled as it passes through the main landing port of Benoa. Length (to the centimetre) and sex are always recorded, and nowadays otoliths are always extracted, although only a length-stratified subset (500 per year in the recent past) are read. Between 900 and 1700 animals were measured per year between 2002 and 2010. Thus the data can be seen as

1. Random samples of length and sex from the entire adult catch
2. Random samples of age, given length and sex.

Before considering how to analyse the POP results, it's instructive to consider what can be derived from analysing the length and age data alone. We work separately by sex, and assume that each animal follows a von Bertalanffy growth curve, with its own personal L_∞ but with k and t_0 fixed by sex. (Note that this relationship makes it simple to back-calculate length at any age for any particular fish, once we have its age and length in a given year.) If the L_∞'s are Normally distributed, then the distribution of length-at-age-and-sex is also Normal with constant CV. Under a steady-state assumption (constant numbers-at-age over time), the expected distribution of length and age in the adult catch of given sex can be reconstructed from these 4 growth parameters (3 for the vonB, and 1 for the CV), plus a survival rate and a residence-length relationship. If either the survival rate or the residence-length relationship is fixed, then the other can be estimated from the data. Figure 24 shows the fits obtained by inputting two different values for survival: 0.7 and 0.9. The fits for both sexes are good at $s = 0.7$, and show a residence-length relationship that

Figure 24: Female (top) and male (bottom) L, A, and L-at-A fits for different survivals (left & right)

looks plausible[66] For $s = 0.9$, the female fit (top row right) is poor, and this is because the residence-length curve has gone completely flat and has no room to manoeuvre; in fact, to get a good fit at $s = 0.9$ residence time would have to be a *decreasing* function of adult length, which is not plausible. For survival rates as low as 0.5 (not shown), the implied residence-length relationship time is again implausible, but in the other direction: exponentially related to length, doubling every 3cm. However, for a broad range of survivals in between those values, equally good fits and plausible residence-length relationships can be obtained.

Separating survival and selectivity using POPs The length/age data alone are evidently not able to disentangle survival from selectivity (residence time). However, the POPs can help. The typical gap between offspring birth and adult capture— assuming that the adult is in fact captured subsequently, i.e. that the pair is an identified POP— is related to survival. If survival rates are low, very few parents will survive to be caught say 8 years later (the maximum gap possible in this study), so most of the POPs that are found will be separated by just one or two years. In fact, the same effects are at work here as in the start of section 19.3.3, so growth and residence time need to be properly accounted for too, but the intuitive basis should be clear. The close-kin data thus has two vital roles: the *number* of POPs (given the number of comparison) essentially sets the scaling of absolute abundance, and the distribution of time-gaps *within* the POPs essentially determines survival. Formally, this is all handled automatically within the adult-assessment model described next.

19.3.4 Adult-assessment model structure

The model is age-based (and sex-based), in that it keeps track of numbers-by-age-and-sex; each year, each fish either gets one year older or dies. However, most phenomena are driven by length, which is assumed to have a fixed distribution at age as in section 19.3.3. A plus-group is used for ages above some threshold (25 in all runs to date) by which a fish has just about stopped growing, and a minimum age for possible spawning also needs to be set (currently 6), at which "recruitment" to the adult population is deemed to begin. There is also a plus-group for length (200cm) and, unusually for stock assessments, a sort of "minus-group" as well, currently set to 150cm. Experience with fitting just to age and length data showed that trying to extend the fit to the small proportion of adults below 150cm gave poor results, in that this small "tail" started to "wag the dog" and distort the fit elsewhere. The focus of the CK

[66] We have assumed that the relationship reaches an asymptote at large lengths, but haven't tried checking this assumption.

analysis is spawners, which are mostly 160cm and up, so it is more important to get a good fit there than to squeeze a last drop of misinformation out of very small adults. However, it is necessary to somehow keep track of the small spawning contribution of fish under the minus-group, and accordingly there is some tedious book-keeping code in the model.

The model has to be able to compute expected proportions at length, and at age given length, by sex in any given year. Once these are available, a binomial likelihood can be computed from the observed length and age data. The sample sizes involved are very large, and to avoid the risk of the model tying itself into knots trying to match unimportant nuances of such a large dataset, it may be advisable to downweight the length and age data, i.e. to reduce the nominal sample size. *Ad hoc* patches like this are not a good long-term solution, but they are a useful sensitivity check for robustness of conclusions. All these steps related to length and age data are fairly standard in stock assessments (apart from the minus-group mentioned earlier).

The model also needs to be able to compute, for every usable pairwise comparison, the probability that the adult will be a parent of the juvenile, given the dates and the adult size (adult age not generally being known); and for those pairs that are POPs (for which adult age is known), it also needs to compute the probability of the adult having the age it did, given its length and the fact of its POPhood. To give a brief flavour of how this is done: the POP probability is computed by first working out the total female (or male, depending on the adult's sex) reproductive contribution in the juvenile's birth-year, using the numbers-at-age in that year, the length-age relationship, the length-residence relationship, and the daily reproductive output-length relationship. Then the length of the adult in question at capture is used to back-calculate its probabilistic length distribution in the birth-year, bearing in mind all the ages it could really be (since its age is generally not known) via Bayes' theorem; and this length distribution is turned into its expected back-calculated reproductive output that year using the fecundity-length relationship(s). The chance of the adult being the juvenile's parent is then equal to

$$\frac{\text{its own reprod output that year}}{\text{total reprod output from fish of that sex that year}}$$

Given the probability, the outcome (POP or not) is just a Binomial random variable with $n = 1$, and contributes to the total log-likelihood in the obvious way. This calculation automatically (after a great deal of programming) takes care of the survival, growth, and residence-length effects mentioned in section 19.3.3.

To actually compute a likelihood, it is necessary to specify somehow various terms:

- numbers-at-age in 2002, and for incoming recruitment (at age 6) in 2003-2010;
- survival rate in each year and age;

- residence/length relationship;
- growth parameters;
- relation between daily reprod output and length *for males*.

The total number of potential parameters is colossal because of the numbers-at-age and survival terms, so of course one needs to specify them parsimoniously given the limited amount of data available. This is done using formulas (sensu R) for each of bullet-point term, describing what covariates are allowed to influence it, and perhaps what functional form that influence might take. For example, we might choose to make survival constant over age and time, except for the plus-group[67]. We might also make assumptions of constant "recruitment" in the 2000s; and/or that numbers-at-age prior to 2000 were in equilibrium with survival; and/or that von Bertalanffy k is the same for both sexes; and/or that the slope of the residence/length relationship (but not its midpoint) is the same by sex; etc. A few examples are given in the Results section.

The final term— male daily reproductive output as a function of length— can *in principle* be estimated provided we are willing to assume that survival rates for males are the same as for females. Without that assumption, there is nothing to anchor the selectivity/survival/fecundity triangle for males. For females, we do not need to estimate this term because we have direct data from the fecundity studies.

The likelihood itself is coded in Pascal, with derivatives computed by an automatic differentiation toolbox a bit like ADMB. The overall data-handling and fitting is done in R, calling the `nlminb()` optimizer to do the fitting. Some care was needed to avoid numerical problems in calculating the log-likelihood, and because of limited time there are still starting-value problems so that some model parametrizations can't get started. However, once a starting value has been obtained, no convergence problems were encountered, at least for the fairly parsimonious specifications (say 15 parameters) that have been tried to date.

19.3.5 Results

It will be apparent from the previous section that an enormous number of different *versions* of the adult-assessment model could be investigated. A full investigation is far beyond the scope of this project (though there is a separate follow-on project to consider how to integrate the CK data with the existing SBT Operating Model i.e. assessment). In this section, we just present a few results from fairly simple versions of the model.

Basic investigations suggest that:

- Mean L_∞ is appreciably larger for males than females. The evidence for any difference in k or t_0 is not overwhelming, but making these two

[67] In SBT as with other top-predators, it must be the case that natural mortality rate increases for old animals, since simple maths shows that the sea would otherwise just fill up with decrepit tuna.

sex-linked as well does not seem to overparametrise the model. CV of length-at-age appears to be the same for both sexes.

- Residence time appears to be lower for males of a given length than for females, so we do need a sex-specific intercept in this term. However, there is not enough data to estimate any sex difference in the *slope* of the relationship. Also, introducing extra flexibility in model form beyond the logistic (asymptotic) can give nonsensical predictions for very large fish. A good choice seems to be `~sex+length`.

- Allowing survival to vary over time makes the predictions of abundance quite unstable, especially in recent years. As with any mark-recapture model, it is difficult to estimate survival from really short periods at liberty, and if for example we decide to let survival change in 2006, then (i) we only have 4 years of data to estimate survival from, and (ii) we have only a portion of the POPs to work from, from a sample size which is not enormous to begin with. Since fishing mortality on *adult* SBT is thought to be low (a thought that is certainly consistent with the results here), the main contribution must be natural mortality, so there is no obvious reason to allow adult survival to change over time.

- With male daily reproductive output, the signal in the data is weak, and direct estimation of the exponent in a power-law relationship tends to give ludicrous estimates. I have assumed instead that male daily output is directly proportional to length (i.e. exponent of 1). There is no good reason for that particular choice, but fortunately the abundance and survival estimates seem not to be much affected by assumptions about male daily output in practice, even though it could matter in theory.

- Based on just one comparison: changing the nominal annual length/age sample size from 300 to 900 did not much affect the abundance estimates much (i.e. by a few percent).

Having got these basic issues out of the way, the remaining questions concern how to set up initial numbers-at-age and incoming recruitments. In a full steady-state model, the age distribution in year 1 (actually 2002AD in our setup) is determined by the survival rate, and the incoming recruitments thereafter are equal to the numbers at recruitment age in year 1. The estimated annual survival for this case is 0.73, and the abundance estimate[68] (age 10+, both sexes) is 1,750,000, or a corresponding SSB of 135,000T based on the length-weight relationship in the OM— note though that the interpretation and appropriate definition of "SSB" might be changed substantially by the CK results. The CK abundance estimate is higher than all the various OM estimates from various scenario, though the survival estimate is similar to the base-case OM estimate; see Discussion.

[68] Since this is steady-state, there is in theory only one estimate. Because of minor imperfections in the steady-state formulation concerned with the plus-group, there is a 1% trend over the period.

Figure 25: Steady-state diagnostics: length

Figure 26: Steady-state diagnostics: sex-ratio

Some diagnostic plots for the steady-state model are shown in Figure 25-28. These pertain to the length and age and sex data only, since the POP data are probably too sparse for diagnostics. The fits to age-at-length are excellent (Figure 27); the fits to length frequency, shown for a sample of years only, are mostly not too bad despite the steady-state assumption (Figure 25) except for 2002 where the data seem completely different from other years; and the residence time plots (Figure 28, which is not actually a diagnostic) seem plausible, showing a strong increase over the main length classes of adults, with males taking longer to appear. However, there is a problem with the fits to sex ratio by length class (Figure 26): in the biggest length classes lower down the graph, where males tend to predominate thanks to their bigger asymptotic size, there is a strong decrease in proportion of females over the 2000s. This decrease is seen overall too (in the black dots), but is not apparent in the smaller lengths, where there is a rise followed by a dip. This difference in trend across length classes suggests that methodological changes in how sex is assessed are unlikely to be the cause. A change in asymptotic size, whereby the incoming old females don't grow as big as the ones dying off, would be one explanation, but there could be many others: e.g. if somehow there were strong sex differences in cumulative fishing mortality on sub-adults during the 1990s that affected the relative proportion of females reaching adulthood.

There is a limit to how non-steady-state the model can be made without giving implausible results. For example, estimating log-linear trends in either initial numbers-at-age (beyond what is suggested by survival and an equilibrium assumption) or incoming recruitments generates implausibly big trends in adult abundance during the 2000s— big increases or big decreases— though the midpoint tends to be fairly stable midway between 1,000,000 and 2,000,000 10+ fish. Of course, it is not very plausible to enforce exponential increases and decreases over decadal time scales.

A modest non-steady-state extension is to allow a change in incoming recruitment in 2002 (or any other year, but that one is particularly easy to set up in the existing structure) to a new, fixed, level. This gives a numerically promising improvement in fit— about 32 units of "log-likelihood" for 2 extra parameters— but the weighting for the length- and age- data is still rather arbitrary at this stage. However, there is only a modest improvement to the sex-ratio fits (Figure 29); there is a trend now, but only in the smaller length classes, which is the opposite of where it's needed. The abundance estimates

Figure 27: Steady-state diagnostics: age at length

Figure 28: Steady-state estimates: residence time

Figure 29: Unsteady-state diagnostics: sex ratio

are slightly lower than for the steady-state model: 1,650,000 in 2002 decreasing to 1,411,000 in 2010. The survival estimate is little changed, at 0.75.

Some further work is needed to calculate CVs for these models (both computationally, and in terms of sorting out the effective sample size of the length and age data, and of course in assessing whether the fit is good enough to begin with), but MVB's *guess* from inspecting the Hessian is that the CV on average abundance might be around 25%. This makes reasonable sense, since the main driver of the abundance estimate is the number of POPs, and Poisson variability around an expected value of 45 amounts to a CV of around 15%; since survival and residence also affect the abundance estimate, and are estimated imprecisely themselves, the overall CV on abundance should be somewhat higher than 15%.

The abundance estimates from the adult-assessment— even though it is not yet perfect, as the sex-ratio fits show— are quite close to the "number first thought" of from the cartoon version. The implication is that the effects of survival and growth/residence must largely cancel out. If correct, this is just an accident of the duration of the study so far and the decision to concentrate on 3-year-old juveniles. With a much longer study, the growth/residence effects would reach an asymptote, and most comparisons would involve adults too young to have been the parent. For this study, at a survival rate of just 0.75, the impact of (say) three years of mortality is quite large; but if Figure 28 is to be believed, the increase in residence time (selectivity) of a typical parent over that timespan will also be large, because the residence curve rises steeply in the length range of many of the parents. Some informal exploration of the magnitude of the effects would be a useful common-sense check.

The fact that the abundance estimates are quite similar in the two models shown here, should **in no way** be taken to mean that these two estimates bound the range obtainable from CK. Model exploration has been quite limited, and there is significant misfit to the sex-ratio trends, although it's not obvious how much this might lead to bias in the abundance estimates.

Fecundity and bodyweight Using the existing daily-fecundity-vs-bodyweight and bodyweight-vs-bodylength data, and the residence-time-vs-bodylength and length-at-age estimates from the CK assessment, we can estimate the (relative) annual reproductive output of females[69] as a function of age, and of course also the mean bodyweight at age. We can plot the two against each other, and see how closely the CK results compare with the OM *assumption* (not estimate) that annual reproductive output is directly proportional to bodyweight for age

[69]With males, this has not been possible because of the lack of direct daily-fecundity data, and the very low precision when estimating it just from the CK data.

Figure 30: Bodyweight vs annual fecundity for females, estimated by CK data. Green line is OM assumption.

10+. The answer is in Figure 30: not very closely. According to CK, the average 12-year-old has about 33% the reproductive output of a 20-year-old, versus about 75% according to OM.

19.3.6 Conclusions

Our tentative adult abundance estimates are a lot higher than the OM estimates— at least 3 times the point estimate from the "base case" scenario, and on the edge of the upper confidence interval of the most optimistic scenario. This document is not the place for a lengthy discussion of the OM, but it should be emphasized that there is very little reliable data in the OM with which to estimate absolute adult abundance (although other quantities such as relative depletion can be estimated more reliably), which of course is in itself is a big part of the reason for undertaking this study. Also, as noted in section 19.3.5, the different notions of effective fecundity in the two models make direct comparisons tricky. So getting an adult abundance estimate that is very different to the OM's is by no means an indication that the CK estimate is wrong, nor that the main conclusions of the OM (which aren't related . Nevertheless, it is important to ask the obvious question: how wrong could these CK estimates be? There are a limited number of issues to consider, given that we are not asking about small changes here— the point is to think of phenomena that could make a huge reduction to the estimate, of the order of 50%.

Is the number of POPs about right? The genetic results strongly suggest that there are few if any false negatives or false positives, given the filtering we have used. It is impossible to "create" large numbers of false positive POPs accidentally by muddling samples, so so the only way that the true number of POPs could be very different from what we found is if there are many false negatives (i.e. POPs that we missed) from some kind of wholesale mixup in the genetics. But there has been a painfully extensive checking process, largely described in previous reports, and there are systems in place to guard against plate mixups and the like, based partly on catching a couple of problem cases early on. Given the way we have handled the genetics, large-scale processing mistakes seem very unlikely. Programming mistakes in the POP-matching code are always a possibility, of course— all the programming for this project has been done by one person (MVB) and has not been checked by anyone else.

How precise is the estimate? If the estimation procedure is valid, then CV on the abundance estimate could be around 25%— reasonably good by fisheries standards, but still woeful by e.g. medical standards. A 25% CV means there is about a 15% chance of being 25% over the truth, which would

put the CK estimate at the upper end of the most optimistic OM results— though again direct comparability is suspect, because of the different age-specific fecundities estimated or assumed in CK or OM. CVs do reflect the amount of noise in the data arising from samples being finite, but do not encompass model uncertainty (which has not been fully explored), nor the implications of the different "definitions" of SSB arising from fecundity-length relationships (which is more something for the OM to consider than CK).

Is the abundance estimate about right, given the number of POPs? If the number of POPs is about right, and the adult sampling was simultaneous with juvenile birth and random, then the cartoon estimate can't go wrong— each juvenile really does have exactly two parents. Most of the other potential problems with close-kin— stock structure, or massive proportions of sibs/halfsibs— don't apply to SBT. So the only other source of possible error is in the adult-assessment model. There are perhaps three points to consider:

1. Has the model correctly estimated the magnitude of the two "cartoon adjustments"? Which are: dilution of the parental-pool by incoming recruitments between juvenile birth and adult capture, versus bias towards sampling parents rather than non-parents because parents tend to be bigger. In practice, these two adjustments seem to have largely neutralized each other, which is possible but not at all inevitable— it must really be just an accidental consequence of the duration of the project. It's not obvious whether back-of-the-envelope estimates from raw data are possible for these terms, but if not then at least some post-model calculations of their sizes would be useful.

2. The entire CK assessment, and the way in which the cartoon adjustments are implicitly calculated, rests on the assumption that selectivity is primarily driven by residence time— the longer you're there, the more likely you are to be caught, all else being equal. The link between residence time and annual female reproductive output rests on the same assumption. It is hard to see how this assumption could actually be wrong, but the caution might be in the phrase "all else being equal". If there are other really major length-based effects on selectivity or on reproductive output (aside from female daily fecundity, for which we at least have some data), then bias could perhaps arise.

3. It must also be allowed that programming mistakes are more likely in the CK assessment than in POP-checking; it is a fairly complicated model that is quite young. Since the CK assessment model will need to be re-coded to marry it with the rest of the OM, any such mistakes should get picked up over the next few months.

Even if there do turn out to be errors in these estimates, they seem more likely to be programming errors, and therefore fixable, rather than being intrinsic problems with the data or its interpretation. The CK data fundamentally do

seem to be extremely useful for SBT: they are bearing out their promise. There is obvious scope for continuing to collect and genotype in future, both to build up the time series and also (thanks to the retrospective qualities of close-kin) to increase the number of POPs found from our already-genotyped 2000s juveniles. The way this might fit into SBT management, and the links with other monitoring possibilities, is far more than can be explored in this study, but the potential value of further CK genotyping is clear.

Finally, we draw attention to the key role of residence time on the spawning grounds— or, to be accurate, how the average residence time depends on size— in getting to an actual abundance estimate, and a selectivity estimate, and an appropriate definition of SSB. Although there is just about enough data in the POPs and the age/length samples to infer the residence/size link indirectly, it would be immensely useful to have direct estimates from a few fish of different sizes, since this could both ground-truth the model and give a basis for estimating further length-dependent effects on selectivity, if required. The data could best come from archival tags (as opposed to e.g. pop-up tags) because they can record over several years, and are not vulnerable to short-term tagging shock. The low fishing mortality on adults means that quite a few tags would be needed to get recaptures, and that we might have to wait a while to get the tags back, but the number of returned tags needed would not have to be at large (even single figures) to give a very useful check on, and input to, close-kin based abundance estimates in future. Such tagging ought not be a very expensive exercise in terms of the value of the fisheries, or indeed the cost of this close-kin project.

20 Appendix 7: Working papers to the CCSBT Extended Scientific Committee

CCSBT-OMMP/1307/info/01
CCSBT-ESC/0709/18

A method for estimating the absolute spawning stock size of SBT, using close-kin genetics

Mark Bravington[1]
Peter Grewe

Prepared for the CCSBT 8th Meeting of the Stock Assessment Group and the 12th Meeting of the Extended Scientific Committee
4-8 September, and 10-14 September 2007, Hobart, Australia

[1] CSIRO CMIS, GPO Box 1538, Hobart, Tasmania 7001, Australia
Email: Mark.Bravington@csiro.au

Contents

List of Tables

Abstract

We describe a method for estimating the absolute spawning stock size of SBT, based on genetic identification of parent-offspring matches in samples from Indonesia and the GAB. The method is related to mark-recapture, and provides an estimate of true adult numbers (not the unrelated genetic concept of "effective population size"). No catch or CPUE data is used, so the estimate is not subject to the biasses and interpretational problems associated with recent SBT assessments. We explain the statistical basis of the method, comment on its robustness, and describe progress with data collection and future plans.

1 Introduction

The CCSBT and its Scientific Committee cannot currently do a stock assessment of SBT. The recent discovery of serious flaws in the historical catch and CPUE data have undermined the main data sources for the assessment, and estimates of absolute SSB— which were always much less precise than relative depletion estimates— are now completely uncertain. Even if a new version of the historical CPUE index can be reconstructed, it seems unlikely (given the inherent difficulties of partial spatial coverage) that uncertainty in the series will genuinely be low enough to allow a precise assessment. It may be possible to develop a new assessment based on different future data sources (e.g. a different CPUE-style relative abundance index), but there will not be enough new data anytime soon to allow this. As to other existing data sources, tagging gives very little information about spawning-age fish, and even for sub-adults it is handicapped by various reporting rate issues in all the major fleets. The juvenile aerial survey in the GAB is of limited value on its own, as it is only a relative index and covers only the GAB-visiting part of the juvenile population, and anyway can say nothing direct about adult abundance[1].

This paper presents a completely different way to estimate, within a couple of years and using fishery-independent data, one key management quantity: the absolute abundance of adult SBT. The basis is counting parent-offspring pairs in samples of juveniles and adults (i.e. fish old enough to spawn). Intuitively, for a sample of fixed size, there will be fewer parent-offspring pairs if the population size is higher. This idea can be developed into a formal statistical estimate of absolute adult abundance (but not juvenile abundance), using ideas from mark-recapture. DNA fingerprinting can be used to actually identify the parent-offspring pairs. The approach was pioneered by Skaug, 2001, who applied the method to a small sample of North Atlantic minke whales, and by Nielsen et al., 2001, who estimated the abundance of male humpback whales in the West Indies. Skaug's dataset exhibits a number of difficult features which don't apply to SBT (see **Discussion**), most notably the inability to distinguish adults and juveniles; these features forced Skaug to make extra assumptions during analysis, which are not required for SBT. Nielsen *et al.* were able to directly compare genetic samples from males with calves, given also a genetic sample from the calf's mother, and were able to use a simpler analysis with fewer assumptions. Although their sample size was very limited, their results were consistent with independently-obtained estimates of male abundance.

[1]However, a significant trend over time in a juvenile abundance estimate is an important *indirect* indicator of spawner abundance.

We are embarking on a project to estimate the absolute abundance of SBT adults, based on identification of parent-offspring pairs between adults in the spawning grounds off Indonesia and juveniles in the GAB. The proposal was already being developed before overcatch was reported, but the urgency of the work is now much greater. Our method is more similar to Nielsen *et al.*'s than Skaug's, in that we are able to distinguish adults from juveniles, but there are a number of important differences, described in this paper. The project includes four years' sampling from the Indonesian spawning-ground fishery (2005/6-2008/9), and at least three years' sampling from the GAB juvenile fishery (2006-2008), with results expected by CCSBT 2009. Our target is to have at least 7000 fish genotyped (about 50/50 adults/juveniles); we have based this on approximate sample-size calculations, aiming for ∼70 parent-offspring pairs to get an overall[2] CV of ∼12%. We have already done some careful preliminary genetics to develop suitable loci, so that we will be able to establish parent-offspring relationships with high confidence in a cost-effective way.

Although underpinned by genetics, our approach has nothing in common with "effective population size"; it is a direct estimate of recent[3] adult abundance, and is based on mark/recapture principles rather than population genetics theory. Genetics is only used as a "mark" (in the juvenile) which can be "recaptured" (in the parent), and the only theory required is that of biparental inheritance.

For clarity, we will first describe the principles as they would apply to a single-year study where all adults have an equal probability of being sampled: how to estimate abundance, and how to estimate the CV. We then describe the current status of our SBT work and our future plans. The setup is a little more complicated than the basic case described in **Basic methods**, so we go on to outline how the basic idea can be modified (or whether it needs to be) to deal with a number of potential wrinkles for the case of SBT:

- sex-specific effects
- multi-year sampling
- age- or size-dependent catchability and fecundity
- additional reproductive variability
- population substructure

The final section includes a summary, some comments on how the immediate results (which pertain only to adults) might be used in management (which pertains to a much wider entire age range), and some thoughts about how the approach might be extended in future as part of a long-term monitoring and assessment framework.

[2]The real CV will likely be different, for several reasons explained later. Mainly, the sample size calculation depends on the abundance, which is of course the thing that is very uncertain before doing the project.

[3]I.e. with a 3-year lag.

2 Basic methods

DNA tests are commonly used to test parenthood. Colloquially, for a typical "gene" with several variants in the population and two copies of the gene in each animal, a parent and its offspring must have at least one identical variant, whereas unrelated individuals might have totally different variants. Formally, a parent and its offspring must have at least one matching allele at every diploid locus. If a locus has a large number of different alleles, there is a low probability that two unrelated animals will have a matching allele at that locus just by chance. If we examine a large number of loci on each animal, the probability that two unrelated animals will have a matching allele at *every* locus is therefore extremely low. Hence, we can in principle completely rule out "false positives", i.e. apparent parent/offspring pairs that are really unrelated. False negatives are almost impossible if scoring is reliable, so from now on we assume that the genetic evidence is an exact indicator of a parental relationship.

Now suppose you have a sample of m_A randomly-selected adults[4] and that, one year later, you collect a sample of m_J one-year-old juveniles. Pick one of the juveniles and one of the adults, and genotype both of them at enough loci to rule out any possibility of false-positives. What is the probability of a "hit"— i.e. that the chosen adult is actually a parent of the juvenile? Since the juvenile must have had two parents, the probability that the chosen adult is one of those two is $2/N_A$, where N_A (or just N) is the number of adults alive when the juveniles were spawned. Now repeat the comparison for the same juvenile and all the other adults. The expected number of hits between that juvenile and the entire set of m_A adults is $2m_A/N$. Now repeat this for all the juveniles: the expected total number of hits, $\mathbb{E}[H]$, is $2m_J m_A/N$. Thus, if h is the actual number of hits, we can form an approximately unbiassed estimate[5] of N in the obvious way (formally, by using the "method of moments") via:

$$\hat{N} = 2m_J m_A/h$$

Note that the method cannot tell us anything about the total abundance of juveniles. The logic doesn't work in reverse: although we know that each juvenile must have had two parents, we don't know how many juveniles on average each parent would have had. In mark-recapture terms, each juvenile "marks" exactly two adults which might subsequently be recaptured, allowing us to estimate the number of adults. Looked at the other way round, though, each adult "marks" an unknown number of juveniles— which makes it impossible to use mark-recapture analysis directly to estimate the abundance of juveniles[6].

There are two crucial points to emphasize. First, the derivation of $\hat{N}$ does require that the adults are randomly sampled, but does *not* require that the juveniles are randomly sampled; in particular, the juvenile samples do not have to be mutually independent. Of course, the juveniles must be selected

[4] Collected just after the spawning season, to avoid removing the very parents that we seek.

[5] As with most maximum-likelihood estimates, the estimate is only *asymptotically* unbiassed, i.e. the bias disappears if the expected number of recaptures is large enough. For $h > \tilde{}10$, the relative bias is about $1/h$, i.e. about 1.5% for the SBT project given the "target" of 70 for h.

[6] Skaug's method estimates adult and juvenile abundance together, and uses the number of half-sibling etc. matches as well as parent-offspring pairs. However, the method is less direct and requires extra assumptions which would not make sense for SBT.

independently of the adults— the method breaks down if applied to mother-calf pairs, for example.

Second, the derivation of $\hat{N}$ does *not* require that all "adults" make an equal reproductive contribution. The key point is actually the random selection of adults. In fact, the "adult" population might be defined as "that set of animals which have equal probability of appearing in our m_A-sample". The trickiest part of applying the method to SBT, is correcting for unequal sampling probabilities among the "adults"; see **Wrinkles**.

2.1 Basic CV & sample size calculations

To get an idea of the uncertainty in $\hat{N}$, one further assumption is needed: that the numbers of hits from different juveniles are independent (see **Wrinkles**). Then some algebra (see **Appendix**) shows that

$$CV\left(\hat{N}\right) \approx \frac{\sqrt{2}}{m}\sqrt{N} \tag{1}$$

where m is the combined sample size (for optimality, split equally between adults and juveniles). Given some *a priori* notion of N, we can use (1) to set the sample size; e.g. a 10% CV requires about $15\sqrt{N}$ samples. For SBT, using a guesstimate from a now-obsolete assessment of $N \approx 350,000$ (the number of fish $\geqslant 160cm$, the lower limit of maturity), a target CV of 12% implies a sample size of 7000, with about 70 hits being expected and about 1% of the adults being sampled. We stress that this is only a sample-size calculation, and the achieved CV will be different for a number of reasons; see **Discussion**.

The remarkable thing about (1) is that it is (inversely) *linear* in sample size. By contrast, in the great majority of statistical settings, CV depends (inversely) on the *square root* of the sample size, meaning that diminishing returns usually set in as more data are collected. With close-kin abundance estimation, though, there is a quadratic gain in efficiency[7], basically because each new (juvenile) sample is compared against *all* existing (adult) samples, hence generating far more than one "data point".

3 Status of SBT work

3.1 Sampling

Our adult sampling program uses the infrastructure of the existing Indonesian catch sampling programme in Benoa, Bali. Samples for genotyping are taken throughout the fishing (spawning) season from all possible SBT 165cm and up. This size limit was chosen based on maturity data, to safely

[7] Unless the sampling fraction becomes "large", or the period of sampling becomes so long that a high proportion of parents of "early" juveniles have died.

encompass all fish big enough to have been parents two years previously, when the youngest juveniles in the corresponding sample were spawned. Mouth tissue is collected by a trained sampler and deep-frozen for shipment to Australia. All genetically-sampled fish are lengthed and sexed (by checking for residual female gonads; see Farley et al., 2007) as part of the regular catch sampling programme, and a portion of the genotyped fish form part of the otolith-collection set and so will be of known age. To make sure we are sampling only spawners, and in the absence of precise information on fishing location, we have excluded all SBT from trips with a high proportion of sub-adult fish, since in the last couple of years, boats from some fishing companies have been fishing further south, outside the SBT spawning ground. Coverage of the spawning grounds and spawning season is good; the fleet that lands into Benoa covers the main part of the SBT spawning grounds, although a much smaller unsampled catch of SBT is taken further west from the Cilicap fleet, in an area of apparently lower SBT spawning density (Proctor et al., 2003; Far Seas Fisheries Research Laboratory, 1985).

To date we have collected two adult samples from Indonesia ($m = 220$ from 2005/2006, and $m_A = 1200$ from 2006/2007), and we plan two more years with similar sample sizes to this year. That number of adults is about the maximum possible given the logistics of sampling.

For juveniles, we collected a sample of 4000 juveniles from Port Lincoln in 2006, and are arranging a similar number for this year (about 800 collected so far) and 2008. The 2006 sample contained mostly 3-year-old fish ($\sim$90%), with about 4% 2-year-olds and 6% 4-year-olds. Only a subset of fish will be genotyped to begin with, in case that gives enough precision overall. Genotyping of juveniles will be restricted to fish within length bands that allow an unambiguous determination of age, to ensure that we can accurately track the birth-year (actually this does not exclude very many fish).

None of the samples have been genotyped yet. Although our current plans are to genotype a total of around 7000 fish, if it does turn out that we get many fewer hits than expected (i.e. much bigger N) then we will have a large number of spare juvenile samples than can be genotyped to improve precision. This is quite possible if the true spawning stock size turns out to be much higher than we assumed in our sample size calculations.

3.2 Genetics

Genotyping costs money, and given that the sample size will be quite large, it is important to minimize the number of loci that need to be tested. The key is to do some careful preparatory work to pick the best set of loci, and to use a two-pass approach: first genotype every sample at some set L_1 of loci and then, whenever a possible matching pair is found, check the pair by genotyping a further set L_2 of loci. L_1 needs careful design; it must include enough highly-informative loci to rule out a high proportion of samples on the first pass, but not so many that expense is pushed needlessly high. An outline of the calculations required is given in the Appendix; the number of loci required depends only on the sample size, not on N. Because only a small proportion of animals will be tested in the second pass, design of the second pass is less critical.

The costs of genotyping go up stair-wise with number of loci (e.g. 6 loci might be almost as cheap as 5, but 7 might cost almost twice as much), so it is worth spending considerable efforts to develop really informative loci and to organize them efficiently for mass genotyping. The ideal loci for close-

kin studies are more variable than is useful for traditional population genetics, for which an excessive number of alleles actually reduces statistical power.

With SBT, we have done a considerable amount of preliminary work to identify new, powerful, and reliable loci (microsatellite library enrichment, locus discovery, primer design, amplification optimization, trialling on a sample of 16 fish, allele frequency calculations). A basic set of 20 tetra-nucleotide[8] loci have been selected and, based on their allele frequencies in the sample, using any 12 of the more powerful of these as L_1 will eliminate well over 90% of unrelated fish on the first pass. In other words, even if some of the best loci fail for some unexpected reason such as strongly preferential amplification of short alleles, we still have plenty of other candidates. A subset of the remaining loci can be used L_2.

4 Wrinkles

The base case described above is very simple, but does not apply directly to our SBT project. There are a number of wrinkles which require attention. To keep the descriptions as clear as possible, each wrinkle is discussd independently of the other wrinkles, usually in terms of adjustments to estimators. In practice, though, the wrinkles interact, and it will be necessary to move to a fully parametric likelihood-based framework for estimation. That will complicate the statistical programming, but the comments below about estimability and precision made below will not change.

4.1 Sex

There is a sex-bias in spawning-ground samples of SBT, and a different bias in the whole adult population (Farley et al., 2007). Nevertheless, each juvenile must have had one male and one female parent. Since the sex of the adult SBT sample is known, the simplest way to deal with any sex bias is just to make independent estimates $\hat{N}_m$ and $\hat{N}_f$ of the adult male and adult female abundances, using the male and female adult samples respectively. This separation-by-adult-sex should be assumed throughout the rest of this paper, but is not mentioned explicitly. There is a small effect on the CV— see Appendix.

4.2 Sampling delays and multi-year sampling

The abundance estimate described above is retrospective: N is the number of adults that were alive in the year when the juveniles were spawned, rather than when the juveniles or adults were sampled. This remains true for the multi-year SBT project, but there are extra complications of modelling, sampling, and interpretation that arise from the multi-year nature of the project. There are really three aspects. The first is that, for a given cohort of juveniles, the potential parents will be sampled

[8]Loci with tetra-nucleotide repeat sequences are easier to score reliably than the di-nucleotide loci which (being easier to find) are more commonly encountered in population genetics.

across several years, rather than in one year. Linked to this is the second aspect: there will generally be a delay of several years before the potential parents of a given juvenile cohort are sampled, during which some of the parents will die. The third aspect, which is quite separate, is that there are multiple cohorts of juveniles.

With respect to delayed adult sampling, it is obvious that a given juvenile could only have been spawned by fish that were big enough to be adult in its birth-year. Therefore, depending on the birth-year of the juvenile and the sampling-year of the adults, it is necessary to restrict the set of potential parents that are checked for hits, to make sure they were all mature in the birth-year; this can be done by using an age or length cut-off projected forwards to cover the delay. In other words, it is important to do all checks against the *same* population of potential parents of a given juvenile across the years of adult sampling; in mark-recapture terminology, the population must remain "closed". This ensures that the abundance estimate pertains to the original size of the adult population in the juvenile's birth-year.

With respect to possible mortality between juvenile "marking" and parental "recapture", it turns out that the date of adult "recapture" does not affect the probability of that adult being the parent of a particular juvenile (under reasonable assumptions, and after addressing some of the other wrinkles, in particular **Multi-year-breeding cycles**; see Appendix for justification). Hence sampling delays do not lead to bias. For a given cohort of juveniles, the basic model could in fact be extended to the multi-year adult-sampling case by simply aggregating the potential-parent samples across years, using the year-dependent size cut-off.

With respect to the multiple cohorts of juveniles, it is necessary to allow for possible changes in adult abundance over the different birth-years of the juvenile cohorts. In the context of the basic model, this could be done easily by constructing independent estimates of adult abundance in each birth-year[9], and then averaging (to reduce noise). In practice, the interaction with age- or size-dependent catchability will necessitate a more complicated likelihood-based multi-year model for SBT.

Overall, the multi-year and delayed-sampling issues entail a small number of extra parameters, but should have little impact on CV. It is worth noting that, because the number of hits is proportional to the square of the sample size, and because each new year of samples gets cross-matched to earlier years as well as to itself, most of the hits will not be found until the final year of this 3-year study.

4.3 Multi–year breeding cycles

It is theoretically possible that SBT— even big ones— have a multi-year breeding cycle and do not turn up to spawn every year. Suppose there was a two-year cycle: then a one-year project that sampled only one cohort of juveniles would either coincide with an "off" year or an "on" year for the parents of that cohort, and there would either be no matches or twice as many as the total abundance suggests— and there would be no way to detect either phenomenon. However, in a two-year program, or a one-year program with two age-classes of juveniles, the overall number of matches

[9]The estimates for different birth-years are effectively independent because there is a negligible probability of any adult matching multiple juveniles from different cohorts.

comes out right and the bias disappears (see Appendix). In practice, the effect of a regular breeding cycle would, if important enough to matter, be obvious in the samples, by comparing sampling-year of identified parents against birth-year of corresponding offspring. If this revealed a 2-year cycle, say, then the method would need an adjustment to differentiate between

$$\mathbb{P}\,[\text{random adult on spawning ground is my parent}|\text{odd number of years since my birth}]$$

and the even-year equivalent. This only adds one extra parameter (or $p-1$ in the case of a p-year cycle), and so would have limited effect on precision. In the SBT project, adult samples will be collected over at least 3 years and will cover at least 4 cohorts of juveniles, so the project should be able to cope with breeding cycles of 4 years or less.

Irregular breeding cycles (for example, breeding on average only one year in two, but at random rather than alternately) don't affect the basic method— if adults are present at random on the spawning grounds, then the chance of any one being your parent is still $2/\!/N$. However, if there is an age-related effect on the probability, some adjustments are required, as described next.

4.4 Age-dependent sampling probability

In our design for SBT, adults are sampled on the spawning grounds. Sampling probabilities will therefore *not* be equal across ages. For example, suppose there is a gradual maturity ogive rather than knife-edge maturity; then, in any given year, the proportion of fish at "age of 50% maturity" that are available for sampling in the spawning grounds will be lower than the proportion available within the fully-mature age classes. This necessitates an adjustment to the basic method, using information on the spawning biology of SBT. The description below is mostly in terms of age, for simplicity, but in fact the main driver is length (or, equivalently, body-weight), and the actual statistical models used on the data will probably need to be length-based.

SBT are multiple spawners, remaining on the spawning ground for days or weeks, with a daily spawning cycle possibly punctuated by rest periods (*Farley and Davis, 1998*); also, some mature fish may simply not visit the spawning grounds in some years. The more time a fish spends on the spawning ground, the more eggs it will produce, *and* the more likely it is to be caught. Bigger/older fish of a given sex seem to spend more time on the grounds (Davis et al., 2003; and consistent with the apparent over-representation of older fish in spawning ground samples found in Farley et al., 2007), and certainly produce more eggs per day[10] (Farley and Davis, 1998).

Although the histological studies above have shown how daily egg production relates to size (and age), there is no independent data on residence time[11] as a function of age. Nevertheless, the quantities required to provide an unbiassed estimate of N can be estimated from three sources: the age profile

[10] Males don't produce eggs. The method used for unbiassed estimation of male abundance is statistically similar to that proposed for females, but differs in biological detail. Some extra collection of male gonads will be necessary, as fewer males have been studied than females.

[11] Residence time has two components: the probability of coming to the spawning grounds in a particular year, times the average residence when actually on the grounds. There is no way to separate these two components, but it is only their product that is important.

of sampled adults on the spawning grounds, the age profile of identified parents, and histological data on daily egg (or sperm) production. In the hypothetical example considered in the Appendix (with single-year sampling and no sampling delays), it is assumed that residence time, daily egg production, and abundance are exponential functions of age with coefficients r, g, and z respectively and only g known *a priori*. The Appendix shows that r and z can then be estimated, and that the basic abundance estimate needs to be adjusted by a factor $\frac{(r+z)(g+r+z)}{z(g+2r+z)}$. Note that if $r = 0$, i.e. that residence times and therefore sampling probabilities are equal for all adults, then the adjustment is 1 whatever the value of g: that is, variations in adult fecundity do not bias the estimate unless correlated with sampling probability.

In reality, residence time and daily egg production will be asymptotic rather than exponential functions of age, so some non-linear estimation will be necessary and more than one parameter will be involved. Also, a length- rather than age-based model will probably be required. A joint likelihood model for all the data will be necessary, and the estimation of extra parameters (i.e. z and r in the hypothetical example) will have some impact on the CV. However, a greater impact on the CV will come from the fact that the expected number of hits depends on g and r. Because of r, our project will sample more heavily from the more fecund fish, so the number of hits (and thus the precision) might actually be better than the "target" even if our guesstimate of N happens to be about right. Further, it might be useful or even preferable to construct an age-weighted version of N, e.g. for direct comparison with spawning ground catches; such an $\hat{N}$ would have yet again a different precision. All these aspects can only be quantified once the data is available.

In principle, we could take advantage of age-specific fecundity and catchability by changing the adult sampling design to concentrate even further on older/bigger fish, which would increase the number of hits per genotype. In practice, though, there are too few adult fish available to make this worthwhile.

4.5 Random reproductive variability

As shown in previous section, systematic reproductive variability between adults does not bias the basic $\hat{N}$ unless correlated with adult sampling probability. Nevertheless, in some fish populations, a small number of mating events can, by chance, contribute a very high proportion of the surviving juveniles. Would this random reproductive variability have implications for a close-kin abundance estimate? The short answer is that there is almost[12] no bias. As noted at the end of **Basic methods**, there is no requirement for adults to make equal reproductive contributions, as long as they are sampled with equal probability (or that any unequal sampling is allowed for, as just described). If you are a juvenile, then the chance that a randomly-chosen adult is your parent is still $2/N$ whether you have no siblings or 1,000,000 siblings.

However, random reproductive variability could affect the precision of $\hat{N}$. If two juveniles happen to be sibs or half-sibs, then the results of comparing the second juvenile against the adult sample are not independent of the results of comparing the first one. Hence, if there are many sibs or half-sibs

[12]Actually there is a small amount of bias because $\hat{N}$ itself is slightly biassed for finite-sized samples, as noted in **Basic methods**. Reproductive variability decreases the effective sample size, so worsens the sub-asymptotic bias—but this should be very small for SBT with 70 hits expected.

in the juvenile sample, then the effective sample size will be substantially reduced. The good news is that this kind of event can easily be detected, by examining the amount of allele-sharing within each juvenile cohort. It is less easy to be sure about individual half-sib identifications than about individual parent-offspring identifications, because the genetic overlap is much less, but nevertheless an overall statistical excess is easy to check. And unless the reproductive variability is actually manifest in the juvenile *sample* (rather than the cohort as a whole), there is no impact on precision.

Seriously high reproductive variability is mostly documented for landlocked species with small populations, such as bass and salmon. With SBT, the large population, prolonged spawning season, pelagic spawning, and multiple mating behaviour all make the phenomenon *a priori* less likely. Further, the juvenile sample is a tiny fraction of the juvenile population. To the extent that the sample is "random" (although this is not a requirement of the method), the incidence of siblings in the juvenile sample should therefore be far lower than the incidence in the cohort (following the same argument used to derive the CV of $\hat{N}$, with $\mathbb{E}[H_{JJ}] \propto m_J^2/N_J$) . Nevertheless, the sample of ~1500 juveniles per year is actually drawn from a much smaller number of schools, and samples from the same school could contain siblings or half-siblings from the same spawnings. Only time, and data, will tell; but the point is that any effect strong enough to reduce precision should be apparent and estimable from the data.

4.6 Closely-related individuals

Accidental hits between related individuals that are not parent-offspring pairs will not be a problem. First, it is only non-parental relationships between juveniles and adults that would matter; within-group comparisons are not used. The number of full-siblings between juveniles and adults will be miniscule, because any such pair would have to result from two matings between exactly the same individuals at least 8 years apart. The number of half-sibs and grandparent-grandchild pairs could be of the same order of magnitude as the number of parent-offspring pair, but such distant relatives only share 1/4 of their alleles, and need not share any alleles at any given locus; hence the chance of say a pair of half-sibs having at least one allele in common in all 18 hypervariable loci is very low. Since there should not be vastly more close-relative pairs than parent-offspring pairs in the adult-juvenile comparisons, and the probability of the former mimicking the latter is very low, false-positives from close relatives are most unlikely to be a problem.

4.7 Population structure

So far, it has been assumed that SBT form a single population with complete interbreeding. Although no previous study has found evidence of population structure, conventional population genetics applied to large populations is a notoriously blunt tool for that task. It turns out (see **Appendix**) that the basic method is unbiassed even when there is population sub-structure, providing that sampling is proportional to abundance across either the sub-populations of adults, or the sub-populations of juveniles. In our SBT project, juvenile samples come only from the GAB, so if there are substantial numbers of non-GAB juveniles out there somewhere, then juvenile sampling will obviously not be proportional. However, adult samples should cover the spawning season and spawning area (, al-

though not necessarily in strict proportion to adult SBT density. Hence, the basic estimator would exhibit population-structure bias if and only if three conditions all apply:

1. adults exhibit fidelity across years to particular parts of the spawning season and/or spwaning grounds;
2. the timing or location of spawning affects a juvenile's chances of going to the GAB (rather than going elsewhere or dying young);
3. sampling coverage of the spawning grounds (in time and space) is substantially uneven, and correlated with the fidelity patterns in (1). (In other words, if adults showed timing-fidelity but not spatial-fidelity, whereas coverage was even across the spawning season but not across the spawning grounds, then the uneven spatial coverage would not matter.)

There is no direct information on condition 1. With respect to condition 2, much the greatest part of SBT spawning occurs within the North Australian Basin (Far Seas Fisheries Research Laboratory, 1985), and particularly towards the east and south of the basin beyond the Australian shelf, where the Indonesian through-flows in summer would tend to push the larvae together into the Leeuwin current. These conditions seem unlikely to induce a strong location-of-spawning effect on most juvenile's subsequent propensity to go to the GAB[13], although a timing-of-spawning effect is possible. With respect to condition 3, the Benoa-based operations that we are sampling coincide well with this main spawning area (Proctor et al., 2003, Figure 4.3.1; note that the fishing range has expanded southwards since then, as per Proctor et al., 2006). Approximate timing-of-effort information could be probably be obtained from the sampling program; spatial information has proved harder to get, but the data obviously do exist somewhere at the company level, and some insights may be obtainable through, for example, the observer program (Sadiyah et al., 2007) or the Fishery High School program (Basson et al., 2007).

Fortunately, there is enough information in the project data to check the first two conditions. If the seasonal/spatial distribution of identified parents of GAB juveniles is substantially different to the seasonal/spatial distribution of all adult samples, then that is a clear signal that the first two conditions do apply. Such evidence of population structure[14] would be of major qualitative importance to management, regardless of its impact on quantitative results.

If and only if the first two conditions do apply, then the third could be checked using timing (and perhaps location) information on Indonesian samples. And if all three conditions do apply, then it should be possible to adjust for the uneven adult sampling probabilities, again using sampling coverage information. That is very much a bridge to be crossed only if we come to it; but because the sampling coverage is at least fairly complete[15] even if not necessarily balanced, we would in principle be able to develop a correction if required.

[13] A small proportion of larvae are found to the north of the NAB and west of it. Different oceanographic conditions apply there, and those larvae could well end up somewhere different as juveniles. However, at least until 1981, this proportion was small.

[14] "Population structure" is probably the wrong phrase, because the behaviour does not have to be heritable; adult spawning preference need not be related to earlier juvenile GABness, even if offspring's GABness is driven by adult spawning preference.

[15] Again: over the great majority of the spawning area.

5 Discussion

5.1 Comparisons with previous close-kin work

It is worth taking a moment to compare our project with Skaug, 2001 and Nielsen et al., 2001. Both studies had very limited sample sizes, since the data were collected for other purposes, and consequently low expected number of hits and low precision. Both studies also had use very limited set of loci, originally developed for different purposes and at a time when genotyping was much more expensive than nowadays. Consequently, the "false-positive" probabilities were so high that both studies had to rely on probabilistic evidence of a match, complicating the statistics. In Skaug's case, it was both necessary (in order to get a bigger number of hits) and unavoidable (because of the equivocal genetic evidence from using a small number of loci) to allow for other close relationships, in particular half-sibs and grandparent-grandchild. This entails further assumptions about reproductive variability and equilibrium age distributions and abundance, which (as Skaug notes) was a major problem for the close-kin approach in that particular example. Nielsen *et al.* had an easier time— and required far fewer assumptions— because the only relationship of interest was parent-offspring, and because knowledge of the mother's genotype makes it much easier to exclude an unrelated father, even with limited loci. They did note that there could be a complication arising from the adult male sampling probability being correlated with reproductive output— noisy males that make a conspicuous display are easier to find, both for females and for biopsy crews. Nielsen *et al.* lacked the data to address that issue, which is essentially the same one that we face with size-specific catchability on the spawning grounds; in our case, though, the histological information about daily egg production should be enough to compensate for unequal sampling probabilities. Overall, in our study as compared to the earlier studies, the much larger expected number of hits (and hence potential precision) should mitigate the need to estimate a few extra parameters.

Close-kin abundance estimation does not seem to have been much used since those two papers. In a marine context, most fish species are simply too abundant to have made the method cost-effective, although this may change as genotyping costs continue to drop. For many fish, the possibility of undetected population substructure would also be a deterrent, as this can bias estimates if not allowed for (see earlier).For cetaceans, where the method was first developed, there are usually alternative methods of estimating abundance, either through line-transect surveys (as with Skaug's minke whales) or through mark-recapture using photo-ID and/or "genetic tagging" from biopsies. SBT is in many ways the ideal species for a close-kin abundance estimate: the spawning population is not that large, there is not thought to be serious population substructure, the species is valuable enough to make mass genotyping affordable, and sampling can be arranged fairly easily on top of existing programmes. Most importantly, though, there is a *need*, because there is not (and never really was) a reliable alternative estimate of absolute spawner abundance.

5.2 Comparisons with conventional mark-recapture

In principle, a conventional mark-recapture program might be a competitive way to quickly estimate adult abundance (although tagging large numbers of huge spawning-age SBT is not that easy). The

potential downside of genetic mark-recapture (either of individuals, or of closely-related animals such as in our study) is that only a limited set of animals are checked to see whether they are recaptures, so the sample size can be greatly reduced in theory compared to an ideal tag-recapture program in which every tag recovered was reported. However, genetic mark-recapture does have one great advantage over conventional mark-recapture, in that there is no confounding between reporting rate and recapture rate. With conventional tags, a non-recovery could be either due to non-reporting or non-recapture, but with genetic tags, non-reporting (i.e. forgetting to send in a sample— it is impossible to tell whether the sample is a recapture or not) simply reduces the sample size without leading to bias. This is an important point which underlines the "fishery-independent" nature of the data.

5.3 Implications of the results for management

The most immediate result for management will of course the adult abundance estimate; given the great uncertainty about the current status of SBT, the obvious first thing to do with it is to compare it against current catches of adults, as a bottom-line check on adult mortality rates. In addition, based on the discussion of age-specific factors, there should be a direct estimate of recent Z among the adults. (This is a Z in the same sense that the slope of a catch-curve is; it combines total mortality with any trend in recruitment.) Combining the N and Z estimates, and using current catch information on older sub-adults (assumed correct in future, but not necessarily in the past), it should be possible to make some inference about likely current mortality rates on older sub-adults, too. There are a variety of ways that such *ad hoc* calculations might be done, and might be extended back to younger fish. Obviously, such calculations do not constitute a full assessment (see next section), but they do allow a sanity check in an environment where there is both great uncertainty and great concern about the status of SBT.

In the slightly longer term, estimates of N (and recent Z) can play a key role in conditioning whatever Operating Model gets developed for testing Management Procedures. (This is how we had originally envisaged the results being used, before the issue of overcatch was raised.) When there are as many dimensions of uncertainty as SBT now has, it is an exceedingly hard task to capture the range of plausible scenarios; there are very many "parameter" combinations are consistent with very uninformative data. Cutting down the "scenario space" is essential part of making MP-testing feasible.

The precision of $\hat{N}$ obviously has implications for how the results feed into management. As per **Wrinkles**, there are a number of model-related reasons why the CV will differ from the basic-case sample size calculations, but probably the dominant factor is that the real abundance might be considerably higher or lower than we have assumed. If the abundance is lower, then the CV will be improved. But even if the actual abundance turned out to be 10× higher than in our sample size calculation, the basic-method CV would still be around 40%, and by tripling the juvenile sample using the "reserve pool", this could be brought down under 25%. This is pretty good precision compared to many fisheries measures, and by virtue of the quadratic efficiency of the method, adding extra years of data will bring down the CV rapidly. This leads on to the question of whether a continuation of close-kin sampling, and an adaptation of the method, could play a larger role in assessment and

management of SBT.

5.4 Scope for future close-kin work

Given the early stage of our close-kin project, it is well beyond the scope of this paper to speculate too far into the future, but a little reflection suggests that close-kin abundance estimation for SBT might be even more useful as an ongoing assessment tool than as a one-off exercise. For one thing, the quadratic gain in efficiency with sample size means that CVs should drop rapidly with the accumulation of more years of data[16]. Second, with far more hits, it may become possible to track individual cohorts through the adult population in terms of their changing proportional contribution to annual juvenile production. Most importantly, though, there becomes a possibility of extending the model to cover earlier age classes. There are several ways this might be done, but conceptually at least, one easy way might be to use the close-kin estimates of absolute adult abundance by cohort (by looking at changes in $\hat{N}$ over time and against adult catches) to set the end-points of a VPA-style back-calculation using catches from pre-adult ages. In turn, this might allow assessment of age-specific selectivity without depending on a relative abundance index. As with the various wrinkles discussed in this paper, the best way to do all this in practice would probably be through an "integrated" likelihood-based assessment model. In that context, it is important to note that the the close-kin estimate provides not just an absolute estimate of abundance, but also an absolute estimate of precision (unlike, say, a CPUE index).

More speculatively, a sampling programme for close-kin genetics could also open up some quite different opportunities. For example, the CCSBT's program of conventional tagging of juveniles could be supplemented by a juvenile biopsy program as part of genetic mark-recapture. The several thousand juvenile samples that could be genotyped annually from Port Lincoln as part of a close-kin adult-abundance estimation, would be a no-cost source of recaptures; one would test the samples against the original biopsy "tags" for individual matches (not close-kin). Additional expense would be incurred in genotyping the "tagged" fish, and obviously the recapture rate would be much lower than is *potentially* achievable with conventional tags, because only a small fraction of the penned fish are genotyped in the close-kin project. On the plus side, though, there would be no problem with unknown/variable reporting rates, nor any need for tag-seeding. And if sample collection from the longline fleets could also be arranged, this would open up even more possibilities for genetic mark-recapture.

5.5 Conclusions

In this paper, we have described how a close-kin abundance estimate for adult SBT can be obtained, and outlined our progress to date. There are a number of important details of sampling and analysis that need to be taken care of, in order to avoid issues that could lead to biassed or hopelessly imprecise estimates. As far as we can foresee, though, just about all these potential issues can be addressed

[16] Eventually this levels out, of course, when there are too few of the original adults left alive to score hits against. By then, though, the focus is no longer on a single estimate of adult abundance, but rather on a time series.

using data that will be available either now or during the project. The sample sizes required seem unlikely to be exorbitant, and the genetic feasibility has been established.

Any proposed new method will, of course, have caveats attached to it until the results are in. However, it is important to bear in mind the problems of all the other data sources on SBT. In comparison, the close-kin estimate (albeit of a limited part of the age range) rests on rather few assumptions. As well as providing fairly swiftly a one-off estimate to calibrate operating models and serve as a bottom-line comparison for catch rates, a close-kin sampling program might even constitute an important part of future management procedures for SBT.

References

[Basson et al., 2007] Basson, M., Andamari, R., Sadiyah, L., and Proctor, C. (2007). An update on the use of the Indonesian Fishery High School dataset to obtain a standardised CPUE series for SBT on the spawning grounds. Stock Assessment Group paper CCSBT-ESC/0709/15, Commission for the Conservation of Southern Bluefin Tuna.

[Davis et al., 2003] Davis, T., Farley, J., Bravington, M., and Andamari, R. (2003). Size at first maturity and recruitment into egg production of southern bluefin tuna. Technical Report 1999/106, Federal Research and Development Council (Australia).

[Far Seas Fisheries Research Laboratory, 1985] Far Seas Fisheries Research Laboratory (1985). *Average distribution of larvae of oceanic species of scombroid fishes, 1956-1981.* S series 12. Far Seas Fisheries Research Laboratory, Shimizu 424, Japan.

[Farley and Davis, 1998] Farley, J. and Davis, T. (1998). Reproductive dynamics of southern bluefin tuna, *Thunnus maccoyii*. *Fishery Bulletin*, 96:223–236.

[Farley et al., 2007] Farley, J., Davis, T., Gunn, J., Clear, N., and Preece, A. (2007). Demographic patterns of southern bluefin tuna, *Thunnus maccoyii*, as inferred from direct age data. *Fisheries Research*, 83:151–161.

[Nielsen et al., 2001] Nielsen, R., Mattila, D. K., Clapham, P. J., and Palsboll, P. J. (2001). Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics*, 157:1673–1682.

[Proctor et al., 2006] Proctor, C., Andamari, R., Retnowati, D., Herrera, M., Poisson, F., Fujiwara, S., and Davis, T. (2006). The catch of SBT by the Indonesian longline fishery operating out of Benoa, Bali in 2005. Stock Assessment Group paper CCSBT-ESC/0609/10, Commission for the Conservation of Southern Bluefin Tuna.

[Proctor et al., 2003] Proctor, C., Merta, I., Sondita, M., Wahju, R., Davis, T., Gunn, J., and Andamari, R. (2003). A review of Indonesia's Indian Ocean tuna fisheries. Australian Centre for International Agricultural Research: Country Status Report ACIAR Project FIS/2001/079, CSIRO Marine Research.

[Sadiyah et al., 2007] Sadiyah, L., Andamari, R., Prisantoso, B. I., Retnowati, D., Proctor, C., and Davis, T. (2007). Trial observer program for IndonesiaŠs tuna longline fishery in the Indian Ocean. Stock Assessment Group paper CCSBT-ESC/0709/??, Commission for the Conservation of Southern Bluefin Tuna.

[Skaug, 2001] Skaug, H. (September 2001). Allele-sharing methods for estimation of population size. *Biometrics*, 57:750–756.

6 Appendix: mathematical justifications

6.0.1 CV calculations

"Baseline" case The expected number of hits is $2m_J m_A/N$ and the maximum possible is $2m_J$ if both parents of every juvenile occur in the sample of adults. The true distribution of H is hypergeometric, assuming all comparisons are independent, but since presumably $N \gg m_A$ (i.e. only a small fraction of the adults are sampled) a Poisson approximation will be fine. Hence the CV of H will be $\sqrt{N/(2m_J m_A)}$, and since $CV(X) \approx SE(\log X)$ and $\log \hat{N} = const - \log H$, this will also apply to $\hat{N}$. To minimize this CV for a fixed total sample size $m = m_J + m_A$, it is most efficient to set $m_J = m_A$, giving

$$CV\left(\hat{N}\right) = \frac{\sqrt{2N}}{m}$$

When there is an unequal sex ratio in the adults, but sampling is in proportion to the sex ratio, then the CV is unaffected. When the sex ratio in the sample is different to the sex ratio in the adult population (which then has to be estimated), the CV worsens slightly. For SBT, spawning ground catches are about 2:1 female:male, and assuming a true 1:1 sex ratio, the CV would increase by a factor of about 1.05. Calculations for these are as follows:

Population proportion known and equal to sample proportion Let p be the proportion of females in the adult population, so that the abundance of adult females is Np and the sample size of adult females is $m_A p$. Then H_f, the number of hits to adult females, is approximately Poisson distributed with mean $m_J(m_A p)/(Np) = m_J m_A/N$. The variance of $\hat{N}|h_f$ is approximately $N^3/(m_J m_A)$, by the Delta-method. An independent estimate can be constructed from the males; note that H_m has exactly the same expectation, so $\mathbb{V}\left[\hat{N}|h_m\right] = \mathbb{V}\left[\hat{N}|h_f\right]$. Since we have two independent estimates with the same variance, we can average them to obtain an overall estimate with

$$\mathbb{V}\left[\hat{N}|h_f, q, p = q\right] \approx \frac{N^3}{m_J m_A}\frac{1}{2} \tag{2}$$

The corresponding CV is the same as for the baseline case.

Population proportion known but not equal to sample proportion Suppose $q \neq p$ is the proportion of females in the adult sample. Similar reasoning shows that

$$\begin{aligned} \mathbb{V}\left[\hat{N}|h_f, p, q\right] &\approx \frac{N^3}{m_J m_A}\frac{p}{q} \\ \mathbb{V}\left[\hat{N}|h_m, p, q\right] &\approx \frac{N^3}{m_J m_A}\frac{1-p}{1-q} \end{aligned}$$

and the inverse-variance-weighted combination has

$$\mathbb{V}\left[\hat{N}|h_m, h_f, p, q\right] \approx \frac{N^3}{m_J m_A}\frac{p\,(1-p)}{q\,(1-p)+p\,(1-q)} \tag{3}$$

For the cases $p = q$ and $p = 0.5$ (any q), the right-hand fraction is still 0.5.

Population proportion unknown but not equal to sample proportion We have $\mathbb{E}\left[H_f\right] = m_J m_A q/\left(Np\right)$. If p is unknown, the best we can do is estimate the product Np from the females, and $N\left(1-p\right)$ from the males, and then add the two to estimate N. Using the Poisson approximation together with the Delta-method, we get

$$\mathbb{V}\left[\hat{N}|h_m, h_f, q\right] \approx \frac{N^3}{m_J m_A}\left(\frac{p^3}{q} + \frac{\left(1-p\right)^3}{1-q}\right) \tag{4}$$

The extra uncertainty compared to equation (3) arises from having to estimate p. For SBT, $q \approx 2/3$, and for $p = 0.5$ (from catch data on adults outside the spawning season) q, the term in brackets is $\sim$0.56, compared with 0.5 used in our sample size calculations. The sex bias will therefore inflate the real CV by a factor of $\sqrt{0.56/0.5} \approx 1.05$.

6.0.2 Sampling delays and probability of being a parent

Suppose we have one juvenile born in year 0 and a sample of adults taken in year y, all of whom were mature when the juvenile was born. What is the probability that one particular adult from the sample will be a parent of the juvenile? Let ℓ be the length of this adult in year 0 (probably inferred from its length in year y), let P be the event that this adult was a parent of that juvenile in year 0, and let Y denote the event that this adult occurred in the adult sample y years after that juvenile's birth. Note that y implies that the adult first survived for y years, and was then captured in the adult sample.

$$\begin{aligned}\mathbb{P}[P|\ell,Y] &= \frac{\mathbb{P}[Y|\ell,P]\,\mathbb{P}[P|\ell]}{\mathbb{P}[Y|\ell]} \\ &= \mathbb{P}[P|\ell]\,\frac{\mathbb{P}[\text{survived to } y|\ell,P]\,\mathbb{P}[\text{sampled in } y|\text{survived to } y,\ell,P]}{\mathbb{P}[\text{survived to } y|\ell]\,\mathbb{P}[\text{sampled in } y|\text{survived to } y,\ell]}\end{aligned}$$

Assume that:

1. survival probability (conditional on length) is independent of whether the adult was a parent of anything in year 0, and
2. sampling probability (conditional on length) in year y is independent of parental status in year 0.

Then the P-conditionals on the top of the fraction are irrelevant, and we have

$$\mathbb{P}[P|\ell,y] = \mathbb{P}[P|\ell]$$

so that year-of-sampling is irrelevant (conditional on length)[17].

The first assumption is probably reasonable for adult SBT, since their annual survival is pretty high— the differential mortality (fishing+natural) associated with spawning or not spawning in a particular year is likely low. The second assumption could be violated if, for example, there was a 2-year breeding cycle; the probability of being in the sample would alternate between low and high according as y was odd or even. This possibility is addressed next.

6.0.3 Multi-year breeding cycles

Clearly, if adults have a two-year breeding cycle, bias will occur if we sample in only one year on the spawning ground and the juvenile ground; either there will be too many matches compared to a no-cycle population of the same adult size, or too few.

To show that bias disappears if we sample a population with a two-year breeding cycle over two years, suppose we sample m_J juveniles and m_A adults overall, split evenly over each of the two years, with a single age-class of juveniles sampled in each year. Crucially, we must also assume that only adults who are going to spawn in a given year will turn up in the adult sample for that year— this is true for SBT. where all fish that have been checked histologically on the spawning grounds have been in spawning condition. When both years' data are analysed, the Year 1 juveniles will match only against Year 1 adults, and the Year 2 juveniles will match only against Year 2 adults. Suppose there are N_1 adults in the "odd-year spawning group" and N_2 adults in the "even-year spawning group"; then

[17] If we did not condition on length, then the fact that the fish was a parent in year 0 provides some information on the size of the fish, and thus on its subsequent survival probability, so the conclusion would no longer be valid.

	Number sampled				Expected number of hits	
Year	Jc1	Jc2	Ac1	Ac2	Same cycle	Other cycle
1	$m_J/2$	0	$m_A/2$	0	$2\left(m_J/2\right)\left(m_A/2\right)/N_1 = m_A m_J/\left(2N_1\right)$	0
2	0	$m_J/2$	0	$m_A/2$	$2\left(m_J/2\right)\left(m_A/2\right)/N_2 = m_A m_J/\left(2N_2\right)$	0
Total	m_J		m_A		$m_A m_J\left(1/N_1 + 1/N_2\right)/2$	

Table 1: Wthin- and between-year hits, given a two-year breeding cycle

For simplicity, suppose that $N_1 \approx N_2$; this is reasonable for a long-lived species where the odd & even breeding groups are made up of multiple cohorts, as with SBT. The total expected number of matches becomes $m_A m_J/N_1 = 2m_A m_J/N$ where the total adult population is $N = 2N_1$. Compare this with sampling m_J & m_A from a freely-interbreeding population of adult size N in a single year; the expected number of hits is again $2m_A m_J/N$. Hence there is little bias as long as we sample both years.

Actually, there is a slight bias arising from the difference between arithmetic means and harmonic means for the odd- and even-year spawning groups. This should not be large for SBT, where so many age classes contribute to spawning. In any case, if the breeding pattern is so clear, it will be possible to detect it, by following cohorts of juveniles and seeing whether they match predominantly against adults sampled in particular years, and then to fit two separate models to remove the arithmetic-harmonic effect.

With a three-year study and a two-year cycle, bias would reappear. However, if the pattern is clear enough to cause bias, it will also be clear enough to detect; matches will only ever occur between samples collected across a gap of a fixed number of years. A more complicated cyclic model could then be constructed.

6.0.4 Number of loci required in two-phase testing

The need to eliminate a high proportion of samples as "definitely non-relatives" in the first pass, sets a stringent limit on p_1, the probability of an accidental hit on the first pass. To eliminate say 90% of samples, we need a 90% probability that a given juvenile will not match *any* of the m_A adults by chance, so that $(1-p_1)^{m_A} = 0.9$. In general $p_1 \approx (1-X)/m_A$ where X is the proportion to be ruled out on the first pass; for m_A of a few thousand and X of around 0.9, this means choosing L_1 to achieve a p_1 on the order of 10^{-5}. Given a set of loci and their allele frequencies, computation of p_1 (the probability that two unrelated individuals will share at least one allele at every locus) is a straightforward exercise in genetic probabilities. Some consideration should really be given to scoring error, which makes the computation more tedious.

6.0.5 Age-specific catchability

The relative fecundity of an SBT aged a is determined by four factors:

$$\text{rel fec}_a \propto \text{average residence time}_a \times \text{spawning frequency}_a \times \text{batch fecundity}_a \times \text{viability}_a$$

Note that average residence time itself has two components: the probability of actually turning up in any given year, times the average residence time given that the fish turns up. There is no need to separate the two for abundance estimation purposes, so we just deal with average residence time.

Of these four factors, we neglect any age-effects on viability[18]. The other three factors determine the number of eggs produced. Previous histological work can be used to estimate the relative spawning frequency while present on the grounds, and the batch fecundity[19]. Therefore we can write

$$\text{rel fec}_a \propto \text{ave res time}_a \times \text{rel eggs per day}_a$$

where relative-eggs-per-day$_a$ is estimated externally. The probability of capture on the spawning grounds is also proportional to average residence time[20]. Until and unless enough archival tag data is found, we do not have any external estimates of average-residence-time$_a$, but these turn out not to be necessary for estimating N. To show this in principle, we will assume (purely for simplicity of presentation) that numbers-at-age, average residence time, and relative eggs per day are all negative exponential functions of age, with coefficients z, r, and g respectively. Then we can perform two "catch curve" analyses, as follows:

1. Use the log-slope of the age profile of adults in the spawning ground to estimate $z + r$.
2. Use the log-slope of age for *identified parents* to estimate $(z + r) + (r + g)$. The first term arises because older fish are more likely to be sampled, and the second because they generate more eggs. Note that there is at least a 2-year gap between spawning and being identified as a parent, since juveniles are not being sampled until they are age 2; hence there is no need to worry about adult-sampling removing potential spawners.
3. The difference between the two log-slopes is therefore an estimate of $r + g$. Since we have an external estimate of g, we can also estimate r.
4. Subtract the estimate of r in (3) from the estimate of $z + r$ in (1) to get an estimate of z.
5. The expected number of hits is $m_J m_A \mathbb{P}[\text{hit}]$. For notational simplicity, define the age-of-maturity as 0, with N_0 animals at that age and N adults in total. Then we have

[18] Whether or not this is correct, it is standard practice e.g. for calculating SSB. In fact, SSB calculations typically just assume that juvenile production is proportional to bodyweight, whereas we "go one better" by estimating the relationship empirically.

[19] For females, batch fecundity is proportional to $W^{2.4}$ where W is body weight. This is based on the change in gonad weight before and after a spawning event. For males, a slightly different approach is needed, based on absolute gonad weight, but the general idea is similar.

[20] A small further source of variability arises from fish of different sizes having different depth frequency distributions while on the spawning grounds, and thus potentially having different catchabilities per unit time. This is tied in with the estimation of spawning frequency earlier on. However, the effect (examined in Davis et al., 2003) is not large.

$$\begin{aligned}
\mathbb{P}\left[\text{hit}\right] &= \sum\nolimits_{a\geqslant 0}\left(\text{propn}-\text{sampled}-\text{adults}-\text{aged}{-}a\right)\times\mathbb{P}\left[\text{adult}-\text{aged}{-}a{-}\text{is}-\text{my}-\text{parent}\right] \\
&\quad \sum\nolimits_{a\geqslant 0}\left(\text{propn}-\text{sampled}-\text{adults}-\text{aged}{-}a\right)\times 2\times\frac{\#\text{eggs}-\text{from}-\text{an}-\text{age}{-}a}{\text{total}-\#\text{eggs}-\text{released}} \\
&= 2\sum\nolimits_{a\geqslant 0}\frac{e^{(r+z)a}}{\sum_{a'\geqslant 0}e^{(r+z)a}}\times\frac{e^{(g+r)a}}{N_0\sum_{a'\geqslant 0}e^{(g+r+z)a}} \\
&= \frac{2}{N_0}\frac{\sum_{a\geqslant 0}e^{(g+2r+z)a}}{\left(\sum_{a\geqslant 0}e^{(r+z)a}\right)\left(\sum_{a\geqslant 0}e^{(g+r+z)a}\right)}
\end{aligned}$$

For ease of exposition, replace the sums by integrals, and note that

$$N=\#\text{adults}=N_0\sum\nolimits_{a\geqslant 0}e^{za}\approx N_0\int_0^\infty e^{az}da=\frac{N_0}{z}$$

We then have

$$\begin{aligned}
\mathbb{P}\left[\text{hit}\right] &\approx \frac{2}{N_0}\frac{(r+z)(g+r+z)}{g+2r+z} \\
&= \frac{2}{(N_0/z)}\frac{(r+z)(g+r+z)}{z(g+2r+z)} \\
&= \frac{2}{N}\frac{(r+z)(g+r+z)}{z(g+2r+z)}
\end{aligned}$$

Note that if $r=0$ (i.e. no age-dependent catchability) then the right-hand fraction cancels to 1 whatever the value of g, and we retrieve the base-case formula. However, if $r\neq 0$, we do need to know g.

6.0.6 Population substructure and sampling bias

Suppose the entire adult population of N is made up of two sub-populations with proportions π and $1-\pi$, and that adults are sampled proportionally from their respective sub-population, so that the overall adult sample contains $m_A\pi$ fish from the first sub-population and $m_A(1-\pi)$ from the second. Juveniles, though, are not necessarily sampled in proportion to sub-population abundance; let m_{J1} and m_{J2} be the numbers sampled from each sub-population.

If the entire dataset is analysed without regard to sub-populations, then the expected number of hits can be calculated by considering samples from each sub-population separately (since there will be no cross-hits between juveniles from one sub-population and adults from the other):

$$\begin{aligned}\mathbb{E}[H] &= \frac{2m_{J1}(\pi m_A)}{\pi N} + \frac{2m_{J2}(1-\pi)m_A}{(1-\pi)N} \\ &= \frac{2m_{J1}m_A}{N} + \frac{2m_{J2}m_A}{N} \\ &= \frac{2m_J m_A}{N}\end{aligned}$$

just as in the case without sub-populations. In other words, the basic estimate is unbiassed provided at least one life-stage is sampled in proportion to sub-population abundance. If both are sampled disproportionately, though, there will be bias.

Update on SBT close-kin abundance estimation, 2008

Mark Bravington
Peter Grewe

Prepared for the CCSBT 5th Management Procedure Workshop 2-7 September and the 13th Meeting of the Extended Scientific Committee 8-12 September 2008
Rotorua, New Zealand

Abstract

We describe progress on estimating SBT spawner abundance using close-kin data, following on from the study proposed last year. Further samples have been collected, and more fish have now been genotyped, allowing us to examine the quality of the genetic data. Results are promising, and we expect to deliver preliminary estimates by CCSBT 2009.

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 last year in CCSBT-SC/0709/18 (Bravington and Grewe, 2007).

Project arrangements

Funding has now been provisionally agreed between CSIRO and FRDC, and the project will be overseen by a steering committee including international experts on genetics, mark-recapture, and tuna assessment. Funding began in July 2008, so there has only been limited time for further genetic analysis (see next section). The revised schedule for the project is described below.

CSIRO has continued to collect samples, as listed in Table 1; we are still receiving samples from Port Lincoln in 2008. Most of the Port Lincoln samples are from age-3 fish, with a substantial proportion of age-2s (based on length measurements). Otoliths are available for almost all the Indonesian samples from 2005-6 and 2006-7 (2007-8 data not available yet), and a number have been aged as part of the standard Indonesian ageing programme (Farley and Proctor, 2008).

Year (Jul-Jun)	Place	Samples held	DNA extracted
2005-6	Indo	216	216
	PL	4000	200
2006-7	Indo	1520	1069
	PL	4000	200
2007-8	Indo	1594	0
	PL	800+	0

Table 1: Samples collected and stored up to August 2008

Genetic progress

Bravington and Grewe, 2007, included preliminary estimates of how many loci per fish would need to be scored to exclude false positive matches (i.e. a juvenile and an adult that are actually not a parent-offspring pair, but that by chance happen to have at least one allele in common at every locus examined). Those estimates indicated that only a modest number of loci would need to be scored, thus making the costs feasible. However, the available data for those calculations came from only 16 fish, so the allele frequency estimates which underpin the calculation were inevitably uncertain. Further, it was not possible to check some aspects of locus reliability (e.g. null alleles) because of the small sample sizes.

We have now genotyped 96 adult fish (all so far from Indonesia 2005-6) at 18 polymorphic loci (mostly the same loci used last year, with some changes for technical reasons). The larger sample size gives us better estimates of allele frequency, and allows us to check for null alleles. All genotyping for this larger batch was done by the Australian Genome Research Facility, using primers and amplification protocols developed at CSIRO.

With the larger sample size, the allele frequency estimates for individual loci do change somewhat but without affecting the overall number of loci required for exclusion *if* genotyping is assumed exact (i.e. no scoring error). A few loci show some evidence of scoring error, in the sense that they have an excess of apparent homozygotic fish (departure from Hardy-Weinberg equilibrium, HWE), likely through inability to score the other allele on that fish— i.e. showing non-amplifying or null alleles. Table 2 shows the sorted p-values for HWE as produced by the GENEPOP program (Raymond and Rousset, 1995); small values indicate possible null allele issues. Given the number of loci being tested, some p-values will turn out small by chance, so in fact only the first 4 or 5 loci are of any conceivable concern; based on the p-values, at least 13 of the 18 loci show no evidence of null alleles.

It is important to note that, even if a locus does exhibit null alleles, the locus may still be useful for DNA fingerprinting; the presence of null alleles simply means that relatedness between a pair of fish cannot be ruled out based on that locus if either fish is an apparent homozygote. This less stringent criterion is less powerful statistically, but robust to null alleles. Depending on the p-value, we can decide whether to use the more stringent or less stringent criterion. If we use the 10 most powerful of the 18 loci in Table 2, and use the less stringent exclusion criterion for the 3 of the 10 with p-values below 5%, then we should still eliminate about 99% of possible false positive juveniles and adults (see section 6.0.4 in last year's paper for basis of calculation). The remaining small number of potential matches— most of which will in fact be true parent-offspring pairs— can be checked by examining a small number of extra loci, at minimal extra cost.

We also examined how often (i.e. in what proportion of fish) each locus failed to amplify

Locus	D111	D232	D139a	D11b	D225	B5	D201	D4D6	B232a
%p-value	0	0	0.14	0.98	2.3	5.0	14.8	22.7	22.9
Locus	3D4	D115	D211	D122	D10	D3	D135	D235	D12
%p-value	24.0	26.6	32.1	42.3	43.6	47.1	58.0	79.3	85.7

Table 2: Testing departures from Hardy-Weinberg equilibrium

at all; this is related to, for example, quality of tissue preservation, and is a serious problem in some genetic tagging studies (e.g. for Spanish mackerel in the Northern Territory, where small pieces of tissue remain uncollected for hours in warm water). In our Indonesian samples, though, there does not seem to be a problem. Depending on the locus, between 0 and 7% of the fish failed to score at all at that locus. However, most of these failures arose from just 3 fish (a small proportion of the 96— and even these may succeed if DNA is re-extracted). If those three fish are excluded, the average unscored genotype frequency per locus is about 1.5% and the maximum in any "top-ten" locus is about 4.5%. Hence total dropout does not seem to be a concern for the Indonesian samples, which are collected under very good conditions. Conditions for Port Lincoln samples should be even better.

The plan from here

Our immediate plans are to:

- genotype a larger set (300 juvenile fish) from the GAB, to examine incidence of siblings and half-siblings;
- finalise locus choice & protocols for mass genotyping;
- present results to date to the project Steering Committee in September 2008, to demonstrate feasibility and plan the next steps.

Although a high proportion (say >30%) of siblings or half-siblings would not bias our abundance estimates (see last year's paper), it would cause problems for precision, so the first check above is important for assessing whether our sample sizes are adequate. Assuming all is well, we will then proceed with genotyping all existing samples and with data analysis as described in last year's paper, presenting a preliminary report to CCSBT 2009. Data analysis is rarely a one-step process, so it is likely that further statistical analysis will be required before a more final report can be presented to CCSBT 2010. Final analysis and write-up will be completed by July 2011. There will also be a further year of sampling in Indonesia (summer 2008-9) and Port Lincoln (harvest 2009), and data from those samples will be available in time for CCSBT 2010.

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 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] Bravington, M. and Grewe, P. (2007). 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.

[Farley and Proctor, 2008] Farley, J., A. R. and Proctor, C. (2008). Update on the length and age distribution of sbt in the indonesian longline catch. Scientific Committee Report CCSBT-ESC/0809/27, Commission for the Conservation of Southern Bluefin Tuna.

[Raymond and Rousset, 1995] Raymond, M. and Rousset, F. (1995). GENEPOP (v.1.2): population genetics software for exact tests and ecumenicism. *Fisheries Science*, 86:248–249.

CCSBT-OMMP/1307/info/01

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.

CCSBT-OMMP/1307/info/01
CCSBT-ESC/1009/Info 2

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 Extended Scientific Committee for the 15th Meeting of the Scientific Committee 4-10 September 2010
Taipei, Taiwan

Abstract

This paper describes progress with the close-kin estimate of spawning biomass. There are now genotypes for about 5000 fish at up to 22 loci spanning 3 years of sampling in both Indonesia and the GAB, and we expect to have 7000 fish done by the end of 2010. The paper gives outcomes for basic feasibility checks: reliability of identifying parent/offspring pairs, and checking for excessive numbers of sibs or half-sibs.

Contents

List of Tables

1 Overview

This paper is an update on progress in the SBT Close-Kin Abundance project. The project began formally in 2008, though data collection started two years earlier. The *modus operandi* is to identify Parent-Offspring Pairs (POPs) via "DNA fingerprinting" (multilocus genotyping), amongst comparisons between adults caught on the Indonesian spawning grounds and juveniles caught in the Great Australian Bight. For samples of given size, the *expected* number of POPs is inversely proportional to the total number of spawning-age adults in the population. This fact can be used to turn the *actual* number of POPs found into a formal estimate of spawner abundance, and thus SSB, using mark-recapture principles. The estimate requires very few assumptions, is fishery-independent, and is not vulnerable to the reporting-rate issues that can plague conventional tagging programs. Background is given in Bravington and Grewe (2007).

Although the basic principles of the project are straightforward, a number of technical issues have had to be addressed. Most notable is the need to collect large numbers (7000+) of tissue samples and genotype them at large numbers of loci, in order to reliably identify enough POPs to permit model checking and to give a precise abundance estimate. The project is the first to attempt genotyping on such a large scale (for non-human subjects). The logistics and quality-control (QC), which are essential to the project, have required a great deal of work.

The project seems to be working. We have now genotyped over 4000 fish with sufficient reliability to eliminate false-positive POPs, and we have indeed found a number of definite POPs. We have checked the incidence of full- and/or half-siblings in the juveniles, and there seem to be too few to have any impact on CV. For reasons explained below, it would be quite inappropriate to estimate SSB using these very preliminary results. The number of POPs so far is also too small to look for phenomena such as skip-spawning. Nevertheless, the results are entirely encouraging, and consistent with the project schedule. We are on track to finish genotyping 7000 fish with full QC protocols this year, and to produce as planned an estimate of SSB for CCSBT 2011.

Extrapolating from the samples analysed so far, the number of POPs eventually found will probably be lower than we assumed when designing the project. This is not surprising, because the number found is dictated by the very quantities that are uncertain and that the project is in fact designed to estimate, in particular SSB. However,

Table 1: Status of samples

Year	Collected	DNA extracted	Genotyped
	Adults (Indonesia, Sep-Apr)		
2005-2006	216	216	
2006-2007	1520	1069	644
2007-2008	1594	1200	1130
2008-2009	1700	1700	736
2009-2010	1840*		
2010-2011	[1500]		
	Juveniles (Port Lincoln, Jul-Sep)		
2006	4000	600	478
2007	4000	800	736
2008	4000	1288	1104
2009	4000	1248	
2010	3300+		
2011	[4000]		
Totals	25230	8121	4828

[]: planned
+: ongoing
*: 640 still in Indonesia awaiting permits to export research samples, for which the rules are currently being changed.

fewer POPs would have two undesirable consequences. First, the CV of the SSB estimate would be higher; second, there would be less ability to check the POPs for phenomena such as skip-spawning, which if detected would necessitate adjustments to the estimation model. The resources of the current project extend only to genotyping the originally-planned 7000 fish, but we do have a further 18000 archived tissue samples (Table 1). Increasing the sample size by genotyping some or all of these would proportionally increase the number of POPs, and the confidence in the final result.

Separately to the current project, we are also continuing the collection of tissue samples from adults in Indonesia and juveniles in Port Lincoln. These could be genotyped in future to develop a time-series of SSB estimates.

2 History since CCSBT 2009

By CCSBT 2009, we had selected and optimized an initial set of loci, and had begun checking for sibship amongst juveniles. This year, we started by genotyping a subset of 500 juveniles from a single year and cohort at 11 loci, to check more thoroughly for high incidence of full- and/or half-sib(ling)s. If the sib incidence was very high, then the number of parents being looked for would be much less than twice the number of juveniles, and the CV of the estimate would be worse than expected[1]. In extreme cases, this could render a close-kin project infeasible, so it is important to check for high sibship levels before embarking on the expense of full-scale genotyping.

Results on locus development and sibship checking were presented to the project's Steering Committee in May 2010; there was no evidence of substantial sibship amongst the 500 juveniles (see also section 3.3). The Steering Committee agreed that the project should go ahead with genotyping the remainder of the planned 7000 fish, aiming to get 5000 done by CCSBT. The set of loci was to be expanded well beyond 11, to avoid swamping genuine POPs with false-positives POPs (see next section and Appendix).

Between May and July 2010, about 4800 fish were genotyped at 22 loci, and the data were entered by late July. After an initial (but not comprehensive) clean-up, a usable version of the dataset was ready by 10 August. The results in this paper are obviously preliminary.

Sample collection and preparation have continued throughout, as shown in Table 1.

3 Results of genotyping

The goal of the genotyping is to find all the POPs, and to do so without ambiguity. In particular, there should be a negligible proportion of false-positive POPs (unrelated pairs that happen to look like POPs), and enough

[1] As explained in Bravington and Grewe (2007), sibs and half-sibs do *not* cause bias in the abundance estimate, but can affect CV.

headroom to distinguish between false-negative POPs (true POPs which appear not to be, due to some error in the genotyping) and almost-false-positives (unrelated pairs that look like false-negative POPs). The solution to both issues is to use plenty of loci. More loci do require more resources, but the marginal cost of scoring a few extra loci is small compared to the difficulties and uncertainties that arise when a substantial proportion of POPs are likely to be false. The Appendix gives further background on false-positives, false-negatives, and criteria for assessing POP status.

Since 2009, we have greatly increased the number of loci used, because it became clear that we would otherwise encounter numerous false-positive POPs. We are currently using 22 loci, although one has proved hard to score reliably on a substantial fraction of the 5000 fish, and has been excluded from the routine checks. Of the remainder, seven show statistically significant evidence of null alleles (Appendix), but at low levels. Fortunately, it is easy to relax the exclusion criterion used for assessing POP status to allow for nulls, and there is little increase in the false-positive rate provided the null-allele rate is small.

Most samples have been scored successfully at most loci. About 600 fish (grouped into 6 Plates of 92 fish) are currently missing an entire Panel (a group of 3-6 loci that are all processed simultaneously). Those plate-panel failures are only temporary; they can be resolved cheaply by re-processing. Aside from those, only about 5% of the samples (228 of 4800) appear to be unusable, i.e. failing to score at large numbers of loci. These encouraging figures reflect the generally very high quality of the tissue and its state of preservation.

The volume of samples and loci is immense: 5000 tissue samples were cut down to size, and then had DNA extracted; DNA was organized into 54 plates each of 96 fish; each plate was processed 5 times, with a different subset of the 22 loci being scored each time; the final dataset contains 200,000 alleles. To avoid mixups, stringent QC is needed at all steps along the way. We are still finalizing the QC process, but by the end of 2010, it will be possible to trace each fish through every step, to confirm that the fish still has the genome it started with.

3.1 POPs

To avoid problems with false-positives, we are currently restricting attention to fish with at least 16 scored loci, which equates to about 2000 adults and 2000 juveniles. Because different fish have different loci missing, the number of loci compared in each pair varies between 10 and 21, so individual comparisons vary considerably in their false-positive probability. However, with this subset of fish, the total expected number of false-positives is just 0.03.

We found 7 POPs that matched at every locus compared. They look genuine; all were based on comparisons of 18 or more loci, and also matched at the extra 22nd locus, which was not used in screening for POPs. There were 2 pairs that failed to match at just one locus. On investigation, both seem to be almost-false-positives (unrelated), rather than false-negatives (true POPs with a mis-scored locus). For one thing, the numbers of loci compared were just 14 and 11, unlike the 18+ comparisons involved with the definite POPs, so the chance of an almost-false-positive is relatively high; also, we have re-checked their mismatching loci, and there is no ambiguity about the genetic signatures.

As a consistency check, to see whether the process would generate false POPs, we also compared the juvenile sample just with itself, and the adult sample just with itself. In each case, the number of comparisons involved is about the same as for the real juvenile-adult comparison. However, since there cannot be any genuine juvenile-juvenile or adult-adult POPs[2], any POPs found would be have to be false-positives. Reassuringly, there were none.

There are some subtleties around the criterion for deciding which fish (or pairs of fish) to use in comparisons. The 16+-loci criterion used so far is temporary, and we will refine it intersessionally. Table 2 shows the effect of increasing the stringency; moving down the rows, a clear gap opens up between true POPs in the leftmost column and the closest almost-false-positives to the right, but at the expense of sample size in the Total column. In terms of an SSB estimate, there would be a bias-variance trade-off: stricter criteria mean fewer false positives and thus less bias, but also lower sample sizes and thus increased variance. Once the criterion is stringent enough to keep the expected number of false-positives well below one, then there is no reason to make the criterion any stricter. For criteria based on minimum-number-of-loci, this occurs at the 16+ level, where the expected number of false-positives is 0.03. In the table, the 16+ row is also where a clear dip between complete matches and single mismatches first appears.

The 15+ row is interesting. There is no dip in the frequency of almost-POPs, suggesting that false-positives POPs are likely, and this is confirmed by probability calculations. There are an extra 7 apparent POPs compared to the 16+ row, but 6 of them compare only either 9 or 10 loci, and may be false-positives. The 7th extra potential

[2] After excluding comparisons of a fish to itself, obviously. Note that it *is* theoretically possible to have an adult-adult POP, but there would have to be an age gap of at least 10 years. The number of such comparisons is very small.

POP is a 15-locus comparison, and its status is unclear. Many of the exactly-15-loci fish will gain extra loci once the failed plate-panels have been re-run, so these particular mysteries will be resolved. On the whole, it seems that the current set of loci are almost sufficient to deal with the false-positive-and-negative issue but, as the Appendix explains, it will be more secure once an extra couple of loci have been added, particularly if larger samples of fish are compared in future.

Table 2: Number of comparisons.

		Number of mismatching loci						
		0	1	2	3	4	...	Total
Min .#loci for inclusion	15	14	35	310	1527	6157	...	4807224
	16	7	2	15	110	896	...	4096196
	17	7	0	6	62	462	...	3840489
	18	7	0	3	28	247	...	3486439
	19	7	0	1	12	106	...	2842784
	20	3	0	0	4	47	...	1964118

Rows are strictness of criterion; columns are number of mismatching loci. The zero-column shows POPs, be they genuine or false-positive; the one-column shows false-negatives or almost-false-positives involving a single locus; the two-column shows pairs that mismatch at two loci, etc. The Total column shows the "sample size".

3.2 Description of the POPs found

The parents have not yet had their otoliths read. That aside, here is a brief description:

- All 7 parents are distinct (i.e. no more than one matching offspring per parent).
- There are 2 female and 5 male parents.
- Most of the 7 parents were slightly above the median size of adults captured in the same year. Sizes ranged between 161cm/93kg and 177cm/115kg. The parents would of course have been smaller when they spawned their offspring.
- 3 parents were caught in 2006/7, 1 in 2007/8, and 3 in 2008/9.

There is less to say about the offspring. All were 3-year-olds, but then we have deliberately concentrated on genotyping 3-year-olds so far. One was caught in 2006, and six in 2008 which, as per Table 1, is also when the most juveniles were sampled[3].

Once sufficient POPs have been found, they can be used to detect and estimate various unexpected aspects of breeding biology, such as skip-spawning, relationship between size and residency on the breeding grounds, and temporal stock structure (e.g. if all parents of GAB juveniles were caught in a limited part of the spawning season). However, with only 7 POPs to date, it is impossible to say anything definite about these such questions yet.

3.3 Checking sibship

The genotype data can also be used to estimate the incidence of sibs and half-sibs within each cohort and capture-year of juveniles. As noted last year, the genotype data is much less informative about sibship, and particularly half-sibship, than about POPs. Studying sibship in large samples is is a hard problem statistically, much harder than the main part of this project. There are no existing statistical algorithms, so we have developed our own. The (extensive) technical details are omitted from this report, but have been reviewed by the Steering Committee.

Since May 2010, we have adjusted the sibship algorithm to cope with null alleles, which can otherwise give spurious evidence of sibship. We have also applied it to the entire juvenile sample (one year at a time), not just to the subset mentioned in Section 2. The results give no indication that sibship is common. The point estimate is zero in two years, and involves just a handful of fish in the third. Because the number of loci used is limited, the results cannot exclude the possibility of small numbers of sibs and halfsibs, but small numbers would in any case pose no problem for the project. Simulations suggest that our algorithm is effective at detecting high incidences of sibship, given the number of loci and the sample sizes used, so the absence of evidence can be taken as evidence

[3]This oversampling was deliberate, to increase the proportion of the adult sample that would have been mature when the juvenile samples were spawned. Appropriate allowances will be made in the full SSB estimation model.

of absence. It is notable that none of the POP parents match to multiple offspring, so sibs/half-sibs cannot be overwhelmingly common.

We will continue to analyse sibship as more samples are genotyped.

4 Discussion

The results so far are very encouraging, given the innovative nature of the project. However, there are a number of reasons why these interim results **should not be used in a formal abundance estimate**. The following list is not exhaustive:

- QC is still incomplete. In particular, there is a real possibility that some as-yet-unchecked plate-panels may have been inadvertently swapped, as actually happened for a couple of the plate-panels that we have been able to check (and correct). Every undetected plate-panel swap would affect about 200 fish, completely disguising any POPs in about 10% of comparisons so far, so it is crucial to finish those checks.
- It is not yet clear what the appropriate number of comparisons is for estimating SSB. For example, there has been no exclusion of those "adults" that would actually have been immature at the juvenile birthdate; a just-mature adult in 2008 could not be the parent of a 3-year-old juvenile that was caught in 2006 and therefore spawned in 2003.
- 7 POPs is far too few to decide on appropriate model structures, e.g. whether or not skip-spawning is commonplace.
- With this small number of POPs, the sampling variability is very high. The 90% confidence interval for the "number of matches that *should* be present" is $[4.0, 13.2]$ and this range would be proportionally reflected in the interval for any SSB estimate.

Notwithstanding the above, it is worth noting that 7 POPs from about 4,000,000 comparisons is within the range to be expected, based on the range of values for 2004-2006 SSB considered in the OM scenarios, and using the naive assumptions of the simplest possible mark-recapture model.

5 Prospects for 2011

In this final year of the current project, we plan to:

- finalize QC procedures, including rescoring missing panels and reorganizing some of the existing panels;
- add a small number of additional loci, to widen the gap between almost-false-positives and false-negatives;
- genotype another 2000 fish (50/50 adults and juveniles);
- identify POPs
- read otoliths of the parents;
- check for patterns in the POPs and develop the full estimation model accordingly;
- estimate SSB and associated CV.

Note that the SSB estimate will be retrospective to juvenile birthdates, i.e. roughly 2004-2006. The precise definition will depend on details of the estimation model, e.g. whether it is time-averaged or disaggregated by year. Those details will not be decided until we have checked for patterns in the POPs.

The CV of our estimate in 2011 will depend primarily, though not exclusively, on the *number* of POPs that we find. All else being equal (sample sizes etc.), a smaller SSB would mean more POPs and a smaller CV. If the SSB was much larger than we assumed when designing the project— and the whole point of the project is that the true SSB is *not* known precisely— then the CV would also be larger[4].

The CV will also depend on any adjustments that need to be made to the estimation method following model checking. In this instance, "model checking" means looking for patterns among the POPs: e.g., if successful parents tend be unusually large, or if there is evidence of skip-spawning. Such phenomena would not present any fundamental

[4] Happily, the lower confidence limit would still be larger, despite the increase in CV.

problem for the project, but would necessitate adjustments in estimation which would affect CV; see Bravington and Grewe (2007). To do the model checking, there must of course be a reasonable number of POPs in the first place.

So far, we have used about 4000 fish in pairwise comparisons. Once we have 7000 fish genotyped, the number of comparisons will actually triple, thanks to the quadratic relationship with sample size. Extrapolating from our 7 current POPs, we might therefore expect around 20 POPs next year, although the final figure might be considerably higher or lower. Using the simplest estimation method, 20 POPs would yield a 22% CV, which sounds respectable. However, the estimation method that ultimately gets used is sure to be more complicated, since it will have to take into account the multi-year nature of the study, etc. The extra parameters required will increase the CV. In addition, the amount of model-checking that is feasible with 20 POPs would be limited.

A particularly important check is on the strength of any relationship between fish size and residency (length of time spent) on the spawning grounds. Residency affects not just catchability and therefore a fish's chance of appearing in the adult sample, but also the number of juveniles it is likely to have contributed three years previously (since most big fish now were also big three years ago). In mark-recapture terms, this induces "heterogeneity of capture probability". If the size-duration relationship is strong, it can bias the abundance estimate, and also the interpretation of SSB. As described in Bravington and Grewe (2007), it is possible in principle to estimate and allow for the size-duration relationship, by comparing the size distribution of parents with that of the general adult sample. A reasonable number of POPs will be needed to do this with confidence. Another way to infer this relationship, and an independent check on this aspect of the SSB estimation model, would be via archival tag recaptures from mature fish. However, this would take some years. In the meantime, genotyping some of the 18000 archived samples would be the fastest and cheapest way to improve model-checking and reduce CVs.

6 Acknowledgements

This project relies on our samplers. In Indonesia, 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 tissue samples. In Australia, Tony Jones of Protec Marine has done an outstanding job in collecting over 19000 tissue samples.

We appreciate the assistance of 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, M. and Grewe, P.: 2007, 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.

7 Appendix: how to identify POPs from DNA

Every fish has two *alleles* (i.e. particular sequence of DNA) at each *locus* (i.e. specific place in the genome). The loci we use are *microsatellite repeats*, where many different types of allele at each locus are found across the population. Usually, a fish has two different alleles at a locus (i.e. it is *heterozygous*), but by chance the two alleles may be the same (*homozygous*). The different types of allele are distinguished by their length, and *scoring the locus* means "measuring the lengths of the two alleles". *Genotyping a fish* means "scoring it at all loci possible"; sometimes a tissue sample simply will not yield a result at a given locus, but this does not matter as long as enough other loci are successfully scored.

An offspring inherits one allele at each locus from each parent, so a POP must *match*, i.e. have at least one allele in common, at every locus. The following *exclusion criterion* can therefore be used to determine POP status: two fish are deemed a POP if they match at every locus, but the presence of even one mismatching locus excludes the possibility of being a POP. Matches at all loci can also happen by chance even for unrelated fish; this would be a *false-positive* POP. However, the per-fish-pair false-positive probability becomes vanishingly small if enough loci are compared: about 2×10^{-10} for a 20-locus comparison in this project. Since the overall number of fish-pairs will only

be about 10^7, complete-match false-positives will not be a concern provided we restrict attention to comparisons with enough loci.

The picture is made more complicated by the possibility of *false-negatives*, i.e. true POPs that somehow appear to have one locus without an allele in common, and that therefore fall foul of the exclusion criterion. False-negatives could arise by either of two mechanisms. The first, *mutation*, can be ignored because of its rarity; estimates are typically $\sim 10^{-3} - 10^{-5}$ per generation. The second is *scoring error*, where the alleles recorded are not the same as the alleles actually present. There are many reasons for scoring error, but its incidence can be minimized by careful choice of loci, optimization of the associated scoring process, and having high-quality samples to begin with. Unlike false-positive rates, scoring error rates cannot be predicted; in this study, they will only be directly estimable after a large number of POPs have been identified. While typical estimates of scoring error in other studies are around 1% (i.e. at least 10 times as common as mutation), we can expect to do rather better, because of high quality samples and careful design. To be cautious, though, if the scoring error rate was as high as 1% per locus, then with over 20 loci scored, a significant fraction of true POPs would mismatch at one locus and be incorrectly excluded. (The proportion mismatching at two or more loci should be negligible.) We do therefore have to somehow relax the exclusion criterion to allow for the possibility of some false-negative POPs. These false-negatives could be confused with *almost-false-positives*, i.e. unrelated fish that fail to match only at one locus.

There are not many pairs that mismatch at just one locus, so they can be re-examined; this can sometimes identify a scoring error, and thus distinguish a false-negative from an almost-false-positive. However, scoring errors cannot always be detected on re-examination. The most secure solution is to further increase the number of loci, so that the expected number of almost-false-positives (one mismatching locus), as well as complete false-positives (zero mismatching loci), becomes negligible. The expected number of false-negatives increases linearly in the number of loci, whereas the expected number of almost-false-positives decreases exponentially, so with enough loci it becomes almost certain that any out-by-one comparison is really a false-negative. The almost-false-positive rate in this project is about 40×higher than the complete-false-positive rate, at around 10^{-8}. Hence, the current set of loci is just about adequate for our intended 10^7 comparisons, provided almost all fish are scored at almost all loci. The planned addition of another two loci should fully resolve any ambiguities about pairs that mismatch at just one locus, and should future-proof the comparisons as and when sample sizes increase.

The above omits several technical details, two of which are worth noting here:

- Certain *null* alleles at some loci may not be scorable, even though heritable in the usual way. Fish with a null allele will look like homozygotes with two copies of whichever other allele they have. This phenomenon can be detected statistically at a population-wide level, and allowed for by a slight relaxation of the exclusion criterion. A similar remark applies to a phenomenon called *long-allele dropout*. There are several different ways to handle these phenomena in the context of POP-hunting. Statistically, the most powerful approach looks to be a relaxation of the per-locus exclusion rule to accommodate those two phenomena only, plus a relaxation of the overall exclusion rule to allow (or at least re-check) single-locus mismatches, the latter being a guard against all types of scoring error.

- For, say, aunt-nephew comparisons, the complete-match probability also turns out to be very small, although much higher than for unrelated pairs. However, there are vastly more unrelated pairs than non-parent-offspring close-kin, by a factor of at least 10^5. Provided enough loci are used to eliminate false-positives from unrelated pairs, there will not be a problem with false-positives from non-parent-offspring close-kin.

CCSBT-OMMP/1307/info/01
CCSBT-ESC/1107/19

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 Extended Scientific Committee for the 16th Meeting of the Scientific Committee 19-28 July 2011
Bali, Indonesia

CLOSE-KIN UPDATE FOR CCSBT SC 2011

MARK BRAVINGTON, PETE GREWE, CAMPBELL DAVIES: CSIRO HOBART, JULY 2011

Introduction

This paper gives an update on progress and plans in 2011 for the SBT Close-Kin Abundance project. The project began formally in 2008, though data collection started two years earlier. The *modus operandi* is to identify Parent-Offspring Pairs (POPs) via "DNA fingerprinting" (multilocus genotyping), amongst comparisons between adults caught on the Indonesian spawning grounds and juveniles caught in the Great Australian Bight. For samples of given size, the *expected* number of POPs is inversely proportional to the total number of spawning-age adults in the population. This fact can be used to turn the *actual* number of POPs found into a formal estimate of spawner abundance, and thus SSB, using mark-recapture principles. The estimate requires very few assumptions, is fishery-independent, and is not vulnerable to the reporting-rate issues that can plague conventional tagging programs. Background is given in CCSBT-SC/0709/18.

Last year's update, CCSBT-ESC/1009/Info 2, described the genotyping (AKA scoring) and POP matching of about 4000 fish[1]. Seven definite POPs were found. Finding those POPs showed that the project was succeeding at a technical level. However, for the number of fish examined in 2010, the number of POPs found was substantially less than expected. In itself, this is not particularly surprising; as with many sample design problems, the actual sample size required to achieve a given precision (which in this case is set by the number of POPs) depends on the very thing that one is trying to estimate (in this case adult abundance), which is of course unknown. Nevertheless, it was clear in 2010 that, if the final sample size stayed at the 7000 originally planned, then there would likely not be enough POPs to give a precise estimate by the end of the project. The real issue is not so much the "CV" per se, but rather the need to get enough POPs to see any important patterns in them and then to formulate an appropriate and unbiassed statistical model accordingly. Even though the basic principles of POP-based abundance estimation are very simple, the application to SBT does require some care, because of the multi-year nature of the study and the interaction between the biology and sampling (fishing). For example, SBT may turn out to exhibit skip-spawning, and/or there may be links between adult size and effective fecundity. Both phenomena can be detected from POPs, and then allowed for in the statistical model, but only if there is a reasonable number of POPs in the first place.

In late 2010, the funding agencies CSIRO and FRDC therefore agreed to double the sample size (i.e. number of fish genotyped) to around 14-15,000. Extrapolating from the 2010 results, this should increase the number of POPs to[2] around 70-80, close to the original intention. The modification was possible because the project had deliberately collected a buffer of extra samples every year since 2006; the marginal cost of collection (as opposed to genotyping, which might never have been required) was low, and the extra samples provided a buffer in case the sample size ever did need to be increased. Even though the samples were already available, it has required a huge effort to process, genotype, and database 7000 samples within 8-9 months. Other tasks completed along the way (see below) have been to extend the set of loci (to give better surety about parent-offspring ID, and to cope with the increased potential for false-positives caused by an increased sample size), to fill in the gaps in the 2010 results, and to set up robust and traceable quality-control checks.

Timeline and further work. We are currently finishing a more formal range of quality-control checks to handle the greatly increased volume of genotyping data, which is now coming in at the rate of about 2000 fish per month and has undergone a number of changes of format since the start of the project. The files containing genotype scores are being linked to the existing CSIRO database, so that (adult) fish can

[1]Partial genotypes were available for another 1000 fish last year, but those data were too sparse for POP-hunting.

[2]This extrapolation is also uncertain, since it is based on a count of just 7, but is based directly on real data.

be cross-referenced for age, date of capture, etc. The QC code does need to be in place before another round of POP-checking can be done, but POP-checking itself is quite quick and straightforward. Once POPs have been found, most of the adult POP members will need to have their otoliths specially read, although some will already have been read in the existing Indonesian ageing project. These processes can happen in parallel with the rest of the genotyping, which we expect to have completed by mid-September (although delays in obtaining samples from Indonesia may compromise this slightly; see below). The statistical model will also be ready by the end of August, at least in preliminary form; its construction has deliberately been on hold until enough new data became available to select a sensible model. The draft final report is due on 30th December 2011.

No formal arrangements have yet been established for continued sampling from Indonesia or Port Lincoln in 2011/12, but the low cost of collecting the samples (even if they end up never being genotyped) presents a good case for continuation. Close-kin studies have the remarkable property of a *quadratic* gain in efficiency with sample size; extra samples now will both enhance the effectiveness of the existing data, and open the door to the development of a fully time-dependent (time-series) estimator of abundance.

Progress with genotyping

Table 1 shows the status of DNA extraction and genotyping for samples from various years of the study. We have now genotyped nearly 9000 fish. The whole procedure is now highly streamlined, with a mixture of in-house and outsourced steps to get the best in both quality control and cost-effectiveness, and the current throughput is over 2000 fish per month.

There is no formal design to our choice of samples from different years, and nor would it be possible to come up with such a design before the results are in. However, we aim for (i) a roughly even split between juveniles and adults, since that maximizes the expected number of POPs for a given expenditure (and in practice means using as many adults as possible, since they are in short supply), (ii) a fairly even spread across years, to give the best chance of allowing for any time-related effects that emerge, and (iii) selection of juveniles of fixed, known age (based on length). Recent changes to Indonesian administrative requirements for export of biological samples have led to unexpected delays in obtaining the necessary approval to ship the most recent samples, and in the worst case we may not have enough time to genotype them within the current study. However, even if that does happen, the shortfall can be made up from the stocks of juvenile fish.

The suite of loci has been expanded and reorganized since 2010. A total of 25 hypervariable microsatellite loci are now used for scoring and one for cross-checking, five more than in 2010. The key is to use enough loci to ensure that the expected number of false positives is much lower than the expected number of true positives, now expected to be 70-80. There will be about $7000^2 \approx 5 \times 10^7$ adult-juvenile comparisons in all, so the average probability of a false-positive needs to be kept to less than, say, 10^{-8} to expect less than one false-positive overall (number of comparisons times the probability of each one being false-positive). Not all loci are scored successfully for all fish, and a typical comparison will involve about 18 loci. The false-positive probability based on comparing 18 "typical" loci is about 7×10^{-10}, so there should be plenty of buffer against false-positives. The overall sample size may need to be reduced somewhat to exclude fish with few successfully-scored loci, since such fish will otherwise greatly increase the false-positive rate, but overall there should be plenty of loci even if the close-kin study continues and the sample size grows over time.

To check the consistency of our genotyping, this year a number of plate/panels have been re-scored (genotyped) by different readers. While there are very few instances of dramatically different scorings (say, one reader scoring as AB and another as AC or even CD), there are a modest number where a locus has been scored AB once and AA once. So far, it seems that that a genuine second allele has been overlooked in these cases, rather than a spurious allele being invented. Most of the fish have only been scored once, for obvious reasons of cost, and it is therefore likely that there are a few mis-scored "homozygotes" among them. The overall effect cannot be large, since scored homozygotes, whether real or not, are uncommon in all our loci (which are deliberately chosen for hypervariability, and thus low homozygosity). Nevertheless, it is possible that the low but non-zero estimates of heritable-null allele frequency at some loci, mentioned last year, are at least partly an artefact of genotyping error.

TABLE 1. Status of sample collection and processing, July 2011

Year/ Place	Collected / Archived	DNA extracted	Genotypes complete	Genotypes planned†
Adults (Indonesia)				
2005-6	216	216		210
2006-7	1520	1520	1520	
2007-8	1594	1594	1564	30
2008-9	1637	1632	1380	252
2009-10	1200+680*	1172	1104	748
2010-11	~1000*			~1000
Total A	**~7850**	**6134**	**5568**	**~2250**
Juveniles (GAB)				
2006	4010	1440	460	920
2007	4065	1472	736	644
2008	4027	1452	1104	276
2009	4103	1440	1012	368
2010	4071	1440		1380
2011	[4000]			
Total J	**24300**	**7244**	**3312**	**~3550**
TOTAL	**32850**	**13378**	**8880**	**~5800**

* : in storage awaiting export from Indonesia
†: approximate, depending on best way to organize genotyping

To illustrate how we plan to handle POP-finding and the issues around false positives, false negatives, and scoring error, we include an exact copy of last year's Table 2; note that it has not yet been updated to include the new data. With more loci scored this year, the gap between the "lucky lookalikes" (the right-hand columns— pairs of fish which by chance share alleles at a lot of loci) and the true POPs (the left-hand column) will be bigger than in the Table. Even if the occasional scoring error does results in a small number of pairs with an apparent mismatch at one locus (i.e. true POPs which should be in the left-hand column, but has moved into the next one), they will still be clearly separated from the lucky lookalikes, and will not be automatically rejected— i.e. they will not become false negatives. Any pairs will small numbers of mismatching loci will be re-scored, and their POP status will be assessed taking into account the nature of the apparent mismatch (as per previous paragraph).

Also, we have now filled in most of the big gaps from last year. These arose when an entire plate/panel failed to work, thus removing about 5 loci from 100 fish and leading to a lot of pairwise comparisons involving rather few loci. This means we can now afford to be reasonably stringent about the "entry requirement" for a comparison (i.e. how many loci must be scored in both fish of a pair, for that comparison to be considered) without sacrificing too many potential comparisons. In terms of Table 2, that means we should have a good-sized "sweet spot" between the bottom row (where the entry requirement was too strict, and the number of comparisons was thus much reduced) and the top rows (where the entry requirement was too lax, and the lucky lookalikes overlapped with the true POPs).

We have also tuned our quality-control procedures this year. With so many fish involved, and each plate of ~96 fish needing to be run through the equipment on several separate occasions (i.e. in different "panels"), it is essential to have some way to check that the fish have not become muddled up. Each plate of fish is uniquely coded by using water "blanks" in specific positions. Each plate also uses positive controls (i.e. two known "standard" fish) in specific positions, to give a controlled product and an additional key to identify a plate across all panel runs. All runs are multiplexed, with a standard tube used for each panel and all plates. This and other protocols have allowed us to detect and fix several problems that could otherwise compromise identification of POPs.

TABLE 2. Number of pairwise comparisons, by number of mismatching loci involved (COPY OF 2010 TABLE, WITHOUT NEW DATA; JUST FOR ILLUSTRATION). Rows are strictness of entry requirement for a pair of fish to be compared, in terms of the minimum number of loci at which *both* fish are successfully scored. Columns are number of mismatching loci with. The zero-column shows POPs, be they genuine or false-positive; the one-column shows false-negatives or almost-false-positives involving a single locus; the two-column shows pairs that mismatch at two loci, etc. The Total column shows the "sample size", i.e. total number of pairs that meet the entry requirement.

		Number of mismatching loci						
		0	1	2	3	4	...	Total
Min .#loci for inclusion	15	14	35	310	1527	6157	...	4807224
	16	7	2	15	110	896	...	4096196
	17	7	0	6	62	462	...	3840489
	18	7	0	3	28	247	...	3486439
	19	7	0	1	12	106	...	2842784
	20	3	0	0	4	47	...	1964118

ACKNOWLEDGEMENTS

This project relies on our samplers and technical support crew. In Indonesia, thanks are due to Craig Proctor (CSIRO), to the scientist team at Research Institute for Tuna Fisheries (Benoa, Bali), and in particular to Mr Kiroan Siregar, and Mr Rusjas Mashar (sampling enumerators) for their continued diligent efforts in ensuring the successful collection of tissue samples. In Australia, Tony Jones and Adam Kemp of Protec Marine have done an outstanding job in collecting over 20000 tissue samples from Sam's Tuna and Tony's Tuna. At CSIRO Hobart, Peta Hill, Rasanthi Gunasekera, Matt Lansdell, Scott Cooper, Danielle Lalonde, Bruce Barker, and Mark Green have made heroic efforts this year to process, genotype, and database what has become an enormous quantity of samples.

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

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

Report of the Close-Kin Project: estimating the absolute spawning stock size of SBT using genetics

CCSBT-ESC/1208/19

Mark Bravington Pete Grewe Campbell Davies

CSIRO Marine Lab, Hobart: July 2012

In confidence to CCSBT SC;
not for circulation

Contents

Abstract

It is in principle possible to estimate the absolute abundance of adult SBT without using catch or CPUE data, via a variant of mark-recapture applied to parents and offspring identified by genotyping large numbers of adults and juveniles. The method was first described in CCSBT-SC/0709/18, and since 2006 we have been running a large project to implement it. The project is now coming to a successful end, and this paper describes the [main] outcomes. We genotyped over 13,000 SBT caught between 2006 and 2010 in the GAB (juveniles) and off Indonesia (mature adults), and found 45 Parent-Offspring Pairs (POPs). Combining data from the POPs (the number found, plus their age, size, sex, and date of capture) with fecundity-at-size studies and Indonesian length, sex, and age-frequency data, we constructed a self-contained assessment of absolute adult abundance that does not require any catch or CPUE data. As well as abundance, we were able to estimate adult survival, selectivity-size relationship, and effective female reproductive contribution as a function of length. This paper explains the method, and presents an example of results for a steady-state scenario. These results, plus those from a limited number of other scenarios explored to date, indicate that adult abundance is considerably higher than current OM estimates. A small amount of work remains to finalise our self-contained assessment and more fully explore the model uncertainties, and we expect to complete it in the coming months as part of final project reporting. Options for the integration of the new data into the OM are considered in CCSBT ESC 1208/21.

Acknowledgements

This project has run for over 5 years with contributions from many people, too many to thank individually. Particularly heroic efforts on genotyping and logistics have come from Peta Hill, Rasanthi Gunasekara, and Matt Lansdell of CSIRO, with Jess Farley of CSIRO providing invaluable help with otolith and fecundity data. Special thanks for their diligent efforts in ensuring the successful collection of Indonesian tissue samples are due to Craig Proctor (CSIRO); Mr Kiroan Siregar, Mr Rusjas Mashar, and other scientists at the Tuna Fisheries Research Institute of Benoa; and Ms Retno Andamari (Research Institute for Mariculture, Gondol, Bali).

We greatly appreciate the continued assistance of the the Australian SBT industry (Port Lincoln), the Indonesian tuna fishing industry (Benoa), and the Research Centre for Fisheries Management and Conservation (Jakarta). Funding has been provided by CSIRO Wealth from Oceans flagship and the Fisheries Research and Development Corporation of Australia, with in-kind support from the Ministry of Marine Affairs and Fisheries, Indonesia.

1 Introduction

The SBT close-kin abundance project rests on two simple ideas:

- modern genetics allows us to tell whether any two fish constitute a Parent-Offspring Pair, via "paternity analysis";
- all juveniles have two parents.

Consequently, if you compare any juvenile with a randomly-chosen adult, there is a probability $2/N$ that you will discover that the adult is one of the juvenile's parents, where N is the adult population size. Given large samples of m_J juveniles and m_A adults, and repeating the comparison for each of the $m_J \times m_A$ pairs, some number P of Parent-Offspring Pairs (POPs) will be observed. The expected value of P is $2m_J m_A/N$, so using the observed value the adult abundance can be estimated from $\hat{N} = 2m_J m_A/P$. The devil lies in two details: doing the genetics well enough to correctly find the number of POPs, and adjusting for sampling biasses in the "randomly-chosen adults" aspect (e.g. selectivity biasses towards larger fish).

The project began collecting adult samples in Indonesia during the 2005/2006 spawning season (November-April), and in Port Lincoln during the 2006 season. Subsequent progress has largely followed the approach described in CCSBT-SC/0709/18, and a progress report has been given to CCSBT each year. Preliminary results presented to CCSBT in 2010 and 2011, after most of the originally-planned sample size had been genotyped, showed that the number of POPs found would end up considerably lower than originally expected if the original sample size was maintained, which would make the precision of the final results very uncertain. It was therefore agreed to substantially increase the sample size. This was straightforward since we have available many more frozen tissue samples from juveniles 2006-2010 than we have budget to genotype, but of course required extra time, so the final results of the project are only available now. Oversight of the project has been provided throughout by a Steering Committee including expertise in genetics, mark-recapture, and SBT assessment.

The main departures from the original project plan are that we have genotyped more fish over a longer period, used more loci to ensure the genetics are unambiguous, and have developed a full stand-alone statistical mini-assessment of adults during the 2000s in order to deal with the complexities of growth, selectivity, and fecundity. The assessment uses the POP data, external studies on fecundity-size relationships, and Indonesian length/sex/age composition data, but no catch or CPUE data.

Table 1 shows the final breakdown of 13,023 genotyped samples by year and site. A few hundred more were genotyped, but excluded in the end for assorted quality-control reasons. Although the optimal scheme for a given budget would have been to genotype equal numbers of juveniles and adults (since this is likely to yield the greatest number of POPs for a fixed amount of genotyping effort), regulatory changes and delays with Indonesian export permits meant that we had to shift the balance somewhat towards juveniles. Almost all the Port Lincoln juveniles were age 3 in the year of sampling (based on clear separation of modes in the length frequency), except for a few in 2006 that were age 4. After 2006, the Indonesian samples were taken from every available fish (almost all >150cm length) alongside the existing length/weight measurement and otolith-collection schemes[1]. Sample collection is continuing in both Indonesia and Port Lincoln, but there are no immediate plans or funding to genotype more samples; they are simply being frozen for possible future use.

[1] In two years, some Indonesian vessels fished further south than usual, off the main spawning ground, and were catching subadult fish. Fish from those vessels and years are excluded both from the genetic study and from the length-frequency data.

Table 1: Final tally of fish genotyped successfully. For Indonesia, "year 2006" means "spawning season from November 2005 to April 2006", consistent with the definition of "SBT birthdays".

	2006	2007	2008	2009	2010	Total
Indonesia	214	1457	1526	1394	1164	5755
Port Lincoln	1523	1707	1448	1338	1432	7448
Total	1737	3164	2974	2732	2596	13203

In the rest of this report, section 2 describes the principles and the results of our quest for POPs; further details of the genetic procedures and QC aspects are given in Appendix 1 (section 6), and further statistical details in Appendices 2 and 3 (sections 7 and 8). Section 3 describes some qualitative findings about the POPs found (e.g. average size), section 4 explains the mini-assessment and shows results from one example of its application, and section 5 is a summary.

2 Genetic results: finding POPs

We use the genetic data to find POPs, by first genotyping all the fish and then comparing every juvenile to every adult, eliminating non-POPs via "Mendelian exclusion" as described next. A brief guide to terminology can be found at the start of the Appendix, which contains a more detailed description of the operational aspects of genotyping.

Every animal has two alleles at each locus, though the two may by chance be the same; one is inherited from each parent. Therefore, a POP must share at least one allele at every locus. If there are one or more loci at which the pair do not share an allele, then the pair is not a POP. Although two non-POP individuals could by chance share an allele at every locus compared, the probability is very low if the number of loci examined is large and the loci are individually highly variable, so that no one allele is particularly common. Therefore, the most basic and most rigid exclusion principle is: a pair is treated as a POP if and only if the two animals have at least one allele in common at all loci.

This project relies on the number of POPs actually identified being close to the true number of POPs in our samples. There are two possible issues. The first is false-positives: an unrelated pair might happen to share an allele at every locus just by chance, and thus look like a POP. This probability can be assessed in advance from the allele frequencies, and this step is essential in determining whether enough loci are being used. Not all loci are successfully scored for all fish, so some comparisons will involve a lot fewer than the theoretical maximum of 25 loci in our study, and those comparisons will have a substantial false-positive probability. By excluding those "weak" comparisons, we can control the overall false-positive rate so that the expected number of false positives is negligible compared to the number of true positives[2].

The second possible issue is false-negatives, whereby a POP *appears* not to share an allele at one or more loci. This could arise through mutation, but only very rarely; published estimates of mutation rate for the kind of loci that we used are of the order of 10^{-4} per generation, so with about 25 loci in our comparisons well under 1% of true POPs would be affected by any mutations. A more likely cause of false-negatives is scoring error, whereby the true alleles are incorrectly recorded. Scoring error rates are highly variable between studies (and to some extent between loci within a study), depending on the quality of the DNA itself (i.e. tissue

[2]When we originally planned the study, our intention was to use fewer loci in the first pass (about 15), but to have a suite of "backup" loci that could be used to double-check possible POPs. However, both on economic grounds and more importantly to minimize the chance of processing mixups whereby the wrong animal gets scored, we decided to use many more loci (about 25) and to deploy them all on the first pass.

preservation), how carefully the loci are chosen, how carefully protocols are followed, and how much checking is done. Careful checking can detect and eliminate large-scale scoring errors involving many fish at once (see 6.5). However, a different approach is required for small-scale errors at the level of single loci on single specimens.

Because there are so many different possible causes of scoring error, false-negative rates cannot be predicted in advance (unlike false-positive rates), and can only be inferred after the fact. This is usually done by re-scoring individuals to see how often the scores change. However, depending on the details there may be a possibility of making the same mistake twice, so re-scoring may underestimate the scoring error rate. With our POP-oriented study, we can use a more direct and robust approach; we are using so many loci that the chance of two non-POPs sharing an allele at all-but-one of (say) 25 loci is negligible, and consequently any pairs that *seem* to share alleles at 24 of 25 loci with a mismatch at the 25th are highly likely to be false-negatives arising from scoring error[3]. The proportion of such cases compared to unambiguous true POPs (where all loci share an allele) can be used to estimate the overall false-negative rate.

2.1 False positives?

Barring errors, a POP must have at least one allele in common at every locus, so if a pair is unrelated we will eventually be able to rule it out as a POP by finding a locus that does not share an allele, provided that we look at enough loci. We have scored 25 loci[4] overall, but not all loci are scored for every fish, so some pairwise comparisons involve many fewer loci. If too few loci are used in a comparison between unrelated fish, there is a substantial probability that all the loci will share an allele just by chance. We therefore need to do some filtering, to exclude comparisons that are too likely to give a false positive. Table 2 shows what happens if we *don't* do any filtering. True POPs— plus false POPs, which just happen by chance to share an allele at every locus compared— are in the leftmost column "F0", i.e. with zero loci compared that do not share an allele. False POPs are obvious in the top-left of the table, where very few loci are being compared.

Note that the Table includes a small proportion of (i) impossible and (ii) useless comparisons, where the adult was (i) caught in a year before the juvenile was born, or (ii) caught in the same year. Type (ii) comparisons are biologically possible, but it's not helpful to include same-year comparisons in abundance estimation, because in the year of its capture an adult will not achieve its normal annual reproductive output. All such comparisons have been removed in subsequent summaries and results.

[3]Note that scoring errors do not increase the false-positive probability— there is no reason to think that an error in scoring one fish will either increase or decrease the probability of it sharing an "allele" with another unrelated fish.

[4]Plus another two that showed occasional anomalies, and were therefore omitted from routine pairwise comparisons, but were used in checking ambiguous possible-POPS.

Table 2: All comparisons, broken down by #loci compared and #loci inconsistent with POPhood. Hash (#) means "number of", dot means zero, plusses mean too big to fit.

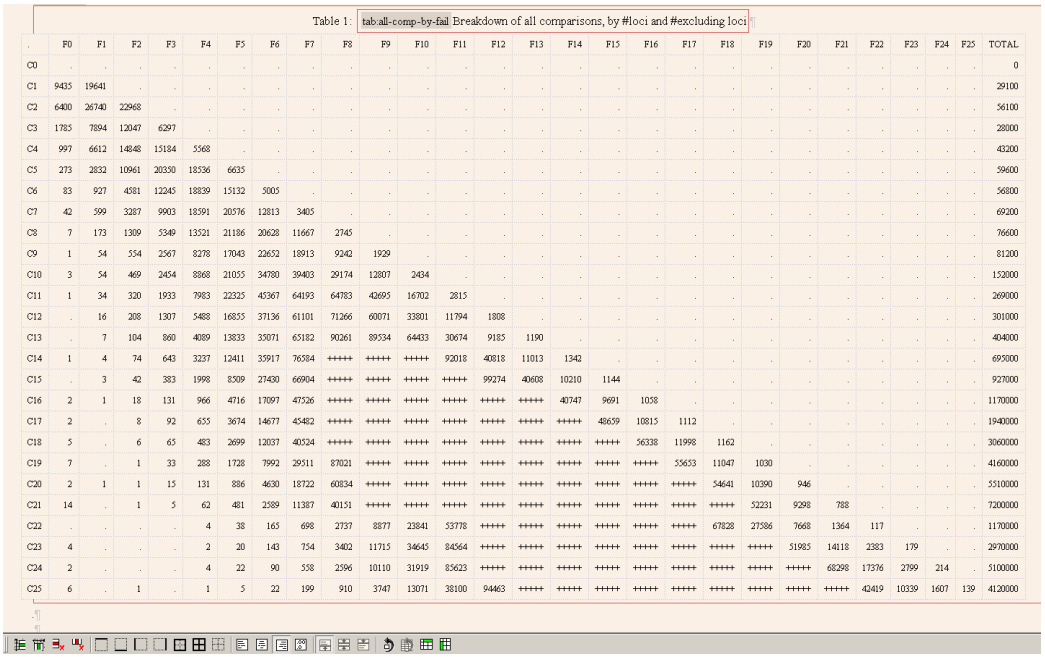


Table 1: tab:all-comp-by-fail Breakdown of all comparisons, by #loci and #excluding loci

.	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	TOTAL
C0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0
C1	9435	19641	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	29100
C2	6400	26740	22968	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	56100
C3	1785	7894	12047	6297	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	28000
C4	997	6612	14848	15184	5568	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	43200
C5	273	2832	10961	20350	18536	6635	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	59600
C6	83	927	4581	12245	18839	15132	5005	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	56800
C7	42	599	3287	9903	18591	20576	12813	3405	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	69200
C8	7	173	1309	5349	13521	21186	20628	11667	2745	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	76600
C9	1	54	554	2567	8278	17043	22652	18913	9242	1929	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	81200
C10	3	54	469	2454	8868	21055	34780	39403	29174	12807	2434	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	152000
C11	1	34	320	1933	7983	22325	45367	64193	64783	42695	16702	2815	.	.	.	.	.	.	.	.	.	.	.	.	.	.	269000
C12	.	16	208	1307	5488	16855	37136	61101	71266	60071	33801	11794	1808	.	.	.	.	.	.	.	.	.	.	.	.	.	301000
C13	.	7	104	860	4089	13833	35071	65182	90261	89534	64433	30674	9185	1190	.	.	.	.	.	.	.	.	.	.	.	.	404000
C14	1	4	74	643	3237	12411	35917	76584	+++++	+++++	+++++	92018	40818	11013	1342	.	.	.	.	.	.	.	.	.	.	.	695000
C15	.	3	42	383	1998	8509	27430	66904	+++++	+++++	+++++	+++++	99274	40608	10210	1144	.	.	.	.	.	.	.	.	.	.	927000
C16	2	1	18	131	966	4716	17097	47526	+++++	+++++	+++++	+++++	+++++	+++++	40747	9691	1058	.	.	.	.	.	.	.	.	.	1170000
C17	2	.	8	92	655	3674	14677	45482	+++++	+++++	+++++	+++++	+++++	+++++	+++++	48659	10815	1112	.	.	.	.	.	.	.	.	1940000
C18	5	.	6	65	483	2699	12037	40524	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	56338	11998	1162	.	.	.	.	.	.	.	3060000
C19	7	.	1	33	288	1728	7992	29511	87021	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	55653	11047	1030	.	.	.	.	.	.	4160000
C20	2	1	1	15	131	886	4630	18722	60834	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	54641	10390	946	.	.	.	.	.	5510000
C21	14	.	1	5	62	481	2589	11387	40151	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	52231	9298	788	.	.	.	.	7200000
C22	.	.	.	.	4	38	165	698	2737	8877	23841	53778	+++++	+++++	+++++	+++++	+++++	+++++	67828	27586	7668	1364	117	.	.	.	1170000
C23	4	.	.	.	2	20	143	754	3402	11715	34645	84564	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	51985	14118	2383	179	.	.	2970000
C24	2	.	.	.	4	22	90	558	2596	10110	31919	85623	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	68298	17376	2799	214	.	5100000
C25	6	.	1	.	1	5	22	199	910	3747	13071	38100	94463	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++	42419	10339	1607	139	4120000

In order to filter out false POPs, we first compute in advance for each possible pair a False-Positive Probability (i.e. the probability that the two animals will share an allele at every locus compared, even if unrelated) based on which loci were scored successfully for *both* fish in the pair, and without looking at the actual genotypes that resulted. We then sort these FPP in ascending order, and find the cutoff such that the *total* FPP from all (sorted) pairs below the cutoff is below some pre-specified threshold T. Only those pairs whose FPP falls below the cutoff are subsequently checked for POPhood, the remainder being deemed too ambiguous. Note that not testing POPhood of an ambiguous pair does not cause any bias in the proportion of included comparisons that yield POPs, because the FPP check is done *before* testing for POPhood, and is unrelated to whether the pair really is a POP or not. The threshold T is by definition equal to the total expected number of false POPs, so we choose it to be a small fraction of the number of true POPs, of which we have a shrewd idea of by this stage. For this report, we have set the threshold at 0.35, below 1% of the number of POPs actually found. Because false POPs lead to a proportional negative bias in abundance estimates, the upshot is that we have kept such bias to under 1%.

The resulting set of filtered comparisons is shown in Table 3. At least 11 loci must be compared to get an FPP above the cutoff, and less than 100 11-locus pairs squeeze in; these occur where the 11 happened to be amongst the most powerful[5] of the 25 loci used for the table. On average, the loci used have about a 0.65 chance of *not* sharing an allele by chance, and the table shows very clearly how (near-)binomial probabilities work; from right to left, the numbers in the columns decline rapidly, except for the leftmost column where true POPs appear.

Importantly, in the bottom-left-hand-corner, the Table shows "clear blue water" between the best-matching unrelated pairs (i.e. with fewest loci that do not share an allele) and the true POPs. The separation is less obvious in the rows above say C16, but by looking at how fast the numbers in each row decline from right to left through the F4-F3-F2 columns, it is clear

[5] I.e. genetically more diverse, and being least likely to share an allele by chance

that very few unrelated pairs would have made it into the F0 column. And of course this is what the FPP calculations suggest: given the filtering rule, we would only expect 0.35 spurious POPs in the F0 column. Given that expectation, it is certainly possible that one ($p = 0.25$) or maybe even two ($p = 0.05$) false POPs could have crept in, but very unlikely that false POPs make up an appreciable proportion of the total of 45.

Table 3: Number of *usable* pairwise comparisons, by #loci and #excluding loci. Comparisons are *not usable* if the adult was caught in or before the year of juvenile birth, and/or the false-positive probability was too high (see text). Columns 8-21 omitted for brevity.

.	F0	F1	F2	F3	F4	F5	F6	F7	>	F22	F23	F24	F25	TOTAL
C11	.	.	.	.	1	4	5	21	>	.	.	.	.	84
C12	.	.	5	42	340	1345	4019	9114	>	.	.	.	.	57,000
C13	.	1	16	151	887	3420	9900	20482	>	.	.	.	.	143,000
C14	1	4	61	587	2876	11277	32947	70962	>	.	.	.	.	652,000
C15	.	3	42	375	1962	8411	27165	66386	>	.	.	.	.	923,000
C16	2	1	18	131	966	4716	17097	47526	>	.	.	.	.	1,170,000
C17	2	.	8	92	655	3674	14677	45482	>	.	.	.	.	1,942,000
C18	5	.	6	65	483	2699	12037	40524	>	.	.	.	.	3,063,000
C19	7	.	1	33	288	1728	7992	29511	>	.	.	.	.	4,158,000
C20	2	1	1	15	131	886	4630	18722	>	.	.	.	.	5,512,000
C21	14	.	1	5	62	481	2589	11387	>	.	.	.	.	7,197,000
C22	.	.	.	.	4	38	165	698	>	117	.	.	.	1,170,000
C23	4	.	.	.	2	20	143	754	>	2383	179	.	.	2,966,000
C24	2	.	.	.	4	22	90	558	>	17376	2799	214	.	5,097,000
C25	6	.	1	.	1	5	22	199	>	42419	10339	1607	139	4,123,000
SUM	45													38,180,182

It is also possible to compute an "expected" version of Table 3, assuming there are no true POPs. That is: for each comparison, taking into account which loci were used, we can compute the probability that there were 0, 1, 2, ... mismatching loci if the pair was truly unrelated. By summing the probability of, say, 1 mismatching loci over all comparisons with, say, 11 loci, we can compute the expected value of the (C11, F1) element corresponding to Table 3. The left-hand columns of the result are shown in Table 4, after filtering out the same comparisons as in Table 3. By definition, the row-totals would be the same as in Table 3; the question is how close the column totals are, as shown in the bottom two rows of Table 4. And they are very close, except of course for the F0 column where we are seeing true POPs. This is good; the laws of probability seem to be working well today. The close correspondence between observed and expected totals for F1/F2/F3 suggests that the calculations leading to 0.35 expected false POPs are sound; of course, the *actual* number cannot be exactly 0.35, but it is most likely 0, and most unlikely to be more than 2.

Table 4: *Expected* number of comparisons with a given number of mismatching loci, given the loci actually used in each comparison, and assuming no true POPs. The TOT OBS row at the bottom is taken from Table 3.

	F0	F1	F2	F3
C11	.	.	0.02	0.17
C12	0.02	0.63	9.43	82.46
C13	0.04	1.27	17.91	149.85
C14	0.15	4.50	60.38	491.97
C15	0.08	2.56	36.35	315.94
C16	0.03	0.98	15.15	144.87
C17	0.02	0.55	9.16	94.78
C18	0.01	0.30	5.32	58.88
C19	.	0.15	2.75	32.39
C20	.	0.05	1.12	14.42
C21	.	0.02	0.47	6.48
C22	.	.	0.02	0.30
C23	.	.	0.02	0.25
C24	.	.	0.01	0.13
C25	.	.	.	0.03
TOT EXP	0.35	11	158	1392
TOT OBS	45	10	160	1496

Using a cutoff to exclude ambiguous comparisons does entail a bias-variance trade-off, because some true POPs may have been overlooked in the excluded comparisons, and any reduction in the overall number of POPs found will increase the uncertainty in our final estimates. However, given the threshold we used, it is only when the number of loci compared is 14 or less that substantial numbers of comparisons are excluded (from comparison of Figure 2 and Table 3), and overall only about 5% of comparisons are excluded. Thus we have managed to achieve less than a 1% bias while only incurring a $\sqrt{5} \approx 2\%$ increase in standard error compared to what we would have gotten from "perfect" genotyping (where every pairwise comparison is usable). This reflects very well on the tissue quality, the processsing, and the selection of powerful, reliable loci.

Of the 45 POPs found, it is interesting that 9 included one locus where the two animals were scored as different homozygotes (one AA and the other BB). We had deliberately relaxed the exclusion rule to permit this situation, in case of "heritable nulls" (see Appendix, section 6.3), and there was no ambiguity about the POP status of these pairs based on the remaining loci[6]. In all but one of the 9 cases the apparent mismatch occurred in one or other of the two loci which exhibited substantial excess homozygosity (D569 and D573; see Appendix), consistent with the "heritable null" possibility.

Note also that close-kin relationships at the level of uncles-and-nephews, while possibly as common as POPs in reality, are not going to lead to false POPs in this study. Between an uncle & nephew, only 50% of loci will share an allele by descent anyway, so with these loci the overall chance of sharing an allele is about $1/2 * 1 + 1/2 * (1 - 0.65) = 0.68$ (compared to about 0.35 for an unrelated pair), and the chance of getting say 20 loci all sharing an allele through chance is about 0.0004— so there would need to be about 2000 uncle-nephew-level pairs to generate a single false POP.

[6]Including additional checks at the extra one or two loci which were not normally used in mass-screening for POPs

2.1.1 Cases where no POPs should be found

As an exercise, we can repeat Table 3 just comparing juveniles with themselves, where true POPs are impossible; see Table 5. The expected total in the F0 column is again 0.35; this time, the observed total is 1 (in C13/F0, so towards the lower end of the number of loci compared) which as noted earlier has about a 25% probability and gives no indication that the false-positive calculations are failing.

We can also compare all adults with all other adults (not shown). This time, POPs are actually possible, albeit likely rare because of the time required to reach maturity— see later discussion. There is indeed one possible POP (C18/F0; unlikely to be by chance, given 18 loci used), and it is plausible biologically. The female "parent" was aged 24 when caught in 2007, and the female "offspring" was 177cm (not aged, but any age from 12 up is plausible, given other length-at-age data) when caught in 2009; this gives plenty of scope for the "parent" to have been mature when the offspring was born.

Table 5: Comparison of juveniles to themselves.

	F0	F1	F2	F3	F4	F5	F6	F7
C11	.	.	.	.	.	.	6	9
C12	.	.	8	45	329	1404	4611	10109
C13	1	.	7	63	399	1574	4935	10697
C14	.	1	36	257	1335	5386	15948	35522
C15	.	1	15	153	872	3307	10661	25493
C16	.	1	6	42	304	1465	5341	14986
C17	.	.	2	31	232	1236	4744	14436
C18	.	.	2	26	169	1010	4318	14160
C19	.	.	6	21	144	888	4136	14761
C20	.	.	1	14	85	603	3025	12153
C21	.	.	1	.	37	275	1644	7109
C22	.	.	.	.	.	22	97	524
C23	.	.	.	.	6	14	98	524
C24	.	.	.	1	2	8	69	403
C25	.	.	1	.	1	6	23	115

2.2 False negatives?

What about accidentally excluding true POPs? That can only happen if there is genotyping error[7]. Large-scale errors involving multiple loci at once would be (and were) detected and fixed by our QC procedures described in the Appendix, so the concern here is about small-scale errors at a single locus and specimen. If such errors lead often to false-negative POPs, these should show up low down in the F1 column of Table 3, as near-POPs that apparently fail to match at one locus (false-negatives at multiple loci being correspondingly rarer). That is not what is seen; rows C17 down have only one entry in F1, compared with an expected total of 1.1 from Table 4.

Prior to producing Table 3, we independently re-scored[8] all the apparent true POPs in F0, all the F1s, and F2s in the rows from C17 down. The original version of Table 3 had 44 rather than 45 POPs; the re-scoring moved one pair from C15/F1 to C17/F0 (changing one existing

[7] Or mutation, but with say ~50 POPs and ~20 loci each, and mutation rates thought to be about 10^{-4} per generation, mutation is unlikely to have happened amongst our POPs.

[8] "Re-scored" means: we re-examined all the peaks and came up with new scores, but did not re-do any of the chemistry.

score, and scoring 2 more loci originally deemed unscorable). The lower left-hand corner of the Table (apart from true POPs in F0) was still empty even without rescoring. Although rescoring changed only about 1 POP, it does give some indication of scoring error rates. Across the 1400 loci that were rescored, there were 8 individual changes, plus deleting one panel of loci for one fish; four of the changes were to delete a score altogether when a locus looked dubious, and the other four were to add a second allele to a "homozygote" (a definite error). Note that all 8-9 changes in the rescoring only unearthed one false-negative (corrected in Table 3), so the *effective* false-negative rate for POP purposes seems to be well under 0.5%. It would also be possible to produce per-locus estimates of scoring error rate based on the partial re-runs and re-used control fish in our QC procedures.

The most important line of evidence to suggest that false negatives from individual scoring errors are not a serious problem, though, remains the absence of entries in the lower left-hand corner of Table 3. Appendix 2 presents a formal statistical approach to estimating false-negative rates by comparing Tables 3 and 4; the point estimate of the overall number of remaining false-negatives is in the range 1-2, and the upper 95% CI in the range 2-3. In any event, false negatives must be at most a small proportion of the 45 POPs.

2.3 Summary of genetic results

Extensive QC procedures were used to ensure consistent and reliable scoring throughout the project. In all, we conducted about 40,000,000 pairwise comparisons to look for POPs. A few pairs had to be excluded because they had too few scored loci to reliably screen out unrelated pseudo-POPs. However, because of the number and quality of loci used, we were able to choose a cutoff for exclusion that implies very little bias (i.e. unlikely to unearth false POPs) while incurring very little penalty in variance (i.e. using nearly all the comparisons). QC protocols were devised to catch large-scale mixups. With respect to small-scale (individual-level) scoring errors, the error rate is too low to cause a substantial proportion of true POPs to be overlooked. In all, we found 45 POPs in about 38,000,000 usable comparisons.

3 Qualititative findings about the POPs

3.1 Sex, age and size of parents vs general adults

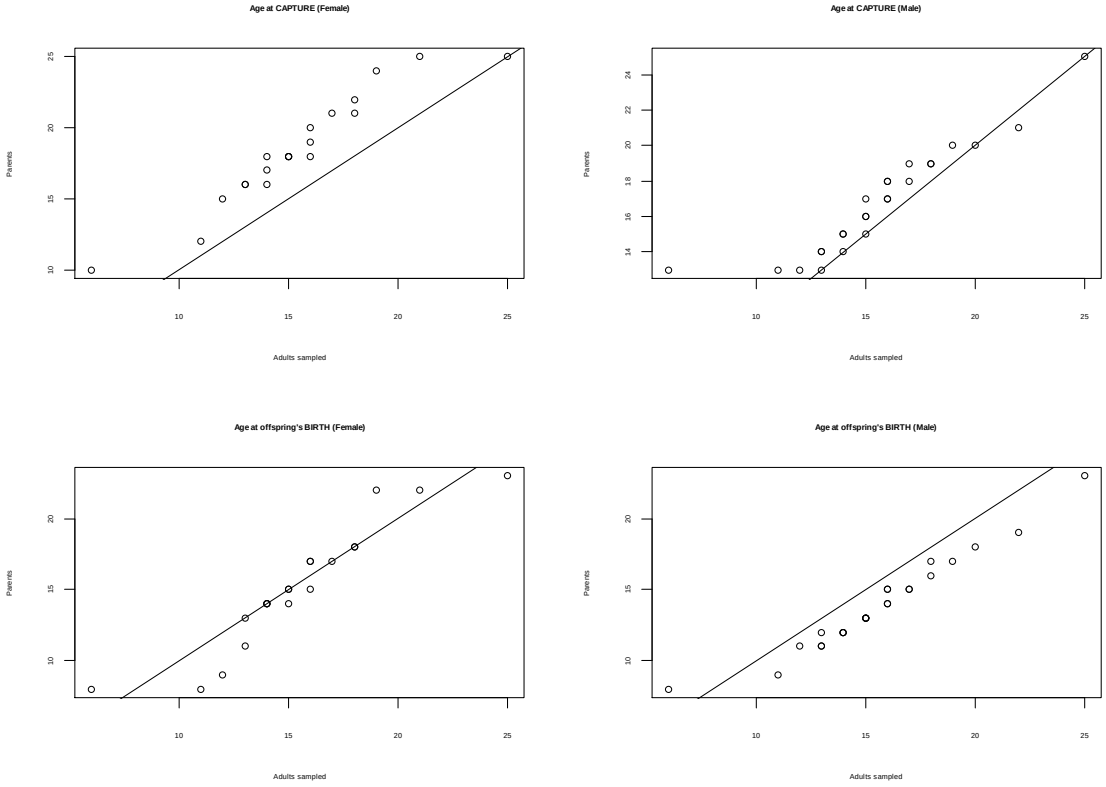
Of the 45 POPs, 20 were female and 25 male. All adults in POPs have now been aged; about 1/3 were aged under the Indonesian/Australian ageing program, and the remainder were aged specifically for this project after being identified through genotyping. On average, parents *at capture* are somewhat older (and bigger; not shown) than typical captured adults of the same sex. However, this comparison is not "fair" because the parents have had the opportunity to grow during the interval between juvenile birth and adult capture, which in this study is on average about $3\frac{1}{2}$ years.

Since juvenile age is known (3 in almost all cases), it is easy to back-calculate parental age when the offspring was born. The youngest successful spawners were aged 8, for both sexes[9]. When back-calculated parental age is used instead of age of capture, the difference between parental and typical adult age distribution disappears for females, and actually reverses for males; But it is important to realize that this back-calculated comparison is also not "fair". Adults are subject to selectivity bias in favour of bigger/older fish, and the selectivity pattern

[9] An earlier version of this document reported an anomalous 6-year-old spawner. On re-checking the otoliths, it turned out that the two otoliths for this fish actually came from different adults, so there must have been a handling error at the collection site. Based on the correct otolith (i.e. the one whose DNA matched the main adult tissue sample), the actual spawning age turned out to be 9.

on the parents would also have changed over the interval between giving birth and being caught. Back-calculated age distributions will be skewed towards younger/smaller fish, compared to what would have been found if the same set of parents had somehow been sampled in the year of offspring birth.

Figure 3.1: QQ plots of parental age vs adult age, by sex . Points left/above the line mean parents are bigger/older. Females on left, males on right. Upper graphs show parental age at year of capture, lower graphs at year of birth.



The upshot of this rather involved argument is:

- parents at capture are older/bigger than typical adults, because they have aged/grown since giving birth;
- back-calculated parental age distribution at offspring birth is similar to typical adult age, but...
- the back-calculated distribution is biassed towards smaller/younger fish, so...
- female parents would actually be bigger than typical adults if it was possible to sample them in the birth-year.
- It's not clear whether the same would be true for males.

These phenomena can only be fully disentangled with the aid of a mini-assessment model.

Similar results are found using length rather than age, but the mini-assessment then has to be fitted beforehand, because of the need to back-calculate length.

3.2 Skip-spawning

From the small number of POPs identified in time for CCSBT 2011, there was no obvious indication of skip-spawning. However, the larger sample of POPs now available does show evidence of biennial spawning for younger fish. The test is to take each POP, and note how many years actually elapsed between juvenile birth and adult recapture, vs how many years *could* have elapsed given the POP was eventually found. For example, if the juvenile in a POP was born in 2007, then only comparisons with 2008/2009/2010 adults would be meaningful, so the probability of matching to a 2008 adult is roughly[10] equal to the proportion of adults checked in 2008 relative to those checked in 2008+2009+2010. Table 6 shows the results, split by parental age at offspring's birth; for younger parents, almost all observed gaps are even-numbered, but not for older parents. The pattern is not sex-specific.

Table 6: Distribution of gap between Juvenile-Birth-Year and Adult-Capture-Year, for young & old parents. Dot means zero. Right-hand table is condensed to odd/even gaps.

Age	Gap (years) ->	1	2	3	4	5	6	7
8-12	Obs	1	6	.	2	.	4	.
	Exp	1.6	2.3	2.7	2.6	1.9	1.1	0.6
13-25	Obs	7	5	10	7	2	1	.
	Exp	4.3	6.8	7.4	6.2	4.5	1.8	1.0

Age	Gap->	Even	Odd
8-12	Obs	12	1
	Exp	6.1	6.9
13-25	Obs	13	19
	Exp	14.9	17.1

Any errors in ageing would obscure patterns such as seen here. Although the sample size is not huge, the difference for younger adults is significant at 1%.

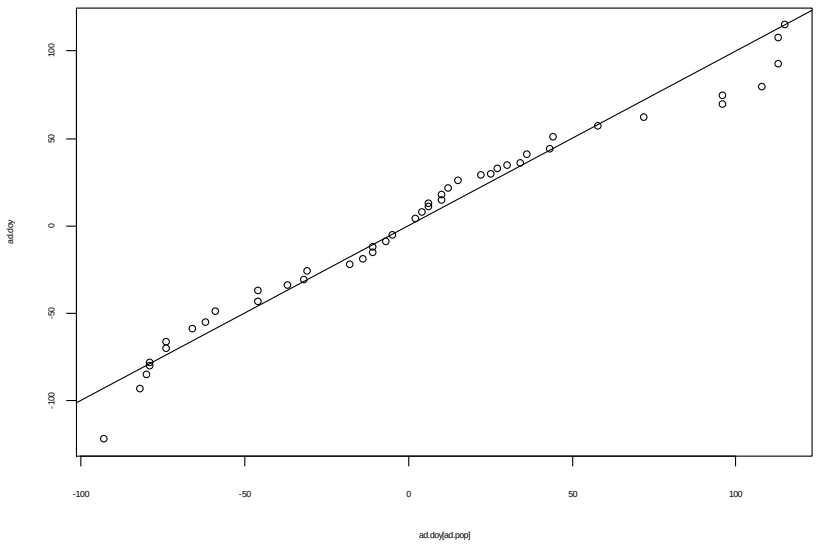
Skip-spawning is not a particular problem for this close-kin study because the study covers many years and the even/odd effect should largely wash out; the general effect of smaller fish being less present is already allowed for in the mini-assessment, because average spawning-ground residence-time (including the probability of not being on the spawning grounds at all) gets estimated as a function of length and sex. However, in a more perfect world, probabilistic size/age-based skip spawning would be allowed for in the mini-assessment.

3.3 Timing in spawning season

Parents of GAB juveniles have the same distribution of capture date within season as do "average adults" (Figure 3.2). Thus there is no evidence of "temporal stock structure" in a way that might lead the abundance estimates to be biassed (eg we *might* have seen that parents of GAB juveniles always spawn early, and we might *not* have had equal coverage through the Indonesian fishing season). Breaking down by sex does not reveal anything either.

[10] Calculations are approximate: e.g. the "expected" rows do not account for growth or mortality, but should reflect any even/odd pattern OK.

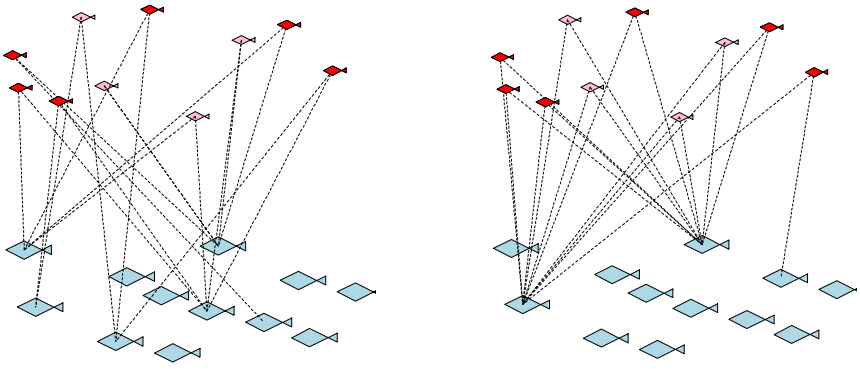
Figure 3.2: QQ plot of day-of-year of capture of Parents (X) vs Adults-in-general (Y)



3.4 Incidence of (half-)siblings among the POP juveniles

There are none. In other words, none of the POP adults match to more than one juvenile. That is a good thing, because if (half-)sibs are common among the *sampled* juveniles, then the pairwise comparisons become non-independent. Figure 3.3 shows what might happen; if there are many (half-)sibs in the juvenile sample, then the number of links to parents remains the same so the abundance estimate is still unbiassed (noting that an adult can "count" in more than one POP), but its variance would increase because the number of POPs actually found would depend critically on whether the "super-parents" were caught.

Figure 3.3: Cartoon depicting the impact that reproductive variability would have. Small fish are juveniles, red ones are sampled.



A preliminary check in 2010 just among juveniles indicated that (half-)sibs could not be *very* common (a critical decision point for the project), and the 7 POPs found in 2010 contained no sibs or half-sibs. Having found none in this much larger set of POPs, we can maybe conclude that (half-)sibs are rare enough *among our juvenile samples* for their effects on variance to be ignored. This is not to say that (half-)sibs are at all rare among *all* 3-year-olds, but simply that our juvenile samples are a very small fraction of the total, and are well-enough-mixed to

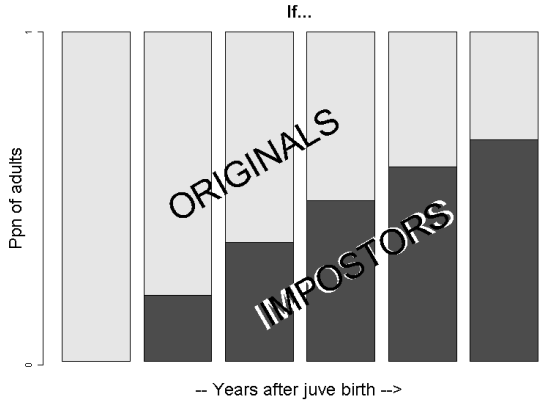
make sib-pairs rare. As an academic exercise, it will at some point be interesting to re-run the juvenile-only sib check with the greatly expanded set of loci now available.

4 Mini-assessment

4.1 What effects need to be considered?

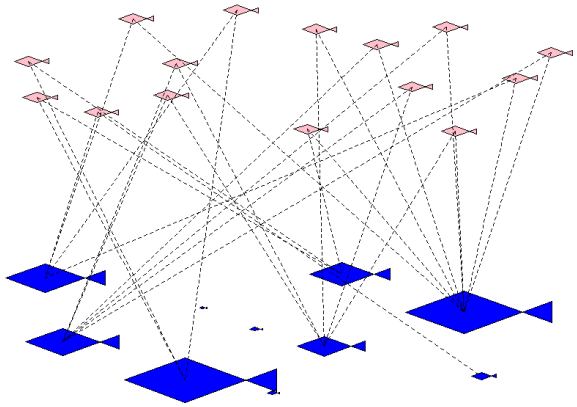
There are two main reasons why the $2m^2/P$ "cartoon" estimator ("cartoon" in the sense of Figure 3.3, for example) would be seriously misleading for SBT. The first is that we cannot do comparisons only against the "parental cohort-group" of each offspring, i.e. the group of adults that were alive at its birth. Figure 4.1 illustrates the main point; if survival rates are the same for all adults, then the cartoon estimator would still be valid even with time lags, *provided* we could restrict comparisons to the light-grey parental cohort-group. But we cannot do so, because (i) we do not know the age of all adults sampled, (ii) maturity is not knife-edge so there is no absolute definition of the parental cohort-group, and (iii) maturity is quite likely length- rather than age-driven. If we are forced instead to sample adults from say the entire 4th column of Figure 4.1, after a 3-year gap, then a high proportion of comparisons will be with "impostor" adults that could not have been parents, and the cartoon estimator would be biassed high.

Figure 4.1: Dilution of original parent-cohort-group by incoming recruitment



The second, linked, reason is that adult sampling is strongly selective towards large/old fish, which are also likely to have been more fecund (even allowing for a 3-year time lag). Because they are more fecund, they have more "tags" per capita (i.e. juveniles that they are parents of), and each tag is more likely to be "recaptured" (i.e. the adult is more likely to be caught) because of selectivity in favour of larger adults. This is the close-kin analogue of "heterogeneity in capture probability", a well-known issue in mark-recapture abundance estimation. Figure 4.2 shows the cartoon version. The upshot for the naive $2m^2/P$ estimator would be that each comparison is more likely to yield a POP than would a comparison with a randomly-chosen adult.

Figure 4.2: Big tuna are more fecund *and* more likely to be caught



Both effects concern not the number of POPs actually found, but rather the difficulty of working out how many comparisons are "relevant" for abundance estimation. The two effects act in opposite directions; the time-lag dilution means that some comparisons are invalid and thus less effective than "random" comparisons, whereas the selectivity-fecundity correlation means that the valid comparisons are more effective than "random" comparisons. The time-lag dilution is also mitigated by growth, since the surviving "original" adults after 3-4 years will be bigger than the "impostors" and thus more likely to be caught. However, there is no particular reason to assume the effects will cancel out, since the time-lag effect is driven primarily by the length of the study whereas the selectivity-fecundity effect is determined by the nature of the fishery and the growth curve. And the effects can be quite large; with an adult survival of say 0.8, after 3 years only 50% of the original adults are still alive to be sampled and the impostors will (in equilibrium, and neglecting selectivity and growth) be involved in about 50% of the comparisons. To deal with these issues properly, we need a mini-assessment.

One further issue arises from of the extended timespan of this study, which spans juvenile birth-years from 2002 and adult capture-years to 2010, as well as the initial age structure of the adults in 2002 which was determined by even earlier events. The 1990s and 2000s have been eventful decades for SBT, and it may be such that steady-state assumptions are simply not viable.

A proper close-kin abundance estimate for SBT therefore has to deal with survival, selectivity, fecundity, and growth, and perhaps with changes in abundance over time. The requisite data come from the length and age-at-length samples from Indonesia, plus fecundity studies explained below. While not strictly "fishery-independent", length and age data are not subject to the same problems as CPUE or total catch. It also makes sense to split the analysis by sex: the cartoon applies equally well if applied to males and females separately, where the chance of a POP comparing to a male adult is $1/N_{\text{male}}$ not $2/\left(N_{\text{male}}+N_{\text{female}}\right)$, and C is split into $\left(C_{\text{male}}, C_{\text{female}}\right)$.

4.1.1 Residence time, selectivity, and fecundity

The tropical waters off Indonesia are really no place for an adult SBT, an animal that is adapted superbly for much cooler temperate waters. Adults arrive on the spawning grounds fat, and leave thin. Of course, the longer they can stay on the grounds, the more chances to spawn they will have, so it seems reasonable to suppose that they will put up with Indonesian conditions for as long as their bodies let them. The key for disentangling the effects of fecundity, survival, and selectivity, is average **residence time** on the spawning grounds, as a function of length. A cursory glance at length distributions from Indonesia shows that few fish under 150cm, and none under 130cm, are caught on the spawning grounds, so there is obviously some link to

length. As per the skip-spawning discussion, "average residence time" already factors in the probability that a fish won't be there at all in any given year. Our mini-assessment specifically assumes that, given length and sex:

- Selectivity $\propto$ residence time
- Annual reproductive output $\propto$ residence time $\times$ daily reproductive output

Except as specifically noted later, we assume that length and sex are the driving influences behind the behaviour of adult SBT, rather than age.

Of course, there could be other "second-order" phenomena which slightly change the above relationships (e.g. different depth distributions by size, and thus different exposure to hooks; different egg *quality* with parental size; etc etc) but these seem likely to be small compared to the dominant effect of residence time. For the rest of this document, it may be helpful to think about selectivity and residence time as directly equivalent, at least within each sex.

We have no direct data on residence time as a function of length, so the relationship needs to be estimated indirectly from data. Independent data on residence time and depth distribution as a function of length, from archival tags placed on big fish, would be extremely useful: both in tightening up parameter estimates in our existing model, and in assessing whether the effects that we hope are "second-order" really are.

4.1.2 Fecundity analyses: daily reproductive output

The canonical reference for SBT (female) spawning biology and fecundity is a study from the early 2000s by Davis et al.[11]. In summary, female SBT while on the spawning grounds have an on-off cycle, consisting of several days of consecutive daily spawning (one spawning event per 24 hours), followed by several days of rest while more eggs are built up. This on-off cycle may be repeated several times. As soon as the final spawning cycle is complete, they leave. The mass of eggs released per daily spawning event can be estimated from the change in gonad weight between just-about-to-spawn and just-after-spawning fish; it is approximately proportional to length to the power 2.47. The average duration of each part of the cycle (and thus the proportion of days on the spawning grounds when spawning actually occurs) can also be estimated as a function of body length using histological data, because the first day of a spawning sequence can be distinguished from the other days, and similarly for a resting sequence. However, the number of cycles per season is completely unknown, and is obviously set by the residence time.

To summarize, the factors involved in daily reproductive output are:

- reduction in gonad weight per spawning event
- duration of consecutive spawning day sequences
- duration of consecutive resting day sequences

A reasonable amount of data is available for all three of these, and the relationship to length can be estimated from fitting three GLMs. (This was already done for the first two factors in Davis et al., and the third factor was addressed during this study.) For now, we have treated the parameter estimates as exact in the rest of the mini-assessment.

We have no comparable data for males, nor on the extent to which male abundance actually influences the number of fertilized eggs per year.

[11] T. Davis, J. Farley, M. Bravington, R. Andamari (2003): *Size at first maturity and recruitment into egg production of southern bluefin tuna* FRDC project 1999:106

4.2 Indonesian length, sex, and age data

A substantial proportion of the Indonesian SBT catch is sampled as it passes through the main landing port of Benoa. Length (to the centimetre) and sex are always recorded, and nowadays otoliths are always extracted, although only a length-stratified subset (500 per year in the recent past) are read. Between 900 and 1700 animals were measured per year between 2002 and 2010. Thus the data can be seen as

1. Random samples of length and sex from the entire adult catch
2. Random samples of age, given length and sex.

Even without the POP data, it is possible to do some steady-state analysis of the age/length/sex data (though it is obviously impossible to estimate absolute abundance), but it is impossible to completely separate selectivity (as a function of length) from average adult survival rate. When the survival rate is very high (e.g. 0.9) or very low (e.g. 0.5) it does become impossible to match the observed length-frequency distributions except by invoking a ludicrous selectivity function, but in the absence of other data reasonable fits to the age and length data can be obtained across a wide range of survival rates.

Fortunately, the POPs can help estimate survival rate, in addition to absolute abundance. The typical gap between offspring birth and adult capture— assuming that the adult is in fact captured subsequently, i.e. that the pair is an identified POP— is related to survival. If survival rates are low, very few parents will survive to be caught say 7 years later (the maximum gap possible in this study), so most of the POPs that are found will be separated by just one or two years. Growth and residence time need to be properly accounted for too, but the intuitive basis should be clear. The close-kin data thus has three vital roles: the *number* of POPs (given the number of comparison) essentially sets the scaling of absolute abundance, the age and length distribution *within* the POPs informs on selectivity/fecundity, and the distribution of time-gaps *within* the POPs essentially determines survival.

4.3 Model structure

The model keeps track of numbers by age and sex; each year, each fish either gets one year older or dies. However, most phenomena are driven by length, which is assumed to have a fixed distribution at age. Each fish has its own personal L_∞, drawn from a sex-dependent log-Normal prior whose mean and variance at age are fixed, while the other von Bertalanffy parameters are the same within each sex. A plus-group is used for ages 25 up, and a minimum "recruitment" age for possible spawning also needs to be set (currently 8). There is also a plus-group for length (200cm) and, unusually for stock assessments, a sort of "minus-group" as well, currently set to 150cm. Experience with fitting just to age and length data showed that trying to extend the fit to the small proportion of adults below 150cm gave poor results, in that this small "tail" started to "wag the dog" and distort the fit elsewhere. The focus of this study is spawners, which are mostly 160cm and up, so it is more important to get a good fit there than to squeeze a last drop of misinformation out of very small adults. However, it is necessary to somehow keep track of the small spawning contribution of fish in the minus-group, and accordingly there is some tedious book-keeping code in the model.

Most of the likelihood is quite standard; multinomial distributions for length-sex frequency data, and for age given length and sex. The effective sample sizes of the length and age data were capped at 300 per year, to avoid these data swamping the information from the POPs. The novel term is the contribution of the POPs. For each comparison made between a juvenile j and an adult i of sex (gender) g_i, the outcome (POP or not) is a Bernoulli random variable

with probability given by

$$\mathbb{P}[j \sim i] = \frac{\text{expected ARO from } i \text{ in year of } j\text{'s birth}}{\text{total ARO from adults of sex } g_i \text{ in that year}}$$

where ARO is Annual Reproductive Output, i.e. daily fecundity multiplied by residence time as in section 4.1.1. This formula replaces the "$2/N$" probability in the simplest possible close-kin implementation.

To actually compute a likelihood, it is necessary to specify various terms:

- numbers-at-age in 2002, and for incoming recruitment (age 8) in 2003-2010;
- survival rate in each year and age;
- residence/length relationship;
- growth parameters;
- relation between daily RO and length *for males*.

The total number of potential parameters is colossal because of the numbers-at-age and survival terms, so of course one needs to specify them parsimoniously given the limited amount of data available. This is done using formulas (sensu R) for each bullet-point term, describing what covariates are allowed to influence it, and perhaps what functional form that influence might take. For example, we might choose to make survival constant over age and time, except for the plus-group[12]. We might also make assumptions of constant "recruitment" (at age 8) in the 2000s; and/or that numbers-at-age prior to 2000 were in equilibrium with survival; and/or that von Bertalanffy k is the same for both sexes; and/or that the slope of the residence/length relationship (but not its midpoint) is the same by sex; etc. One example is given in the Results section.

The final term— male daily reproductive output as a function of length— can *in principle* be estimated provided we are willing to assume that survival rates for males are the same as for females. Without that assumption, there is nothing to anchor the selectivity/survival/fecundity triangle for males. For females, we do not need to estimate this term because we have direct data from the fecundity studies.

The likelihood itself is coded in Pascal, with derivatives computed by an automatic differentiation toolbox similar to ADMB. The overall data-handling and fitting is done in R, calling the `nlminb()` optimizer to do the fitting. Some care was needed to avoid numerical problems in calculating the log-likelihood, and because of limited time there are still starting-value problems so that some model parametrizations can't yet be fitted. However, once a starting value has been obtained, no convergence problems were encountered, at least for the fairly parsimonious specifications (say 15 parameters) that have been tried to date.

4.4 Results

It will be apparent that an enormous number of different *versions* of the mini-assessment could be run. A full investigation is far beyond the scope of this project, and should probably be undertaken in conjuntion with other data sources, particularly to provide context on recruitment levels. A limited set of versions has been run, but the results actually shown here come from an almost-steady-state version of the model, with constant adult survival and constant recruitment

[12] In SBT as with other top-predators, it must be the case that natural mortality rate increases for old animals, since simple maths shows that the sea would otherwise just fill up with decrepit tuna.

from 2002 onwards but an age composition in 2002 that need not correspond to a steady-state prior to 2002.

Basic investigations across suggest that:

- Mean L_∞ is appreciably larger for males than females. The evidence for any difference in k or t_0 is not overwhelming, but making these two sex-linked as well does not seem to overparametrise the model. CV of length-at-age appears to be the same for both sexes.

- Residence time appears to be lower for males of a given length than for females, so we do need a sex-specific intercept in this term. However, there is not enough data to estimate any sex difference in the *slope* of the relationship. Also, introducing extra flexibility in model form beyond the logistic (asymptotic) can give nonsensical predictions for very large fish. A good choice seems to be `~sex+length`.

- There is not much information for estimating male daily reproductive output as a function of body length. We have assumed instead that male daily output is directly proportional to length (i.e. exponent of 1). There is no good reason for that particular choice, but fortunately the abundance and survival estimates seem not to be much affected by assumptions about male daily output in practice, even though it could matter in theory.

- Based on just one comparison: changing the annual effective sample size for length/age data from 300 to 900 did not much affect the abundance estimates much (i.e. by a few percent).

Having got these basic issues out of the way, the remaining questions are how to set up initial numbers-at-age and incoming recruitments. In a full steady-state model, the age distribution in year 1 (actually 2002AD in our setup) is determined by the survival rate, and the incoming recruitments thereafter are equal to the numbers at recruitment age in year 1.

Some diagnostic plots for the steady-state model are shown in Figures 4.3-4.4. These pertain to the length and sex data only, since the POP data are really too sparse for diagnostics. The length-frequency data, shown for few years only in Figure 4.3, are mostly not too bad despite the steady-state assumption, except for 2002 where the data seem completely different from other years. The fits to age-at-length are very good (not shown). However, there is a problem with the fits to sex ratio by length class (Figure 4.4): in the biggest length classes lower down the graph, where males tend to predominate thanks to their bigger asymptotic size, there is a strong decrease in proportion of females over the 2000s. This decrease is seen overall too (in the black dots), but is not apparent in the smaller lengths, where there is a rise followed by a dip. This difference in trend across length classes suggests that methodological changes in how sex is assessed are unlikely to be the cause. The underlying cause needs some further thought; it has nothing much to do with close-kin, and is a question for the OM as well as this mini-assessment.

Figure 4.3: Steady-state diagnostics: length. The unreadable parts do not need to be read; they show details of the particular model version fitted.

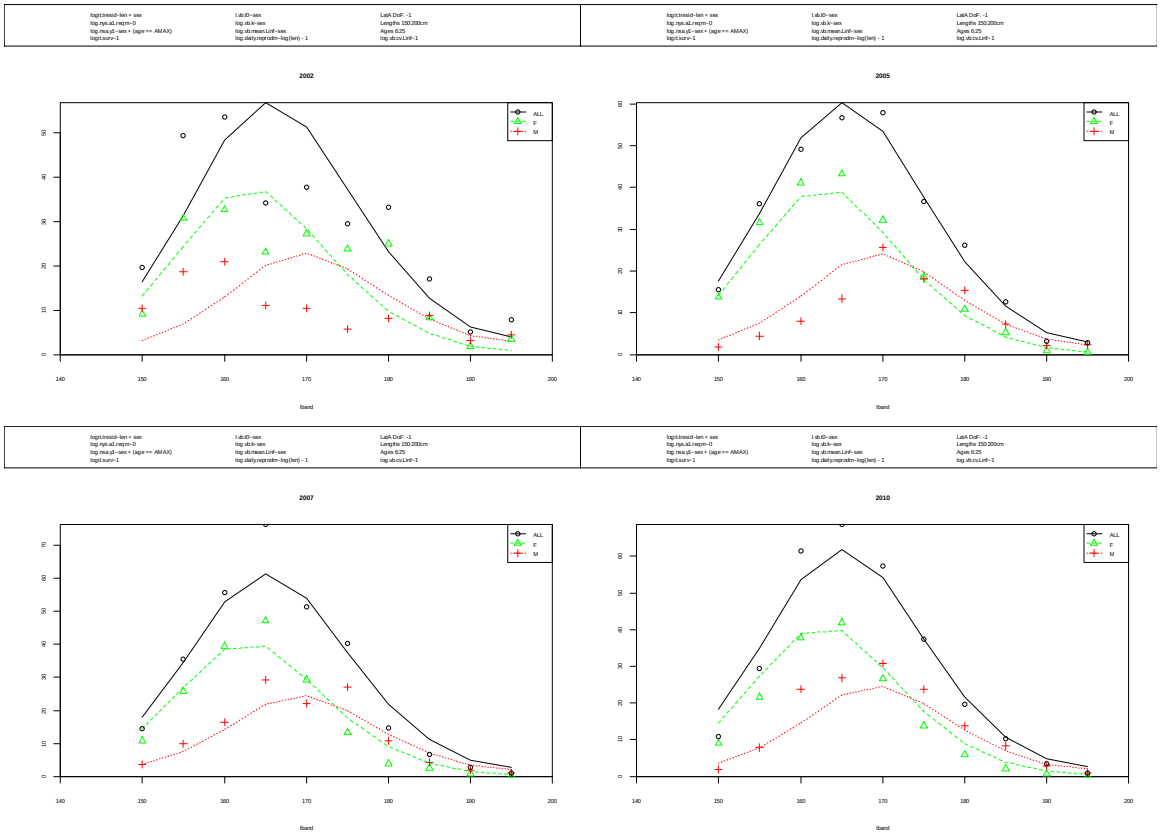
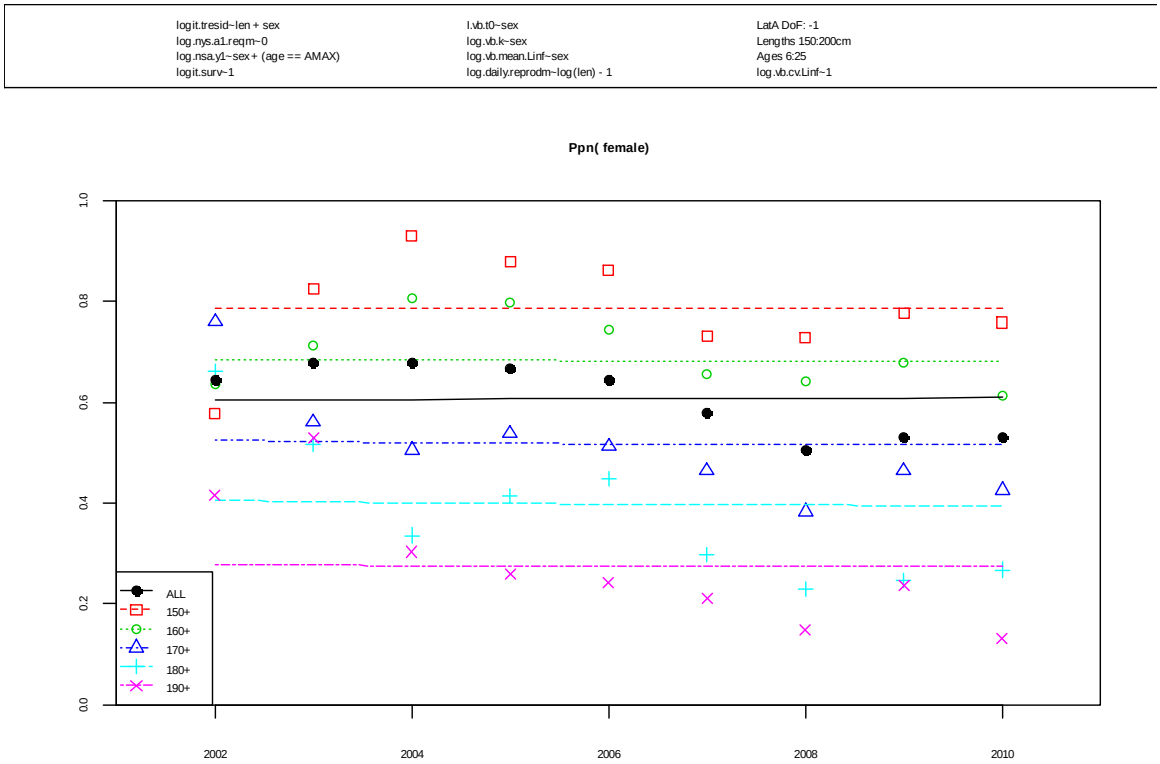


Figure 4.4: Steady-state diagnostics: sex-ratio. The unreadable parts do not need to be read; they show details of the particular model version fitted.

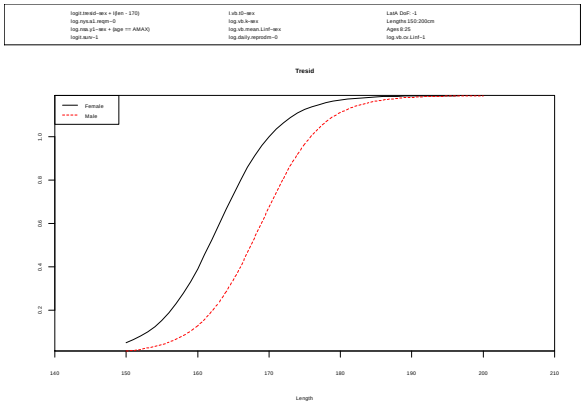


4.4.1 Parameter estimates and uncertainty

Annual adult survival for the steady-state model was estimated at 0.73, fairly close to OM estimates. However, the estimated abundance of 10+ adults in 2004 is much higher than in the OM: 2.04M fish, with a biomass of 157kT. This happens to be fairly close to the simple "twice the comparisons divided by the POPs" estimator, but only by coincidence; the competing effects of dilution by incomers, growth, and selectivity are all strong, and merely happen to largely cancel each other out.

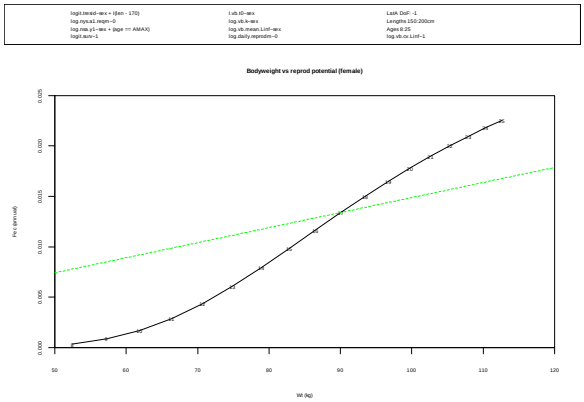
The estimated relationship between residence time (i.e. selectivity, in this model) and length is shown in Figure 4.5; the curve climbs steeply from 160cm for males and about 155cm for females, with males taking longer to "mature". The apparent asymptotic slowdown around 180cm may be a consequence of the functional form chosen (a logistic curve), and warrants further study.

Figure 4.5: Residence time as a function of length



By combining the estimated residence-time with the estimated growth curves (which have average L_∞ of 191cm for females, and 201cm for males) and the fecundity data, it is possible to infer the average female spawning contribution at age. The results are very different to what is assumed in the existing OM, i.e. that spawning contribution is proportional to biomass for ages 10 and up (Figure 4.6); the mini-assessment suggests that older fish are comparatively much more effective spawners than younger fish. This underlines the point that the mini-assessment is structurally different to the OM, and results need to be compared with that in mind.

Figure 4.6: Relative spawning contribution as a function of female bodyweight. Average bodyweight at ages are indicated. Green line corresponds to current OM assumption.



The nominal CV of $\hat{N}\,(10+, 2004)$, obtained by inverting the Hessian, is 17.3%. This is only slightly higher than the theoretical lower limit of 14.9% set by the intrinsic sampling noise associated with the observed total of 45 POPs (see Appendix 3 for why the POP count is unlikely to be overdispersed). The concurrent estimation of survival, residence time, etc., as well as abundance, has contributed only an additional 2.3% to this nominal CV. However, the CV is still "nominal" because that additional contribution is not finalised yet; for one thing, it depends on the "effective sample size" used for the length frequency data, which we simply assumed to be 300 fish per year. Given that there is substantial variability (noise) between observed and predicted length- and sex-compositions, the additional CV should probably be higher than 2.3%. Even so, in an exploratory run with the effective sample size for length frequencies set to just 30 fish per year, the additional CV still only reached 5%, i.e. a total CV around 20%.

To improve and fully de-nominalize the CV, it will be necessary to handle better the overdispersion/variability in the non-POP data, so that the "likelihood" in the mini-assessment accurately reflects the uncertainty. Also, experiments with non-steady state versions of the mini-

assessment indicate that the model uncertainty associated with possible past & present *trends* in recruitment is substantial (though less so around 2004 than in the most recent years). Moving to a random-effects rather than fixed-effect framework for recruitment should mitigate this, and help to accommodate the model uncertainty automatically in the final result. These changes will push the CV up somewhat but, from what we have seen so far, we would be surprised if the final CV[13] exceeds 25%.

Accordingly, we plan to make just a few further changes to the mini-assessment model in 2012, in the course of submitting for peer-reviewed publication and finalizing the CSIRO/FRDC project report. The main ones are:

- allow recruitment (annual incoming 8-yr-olds) to be a random-effect, rather than constant or a trend;
- formally estimate the "effective sample size" for the length-frequency data;
- formally propagate the uncertainty associated with the fecundity GLMs.

With our current mini-assessment, these issues would have to be treated as "model uncertainty", which we have not attempted to explore in this report. Preliminary explorations suggest that (i) the first two can have an appreciable effect on the point estimates of abundance, though by no means enough to change the qualitative conclusions, and (ii) the CV is probably not going to increase much. Once we have tackled these issues inside the mini-assessment, the associated uncertainty should be reflected directly in the final CV.

5 Discussion

This project has successfully managed to complete an enormous amount of genotyping with tight quality control. The data do seem to be able to deliver an internally-consistent fishery-independent[14] estimate of adult abundance, just as planned. The stand-alone estimates are still somewhat preliminary, but are clearly considerably higher than the OM estimates— at least 3 times the point estimate from the "base case" scenario, and on the edge of the upper confidence interval of the most optimistic scenario. While this may seem surprising, it should be emphasized that there is very little reliable data in the OM with which to estimate absolute adult abundance (although other quantities such as relative depletion can be estimated more reliably), which of course is in itself is a big part of the reason for undertaking this study. Also, as noted in section 4.4.1, the different notions of effective fecundity in the two models make direct comparisons tricky. So getting an adult abundance estimate that is very different to the OM's is by no means an indication of a serious problem with either the CK estimate or the main conclusions of the OM. Nevertheless, it is important to ask the obvious question: how wrong could these CK estimates be? There are a limited number of issues to consider, given that we are not asking about small changes here— the point is to try to think of any phenomena that could make a huge reduction to the estimate, of the order of 50%.

5.1 Is the number of POPs about right?

The genetic results strongly suggest that there are few if any false negatives or false positives, given the filtering we have used, the proven effectiveness of our large-scale QC checks in detecting and fixing problems, and the absence of "near-misses" in the lower left-hand corner of

[13] It is important to choose a sensible quantity to estimate the CV of. For example, once recruitment is allowed to vary annually as planned for our final model, the average 10+ abundance across 2002-2008 will have a lower CV than the corresponding abundance in any single year.

[14] Strictly: catch and CPUE independent

Table 3. An independent implementation of the QC software checks would be useful, particularly when the data migrate to our existing SBT database, and this conceivably might unearth a few further problems. However, while it is certainly possible that there are a small number of false negative or false positives in (or not in) our POPs, there is strong evidence that the proportion is not substantial.

5.2 How precise is the estimate?

The nominal CV of our steady-state example estimate is 17.3%, driven mostly (15%) by sampling variability in the number of POPs found. That 15% component seems solid, since there is no reason to expect substantial overdispersion in the number of POPs; see Appendix 3 (section 8). The effective sample sizes we assumed for the length data seem to have been somewhat high, so the current additional contribution of 2.3% from the length/sex/age aspects is probably too low but not by all that much, going by our explorations so far. The remaining changes planned for the mini-assessment will include uncertainty related to those effective sample sizes, and will accommodate aspects of model uncertainty not currently captured in the nominal CV. Based on our explorations so far, though, we will be quite surprised if the final CV exceeds 25%. And, apart from the proviso about selectivity discussed next, the final CV should be a statistically defensible measure of overall uncertainty.

5.3 Is the abundance estimate about right, given the number of POPs?

If the number of POPs is about right, and if the adult sampling is simultaneous with juvenile birth and random, then the cartoon estimate can't go wrong— each juvenile really does have exactly two parents. Most of the other potential problems with close-kin— stock structure, or massive proportions of sibs/halfsibs— don't apply to SBT. So the only other source of possible error is in the adult-assessment model. As mentioned at the end of section 4.4.1, the model is not completely finished and the remaining modifications will change the point estimates somewhat, but we do not expect those changes to be very large. So, aside from possible programming mistakes (this is still a very recent assessment, all coded by one person), there are two main points to consider:

1. The entire CK assessment, and the way in which the cartoon adjustments are implicitly calculated, rests on the assumption that selectivity is primarily driven by residence time— the longer a fish is on the spawning grounds, the more likely it is to be caught, all else being equal. The link between residence time and annual female reproductive output rests on the same assumption (more spawning opportunities). It is hard to see how these assumptions could actually be wrong, but the caution might be in the phrase "all else being equal". If there are other really major length-based effects on selectivity or on reproductive output (aside from female daily fecundity, for which we at least have some data), then bias could perhaps arise.

2. The only other way that an abundance estimate could be biassed, is if there is some type of heterogeneity between adults that is *not* just due to length and sex, and which results in some adults (i) being more likely to spawn offspring caught in the GAB, *and* (ii) more likely to be caught themselves in the Indonesian fishery at least one year later. It is hard to imagine what might cause such heterogeneity.

Even if there do turn out to be errors in these estimates, they seem more likely to be programming errors, and therefore fixable, rather than being intrinsic problems with the data or

its interpretation. The CK data fundamentally do seem to be extremely useful for SBT: they are bearing out their promise. There is obvious scope for continuing to collect and genotype in future, both to build up the time series and also (thanks to the retrospective qualities of close-kin) to increase the number of POPs found from our already-genotyped juveniles from 2006-2010. The way this might fit into SBT management, and the links with other monitoring possibilities, is far more than can be explored in this study, but the potential value of further CK genotyping is clear.

Finally, we draw attention to the key role of residence time on the spawning grounds— or, to be accurate, how the average residence time depends on size— in getting to an actual abundance estimate, and a selectivity estimate, and an appropriate definition of spawning potential. Although there is just about enough data in the POPs and the age/length samples to infer the residence/size link indirectly, it would be immensely useful to have direct estimates from a few adult fish across different sizes, since this could both ground-truth the model and give a basis for estimating further length-dependent effects on selectivity, if that turned out to be necessary. Pop-up satellite tags could yield limited information quickly, but the best data would come from archival tags because they can record over several years, and are not as vulnerable to short-term tagging shock. The low fishing mortality on adults means that quite a few archival tags would be needed to get recaptures, and that we might have to wait a while to get the tags back, but the number of returned tags needed would not have to be at large (even single figures) to give a very useful check on, and input to, close-kin based abundance estimates in future. Such tagging ought not be a very expensive exercise in terms of the value of the fisheries, or indeed the cost of this close-kin project.

6 Appendix 1: The genotyping and QC process

6.1 Terminology

This section is meant as a guide for a non-geneticist in a CCSBT context. It is not intended as an authoritative set of definitions from a genetic perspective, which are widely available in the genetics literature. These definitions include forward-cross-references in *italics*, and use **bold** to indicate additional definitions. In the text after this section, a few technical genetic terms have been used and marked with an asterisk, but deliberately not defined since their relevance will only be apparent to those who already understand them.

Locus: an identifiable place on the genome with characteristic start and end sequences of DNA, and a variable DNA sequence between them. The loci we used are **diploid,** so that each individual has two versions (**copies**), one copy being inherited from each parent. The sequences of the two copies might be different or might by chance be the same. We used **microsatellite** loci, whereby each sequence is characterized simply by its length or **size** (i.e. the number of *nucleotide bases* it contains), which will be a integer in the range say 80-600 depending on the locus and how it is to be purified away from the rest of the genome in any particular study.

Alleles: the set of possible sequences a locus can have, i.e. for microsatellites a set of integers. Alleles at different loci might happen to have the same length, but are in no sense comparable— it only makes sense to refer to an allele for a specific locus. The **allele frequency** for the locus is the frequency distribution of the different alleles across the population under study. A **highly variable** locus has a large number of different alleles and an allele frequency that is not dominated by just one or two common alleles. The probability that two unrelated animals will have an allele in common is lowest if the locus is highly variable, so such loci are preferred for close-kin work. A **null** allele is an allele

that is present in the animal, but is not revealed by genotyping; possible causes include scoring error, and a mutation in or near the locus that causes the DNA amplification process to fail for that copy.

Genotype: which alleles an animal has. Usually means for all the loci together (sometimes called a **multilocus genotype** or **DNA fingerprint**), but can mean just the alleles carried at a single locus if specified.

Homozygote/heterozygote: An animal is said to be a **heterozygote** at some locus if the two copies are different alleles, or a **homozygote** if they are the same.

Scoring/genotyping/calling: deciding which alleles are present at a locus for a particular animal. This really involves many steps, but sometimes "scoring" just refers to the final step of adjudicating on the possible alleles proposed by the *GeneMapper* software. The protocol in our study is that, if the genotype at a particular locus for a particular specimen is ambiguous, no score is recorded (rather than trying to make a subjective best-guess).

Scoring error: Recording the wrong genotype at one or more loci. Large-scale scoring errors affecting many fish and loci simultaneously can arise from inadvertently swapping or rotating entire plates of fish, or from miscalibration of the *sequencer* for a particular *run plate*. Small-scale scoring errors affecting individuals most commonly involve failure to detect a small second peak in a heterozygote, so that the locus is mistakenly scored as a homozygote instead. At least in this study, actual mislocation of peaks were very rare (based on a subset of the fish which were independently re-scored).

Amplification/PCR: the chemical process by which the DNA from certain desired loci only is selected and amplified for input to the sequencer.

(Nucleotide) base is one genetic "letter" (C/G/A/T), the molecular building-blocks which are linked together to form a DNA molecule. DNA occurs in two strands, and each base is paired with its complement on the other strand, so the term **base pair** is often used instead.

Tetranucleotide: The sequences within microsatellite loci are mostly repeats of some short subsequence of base pairs, such as GATA (four base pairs, so a tetranucleotide locus) or CA (a dinucleotide). Dinucleotide loci are more common in most genetic studies, but are more prone to scoring error. In this study we used only tetranucleotide loci.

Panel refers to a set of loci (usually 4-7) which can be analyzed simultaneously by the sequencer.

Plate is a group of 96 DNA samples (including a couple of controls— standard specimens included on every plate— and blanks) placed in wells numbered A1-H12 in an industry-standard format on a small rectangular tray ("**96-well microtitre plate**") ready to load into a sequencer. Each group of 96 fish is originally set up on a **template plate** from which are prepared several **run plates**, all with the same layout of specimens in the 96 wells, but with each run plate specific to a particular panel of loci.

Sequencers are the machines that physically do the genotyping. One run plate is **run** or **sequenced** at a time. For each locus on each specimen, the output is a graph with X-axis corresponding to allele length (as a continuous variable) and "signal intensity" on the Y-axis. Alleles are visible as peaks with a characteristic shape.

GeneMapper is software which identifies possible alleles from the sequencer's output. In most cases, GeneMapper will propose the correct peaks, but each sequencer graph and proposed scores is scrutinized by an experienced scorer who makes the final decision on which peaks truly represent alleles, and which peaks are artefacts.

Bins and binsets: Because of slight variations in run conditions, the locations of peaks reported by GeneMapper will vary fractionally between sequencer runs, even for the same sample. **Bins** are therefore used in GeneMapper to provide tolerance and to convert the continuous-valued peak locations into an integer-valued allele size. Each bin is a continuous-valued range such as $[137.2, 138.6]$, which should span the range of peak locations found for that allele across many runs. The **binset** for each locus is the collection of all its bins. The binset needs to be consistent throughout a study. In this project, we initially developed bins and binsets from genotyping the first 500 individuals, then revisited them after 5000 specimens had been genotyped.

6.2 Flowchart of genotyping

For the last 9000 of the 14000 fish genotyped (from both sites), the procedure was as follows.

1. Tissue biopsy samples from each fish are collected, labelled, and stored in boxes of 100 fish, with corresponding information on fish length, date of capture etc, and in the case of the Indonesian fish are cross-linked to the existing data (sex, otolith if collected, etc).

2. The original tissue from each fish is subsampled down to the 10mg size suitable for DNA extraction. The remaining biopsy tissue is archived, so this (relatively costly) step can be repeated if necessary.

3. The DNA of 96 subsamples at a time is extracted into solution. Part is kept frozen as an *archive plate*. Part is used as the template plate, incorporating two controls (in specified positions) and two water blanks (in known positions, variable from plate to plate).

4. For each template plate and panel of loci, a small amount of fluid is used to prepare a run plate. Enzymes are used to amplify the desired loci (for the panel) from the rest of the genome, and their DNA is PCRed in a multiplex(*) reaction at CSIRO.

    a) The first column of each run plate is then duplicated as a column on a *check plate*, of which there is one per 12 run plates.

5. Run plates are sent to AGRF[15] in Adelaide for sequencing. There are four run plates per 96 fish, labelled H/I/J/L depending on panel of loci are involved. For each run plate, the result is a set of 96 "FSA files" suitable for input to GeneMapper.

6. FSA files are scored at CSIRO by an experienced team; we have used only four scorers, each of whom scored several thousand samples. Results from the various scorers have been cross-checked for consistency on some plates.

    a) The check plates are sequenced at CSIRO using similar machinery to AGRF's, and the results are compared to the corresponding columns of the FSA files from AGRF. This provides a safeguard against plates being swapped or rotated, and against miscalibration of the sequencer.

[15] Australian Genome Research Facility

b) The panels all include a common locus B8B, so by comparing the B8B scores across run plates ostensibly from the same template plate, we could check whether the files for each run plate really did come from their nominal samples.

7. A variety of QC checks are run on the FSA files, to detect plate-level phenomena such as rotation/swapping/miscalibration (see 6.1), atypical allele frequencies, and excess homozygotes, and individual-level phenomena such as duplicate genomes which arise if samples are inadvertently double-sampled at the point of collection.

For the first 5000 fish genotyped, a slightly different and less streamlined procedure was used in steps 4-5. Only 20 of the 27 final loci were used. The PCR for some of the 20 was done by AGRF in single-plex(*) reactions which were then grouped into three panels A/B/C to make run plates, while the remaining loci were multiplexed at CSIRO as above into two run plates corresponding to a pair of panels D & E. [The A/B/C loci were subsequently reorganized into two of the panels (I & J) used for the last 9000 fish; the D/E panels became the H & L panels after the 7 extra loci were added.] After the FSA files returned to CSIRO, we used a shared locus on the D & E panels to check their "alignment", as in 6b. To check alignment of the A/B/C panels with each other and with the D/E panels, we put DNA drawn from the first column of the template plates for D/E panels into one column of an extra template plate, which was then used to make run plates for the I & J panels (containing the same loci as A/B/C, but organized differently). These were sequenced, scored, and the genotypes compared against the corresponding columns in the original A/B/C plates. Although this process was somewhat cumbersome and led to some duplication in scoring (about 10%), it provided an important safeguard against the handling errors that become almost inevitable with such large sample sizes.

6.3 Selection of loci

Loci for this project went through an particularly extensive checking process. Past experience indicated that, to be conservative and to facilitate automated genotyping, we needed to strictly focus on using tetranucleotide repeats that gave solitary, sharp, allele peaks. In short, we wanted a set of loci that: were highly variable but not so variable that the longest alleles failed to amplify well; had simple peak structure with minimal shoulder to the peaks and little stutter; and had clear gaps between alleles. Over time, as more fish were scored, some of our best tetranucleotide loci turned out to have some two-base-pair insertion/deletions, which meant that some alleles were separated by only two base pairs (though usually at least one of the alleles involved was rare). This was tolerated, provided there was at least a one-base-pair gap between bins. Loci were immediately discarded during the initial testing phase if they showed alleles separated by just one based pair, indicative of poly-nucleotide tracks in the amplified allele. After genotyping 5000 fish, we had developed 20 loci organized into 5 panels A-E, with very comprehensive scoring binsets into which almost all detected alleles fell. At this point we included an additional 7 loci (total 27) which were re-organized into four multiplex panels H, I, J, and L. We scored all 27 loci where possible in the remaining 9000 fish, but used only 25 loci for finding POPs; the remaining two loci, with slightly less reliable scoring, were used only for QC purposes, as per 6b above. When scoring, our protocol was not to record a score if in doubt, which is safe for purposes of POP-finding.

An important check in genetic studies, is on the proportion of homozygotes found at each locus. In theory, provided a number of assumptions hold, this can be predicted from the allele frequencies, and the extent to which there is an excess of *apparent* homozygotes is one indication of the reliability of a locus. As shown in Table 7, all but 3 of the 25 primary loci have both low expected homozygosity (which corresponds to being highly variable, and thus powerful for POP

identification), and at most a small excess observed homozygosity as given by the difference between the EXP and OBS rows; this suggests relatively few cases of failure to see the 2nd peak in a heterozygote, for example[16]. The exceptions are in the bottom right of table: D569 and D573. It appears (as described shortly before section 2.1.1) that the excess of homozygotes in those two loci is due to "heritable nulls" (eg from a mutation in the flanking sequence(*) so that primers(*) don't bind), so that some alleles simply don't amplify. No loci showed appreciable evidence of Short-Allele Dominance(*).

To guard against the possibility of heritable nulls in *any* locus, a comparison of two different apparent homozygotes (AA in one fish vs BB in the other) was *not* used to exclude a POP even though there is ostensibly no shared allele, in case the real score was "A-null vs B-null" with the null being inherited. This relaxation has only a small effect on the false-positive probability. However, it is not feasible to relax the exclusion criterion further to allow for the commonest (but still fairly rare) scoring error whereby the second peak of a heterozygote is missed, i.e. by treating AA vs BC as not necessarily excluding. Such a weakened criterion would generate many false positives with the existing set of loci, so many more loci and more expense would be required.

Table 7: Homozygote percentages, "expected" (ignoring nulls) and observed; see text.

	3D4	B5	D10	D111	D11B	D12	D122	D201	D203	D211	D225	D235	D3	D4D6
EXP	19.8	6.8	7.1	11.8	10.7	10.8	9.7	11.7	7.5	11.4	3.4	8.5	16.8	5.5
OBS	19.8	7.3	7.3	12.2	12.3	10.9	11.4	12.4	9.0	17.0	3.7	14.8	16.8	6.7

	D541	D524	D549	D570	D592A	Z3C11A	D517	D534	D582	D569	D573
EXP	14.0	12.4	11.9	7.3	9.8	13.0	3.1	9.3	7.6	9.9	4.9
OBS	14.0	13.5	11.9	7.3	10.2	13.4	3.4	10.1	7.7	45.5	30.9

6.4 QC for Consistency of Allele Size Calling

Examining the consistency of allele-size calling is fairly straightforward, and is mostly dealt with by use of an internal standard and use of an automated genotyping program developed by ABI-Life Technologies (supplier of the DNA sequencer used for fragment separation). To further minimise inter-run variation, all size fragmentations were run on only one DNA sequencer located at the Australian Genomic Research Facility (Adelaide node). This eliminated variation occasionally observed when the same samples are run at two facilities even on the same model of sequencer.

In addition, the ABI system uses an internal size standard added to each sample from which the size curve is extrapolated for estimating allele peak length relative to the standard curve. ABI states that variation using this system ensures +/- 0.5bp accuracy from run to run. Furthermore, the GeneMapper program analyses each individual size curve for peak quality and general fit to the theoretical ideal size curve. Any discrepancies detected by the software raise flags in the analysis window and can be scrutinized in further detail. We also examined each size curve analysis as well as the individual peaks that were used to generate the size curve for each individual in a run plate to ensure another level of QC in addition to that used by the GeneMapper software.

GeneMapper uses a standard set of allele size bins used to smooth out further subtle variation and ensured easy comparison among alleles from different individuals and provided another

[16] Other reasons for deviation from Hardy-Weinberg, such as population structure, are unlikely for SBT, and in any case no deviation is seen for many of the loci despite the very high sample sizes and consequent high power to detect any deviation.

level of QC among plates. Bin sets are developed for each locus to permit automated genotyping using the GeneMapper software. Individual bins represent a value range centred on the median length value of each allele as ascertained following sizing of an initial set of individuals. Preliminary bin sets were developed following detailed analysis of about 500 fish. These sets were designed to encompass slight variations to permit detection of gross deviations from the norm greater than +/- 1.0 bp. After genotyping about 5000 fish, the bin sets were re-assessed for consistent allele calls, and a final consensus adjustment was determined. Bins permit assignment of an integer value to the continuous-valued allele length based on the GENESCAN size standard, and permit simple comparison of allele identities among individual genotypes. A gap of one to three base pairs between bins ensures that an objective decision rule can be consistently applied to a genotype for inclusion of an allele into a designated integer bin. Alleles falling in the gap were rare and presumed to be a result of an insertion or deletion event on an individual's DNA. These were scored as "unknown genotype" but the real value could still be used for confirmation of parentage should it be required to confirm identity (not required with our samples to date).

The use of automated genotyping with a single set of GeneMapper bin-sets allowed us to detect if peaks were consistently falling outside of predetermined bins and would highlight a general problem with the running of a plate (eg. old buffer or polymer in the sequencer leading to general failure of proper electrophoresis and inconsistent separation). Runs where problems were found were re-run with new buffer and polymer; this rectified the problems in every case.

6.5 Avoidance of chimeras

Chimeric genotypes are (in this study) a composition of DNA from more than one fish, rather than (as in some other studies) DNA profiles resulting from multiple DNA in a well (two or more contaminated DNA leading to more than two alleles present for each locus). There are only two possible sources. First, a chimeric error will result from turning a run plate 180 degrees, whereby e.g. the A1 position became the H12 position. This error produces what looks like a legitimate DNA profile but made up of some loci from fish A1 mixed with the remainder of loci from H12 from the run plates that were not rotated. Second, if two run plates are swapped, the loci for those panels (but not for the other panels on the same fish) will be swapped. Clearly, these errors will lead to any POP members on the plate being overlooked, affecting 100-200 fish at a time, so it is important to catch them. Fortunately, once one is aware of these possibilities, it is fairly easy to write QC software using the check-plate results and/or the controls to detect and fix the problem. We did find both types of chimera in this study (rarely), but thanks to the QC protocols we were able to detect and fix them.

6.5.1 Further processing details for the first 5000 fish

For the first 5000 fish we developed a unique system to cope with the potential issues arising from PCR and fragment separation methods used at the outsourcing facility (AGRF). The first 5000 fish were run at AGRF as three single-plex (A, B, and C) and two multiplex (D and E) panels. At this point the multiplex PCR was clearly the most optimal solution and we included 7 additional loci that were incorporated into an optimised set of four multiplexed panels (H, I, J, and L). The A, B, and C panels were combined into the I and J panels while D and E were combined into H and L. To check for generation of chimeric genotypes we used the set of template plates that were the source of DNA for the D and E panels. Since D and E had a common locus scored for both plates we were able to ensure that there were no chimeric individuals there. We then ran the first column of each template plate for panel-I and for panel-J. This checked the genotype calls of 8 individuals that should be identical if no mix up had occurred. We verified that all 8 genotypes for each locus was congruent across all tested

plates indicating that no single-plex mixups had occurred. Since the template plates used were those used to set up D and E we were then assured that there were no chimeric fish generated in the first 5000 genotyped individuals.

6.5.2 Further processing details for the last 9000 fish

For the balance of the fish, a unique system to identify individual template plates was developed to ensure that the fish on the plate could be identified, and that it was not accidentally rotated prior to sequencing. The four panels had a common locus to check on plate to plate variation, and also to detect PCR contamination via negative water controls. Template plates were created in a specific routine fashion with four positions in each plate reserved for positive and negative controls. We used two positive control individuals on every plate with position A01 being control fish #1 (TC-2005, male) and G12 being control fish #2 (TC-2205, female). The positions of the negative water controls were used to uniquely identify each plate. For example, one plate would have water controls in position A02 and A07, while the next plate would have A02 and A09. Care was taken to ensure that the water was placed in one odd-numbered and one even-numbered well row due to the way the 48 capillary sequencer picked up the samples; every dip of the sequencer thereby had one positive and one negative control, so that each electrophoresis had internal controls to check run quality. The internal common locus control for each individual checked to see that each fish was scored with consistent fragment separation for each of the panels. By use of this system for the final 9000 fish, we were able to QC for chimeric individuals, check for PCR contaminants in the master mix, ensure that run conditions did not affect genotype scoring among the four panels, and also ensure that plates were not mislabelled or loaded into the sequencer incorrectly. Our QC caught a few errors but these were few and subsequently dealt with by a quick rerun of the PCR or fragment separation or both.

7 Appendix 2: Rigorous estimation of false-negative (FN) rates

The question of interest is: what proportion of true POPs could have a scoring error that leads to the POP being overlooked? We can estimate this directly by comparing Table 3— observed numbers of (loci compared, loci failing to match)— with Table 4 (expected version of Table 3, assuming zero POPs and therefore zero FNs). If the expected-value calculations behind Table 4 are correct, and if there are numerous true POPs without FNs, then Table 3 should resemble Table 4 except for numerous entries in the F0 column— which is pretty much the case. If the Table 4 calculations were wrong for some reason[17], then the upper-right-hand triangle of numbers in Table 3 would be stretched to the left compared to Table 4— which is not the case. Therefore, we can take the expected values in Table 4 as correct if there were no POPs, and use the differences between the tables to make inferences about the true number of POPs, and about how many FNs are in Table 3. We can do this because FNs will appear in Table 3 as an "echo" of the F0 column, predominantly in column F1, and somewhat weighted towards the lower rows because there is more chance of a scoring error when more loci are involved. Apart from chimeras and mass failures of PCR on a run plate, as described and ruled out in Appendix 1, there seems no reason why scoring errors should not be independent across loci

[17] The only theoretical reason we can see why the calculations in Table 4 might ever go wrong, is if genotypes at different loci within each fish are not independent, something which could arise from substantial cryptic stock structure, with different allele frequencies in the different stocks. That situation is *a priori* unlikely for SBT, and happily there is no suggestion of it in Table 3.

on the same fish; hence, provided scoring errors are uncommon to begin with, FNs are most likely to be in the F1 column, less likely to be in F2, and rapidly less likely beyond that.

The numbers in Table 3 actually result from a second round of checking; we re-scored all the pairs in the F0 and F1 column, and in the lower rows of the F2 column. However, only a small percentage of the fish were re-scored during the second round, and the level of attention paid to these fish may not be typical of the rest of the sample. In this section, we have therefore analysed the data from the *preliminary* version of Table 3, before any selection of fish to re-score took place. This makes the analysis general, but also means that the results are pessimistic in terms of FN likely FNs compared to the final data, because the FN/near-FP status of many would have been cleaned up during re-scoring. The preliminary data, shown in 8, is very similar to Table 3, the main difference being that the C23 row starts (3,1) rather than (4,0); this is one case where a scoring error did cause a false-negative, though this was subsequently detected and fixed on re-scoring. The other differences did not affect POP status of any pairs.

Table 8: **Preliminary** number of *usable* pairwise comparisons, by #loci and #excluding loci, **before** re-scoring. First three columns only.

.	F0	F1	F2
C11	.	.	.
C12	.	.	5
C13	.	2	16
C14	1	4	61
C15	.	3	42
C16	1	1	18
C17	3	.	7
C18	5	.	7
C19	7	.	1
C20	2	1	1
C21	14	.	1
C22	.	.	.
C23	3	1	.
C24	2	.	.
C25	6	.	1
SUM	44		.

7.1 Likelihood for estimating false-negative rate

Let θ be the probability that a pair of fish will be a POP (so θ is inversely related to abundance, etc), and let e be the probability that one shared locus in a POP will fail the parent-offspring compatibility test[18], either through mis-scoring or mutation. Assuming scoring errors at different loci are independent[19] and equally likely[20], then the probability of f loci failing in a POP where c loci are compared, is a simple Binomial probability. Also, for a non-POP pair where c loci are being compared, let p_{cf}^{NON} be the probability that f of the loci will fail the test. For

[18]The basic test is: do they share a visible allele? We used a more relaxed version, so that AA vs BB homozygotes are also deemed (potentially) compatible.

[19]Apart from chimeras, as described and ruled out in Appendix 1, and mass failures of PCR on a run plate which would be picked up by our other QC checks, there seems no reason why independence could fail.

[20]Strictly, the probability of a scoring error that leads to rejection of POPhood probably varies somewhat across loci, but there is not nearly enough data to estimate this; and since the set of loci that actually get used in a comparison is a random variable, and we are only concerned with one or two errors here, the approximation is statistically negligible.

any given pair, this actually depends on the particular loci involved, and is already calculated to form the basis for the expected values in Table 4. Any given pair with c loci compared is either a POP or not, and the probability p_{cf} that the pair will fail at f loci is therefore

$$p_{cf} = \theta \binom{c}{f} e^{f} (1-e)^{c-f} + (1-\theta)\, p_{cf}^{\text{NON}}$$

Therefore, if n_c denotes the number of comparisons using c loci in Table 3, the expected value of cell (c, f) is $n_c p_{cf}$. Strictly, the distribution within each row is Multinomial, but in the first few columns the multinomial "size" is enormous (millions) and p_{cf} is small, so a Poisson approximation is perfectly adequate. If y_{cf} denotes the observed number of pairs in the (c, f) entry of Table 3, then the likelihood of the first few columns up to F failures is (up to a constant)

$$\prod_{c=11}^{25} \prod_{f=0}^{F} e^{-n_{cf} p_{cf}} \left(n_{cf} p_{cf}\right)^{y_{cf}}$$

The term p_{cf} involves the parameters θ and e, which can be estimated via maximum likelihood.

The bulk of the information on false-negative rates is contained in the F1 column (and the F0 column, which is needed for estimating θ), with a little coming from the F2 column. To the right, the noise from the increasingly large numbers of almost-false-positives swamps any signal related to false-negatives with 2, 3, etc number of failures, which will be increasingly rare.

7.2 Confidence intervals on actual FNs

Although the Hessian from the above likelihood could be used in the standard way to derive a confidence interval for the *expected* number of FNs in a *replicate* of this study, that would be solving the wrong problem. Our interest lies in the *actual* number in *this* study; so, if FNs were very unlikely beyond the F1 column, then the number of FNs would be capped above by the total F1s seen, regardless of how many might be found if the study was repeated. This makes quite a difference in practice. A Bayesian argument is required to get the answer we need.

We need the probability distribution of the number of false-negatives $\#FN$ given the observed data, i.e. $\mathbb{P}[\#FN|y]$ where $\#FN$ is the total number of False Negatives and $y = (y_{cf} : c \in 11 \cdots 25,\ f \in 0 \cdots 1)$ is the observed numbers in the F0 and F1 and possibly F2 columns (F3 onward are irrelevant because the chances of 3 or more scoring errors is negligible). For simplicity of argument, say for now that we neglect the F2 column as well. Obviously, the maximum possible value of $\#FN$ is the observed number of F1s, in this case 12. Each of these F1 pairs is either a near-FP or an FN. The probability that an F1 pair with c loci compared is actually a FN rather than a near-FP, is

$$\frac{\mathbb{P}[1 \text{ error in } c \text{ loci}] \times \mathbb{P}[\text{is POP}]}{\mathbb{P}[1 \text{ error in } c \text{ loci}] \times \mathbb{P}[\text{is POP}] + \mathbb{P}[\text{match at } c-1 \text{ of } c \text{ loci}] \times \mathbb{P}[\text{is not POP}]}$$

One implication is that a (C12,F1) fish is much more likely to be a near-FP than a (C25,F1) is, because (i) the probability of a non-POP matching by chance at 11 of 12 loci is much higher than for 24 of 25, and (ii) the chance of a scoring error is about twice as high with 25 loci as with 12.

The FN-status of the pairs are independent[21], θ and e, so the total number of F1 pairs that are FNs is the sum of (in this case) 12 independent Bernoulli (0/1) random variables, with

[21] I.E. the probability that a given F1 pair is actually FN or near-FP is unaffected by the FN-status of the other F1 pairs, given θ and e.

probabilities depending on the number of loci involved. There is an algorithm for calculating the Bernoulli-sum probability distribution, which is already used in the expected-FP calculations[22]. Hence, given a pair of values (θ^*, e^*), we can easily compute $\mathbb{P}[\#FN = x|y, \theta^*, e^*]$ for $x \in 0 \cdots 12$. What we actually need, though, is

$$\mathbb{P}[\#FN = x|y] = \int \mathbb{P}[\#FN = x|\theta, e, y]\, f(\theta, e|y)\, d(\theta, e)$$

which can be estimated by repeatedly drawing pairs (θ^{*j}, e^{*j}) from the posterior distribution of $(\theta, e|y)$ via importance-sampling, and then averaging the $\mathbb{P}[\#FN = x|y, \theta^{*j}, e^{*j}]$ across all the draws. This requires a prior for (θ, e), which we took to be independent uniform on $\log\theta$ and logit$e$, plus of course the likelihood from section 7.1. A fully-conditioned confidence interval on $\#FN|y$ can then be found simply by inverting the cumulative distribution of $\#FN|y$.

7.3 Results of FN analysis

We ran the above algorithms first on just the F0 & F1 columns of Table 8, and then on the F0, F1, and F2 columns. In the first version, the Maximum Likelihood Estimate on #FNs was 1.95 and the 95% UCI was 2.46; in the second version, the numbers were 3.19 and 4.0. The difference is entirely driven by the (C25, F2) entry, discussed further below; without it, the two versions are almost identical. Both versions indicated a very low *expected* number of FNs in the F2 column or beyond (less than 10% of the number expected in F1), although the second version clearly identified an *observed* likely-FN at (C25, F2).

As noted above, these FN estimates are *prior to* rescoring the F0, F1, and F2 (from C16 down) columns. Rescoring certainly fixed one FN, at (C23, F1), so the appropriate estimates and limits for the number of FNs in our final dataset (after re-scoring) are no more than (MLE 0.95, UCI 1.46) or (MLE 2.19, UCI 3.0).

The nature of the mismatching loci for any pair provides additional information on whether an F1 or F2 pair is really a FN, as opposed to just being a lucky near-FP from an unrelated pair. This is because one type of mismatch arises from a comparatively common scoring error (overlooking one allele, so a fish is recorded as AA when it should be AB), whereas the other type (incorrect size for an allele) is extremely unlikely; this was apparent in the results from our routine QC rescoring exercises of individual fish. In particular, after carefully rescoring the (C25, F2) pair, the only way it could be a FN POP would be to have a mutation at one locus and a scoring error at a second— a very unlikely conjunction of events. However, this pair is also a very unlikely event under the only two other possible scenarios: an exceptionally-matched unrelated pair, or a well-matched uncle-nephew-pair (which must be much, much rarer than unrelated pairs). In the end, the only way to resolve the true status of the (C25, F2) pair will be to use more loci, which we plan to do as part of a different project. We cannot at present decide whether to treat (C25, F2) as a FN (in which case we should use the second version of the FN analysis, including the F2 column, to get a point estimate of about 2 FN), or not (in which case we should use the first version, with a point estimate of about 1 FN).

Thus, further detailed investigation of the rescored F1s and F2s might eventually shed some light on whether we should expect 0, 1, or 2 FNs in addition to our 45 POPs. However, whichever the answer, the analysis in this Appendix demonstrates that the proportion of FNs to true POPs must be small, and is certainly not going to affect the qualitative conclusions of this project.

[22] K Butler, M Stephens (1993): The distribution of a sum of Binomial random variables. Tech Rep 467, Department of Statistics, Stanford University

8 Appendix 3: What might cause overdispersion in the POPs?

The CV of the "cartoon" abundance estimate is just the CV of the number of POPs found. We have treated this as "count data", so that its variance is equal to its mean. The question arises: under what circumstances might there be overdispersion in this count?

Overdispersion would arise when the 38,000,000 comparisons are substantially non-independent. It's easy to see why a high frequency of (half)sibs would do that: if every juve had one full-sib partner in the sample, then the results for one sibling completely predict the results for the other, and the information content would only be that of 19,000,000 independent comparisons. (Recall that each POP is counted, even if the same adult is involved in several POPs— so there's no bias, only a loss of precision.) Fortunately, (half)sibs do not seem to be common in our juvenile samples, and for clarity we therefore ignore the possibility of (half)sibs in the discussions below.

There are other phenomena that might at first be suspected of causing overdispersion, but careful thought is required. For example, the 38,000,000 SBT comparisons are based on "only" 13,000 fish, each being used in multiple comparisons. Does this somehow mean that the "effective sample size" is much smaller, i.e. that there is somehow serious non-independence amongst the 38,000,000 comparisons? No— but the reasoning is subtle. Ignoring sibs as per above, consider a comparison of two fish, juvenile J and adult A, in the "cartoon" version. With no further information except the population size N, the chance of a POP would be 2/N. Assume (as with SBT) that N is large, the sample is moderately large, and the number of POPs is small. Independence amounts to the following question: does knowing that (i) J is not in a POP with any of the *other* non-A adults, and (ii) A is not in a POP with any of the *other* non-J juveniles, help us to predict the outcome of the J-A comparison?

The information in (ii) is irrelevant (given that the other juveniles aren't halfsibs of J), because if N is large then the number of non-J offspring of any adult in the sample will almost always be zero anyway, so knowing that it really is zero for one particular adult is not informative. And as for (i): knowing that the other sampled adults aren't J's parents tells us almost nothing almost nothing about whether A will be J's parent[23]. Finally, comparisons that don't involve either J or A are obviously irrelevant. So, at least in the more than 98% of comparisons that don't involve a member of a true POP, knowing the result of all the other comparisons doesn't help us predict the outcome of this one— which is the definition of independence. [If the sampled fraction of fish was a substantial proportion of the total population size, and/or if a substantial proportion of the sampled fish turned up in POPs, and/or if there were many sibs in the samples, this argument would break down.]

Another phenomenon that might superficially seem like a source of overdispersion but actually isn't, is the non-random sampling of juveniles, e.g. shifts in sampling locations within the GAB between years. Non-random juvenile sampling has in fact been a deliberate aspect of the design all along, from the 2007 CCSBT paper onwards; for example, we don't sample any juveniles off South Africa. However, as noted in that paper, the only things that matter in order to keep the comparisons statistically independent, are that (i) there are few (half)sibs among the juvenile samples, and (ii) that the *adults* be sampled randomly (apart from selectivity and other effects that are specifically allowed for in the mini-assessment). Even then, all that "randomly" has to mean is: "a parent of one of the sampled juveniles is just as likely to be sampled X years after that juvenile's birth, as is another adult of the same sex, age, and size".

There is one other phenomenon which theoretically could be important for CK abundance

[23] "Almost" because this information does slightly reduce the potential pool of parents, from N to [N minus the adult sample size].

estimates, not so much for overdispersion as for bias: an unholy trinity of cryptic stock structure, biassed sampling of adults, and biassed sampling of juveniles. A lengthy explanation was given in our 2007 CCSBT paper, and is copied below. The key point to add in 2012, is that we have now checked as suggested in 2007 for any *temporal* substructure on the spawning grounds (see 3.3), and found none; we have not checked *spatial* substructure, but as below this seems *a priori* unlikely.

[4.7 from CCSBT 2007 CK paper] Population structure

So far, it has been assumed that SBT form a single population with complete interbreeding. Although no previous study has found evidence of population structure, conventional population genetics applied to large populations is a notoriously blunt tool for that task. It turns out (see **[6.0.6]**) that the basic method is unbiassed even when there is population sub-structure, providing that sampling is proportional to abundance across either the sub-populations of adults, or the sub-populations of juveniles. In our SBT project, juvenile samples come only from the GAB, so if there are substantial numbers of non-GAB juveniles out there somewhere, then juvenile sampling will obviously not be proportional. However, adult samples should cover the spawning season and spawning area, although not necessarily in strict proportion to adult SBT density. Hence, the basic estimator would exhibit population-structure bias if and only if three conditions all apply:

1. adults exhibit fidelity across years to particular parts of the spawning season and/or spwaning grounds;

2. the timing or location of spawning affects a juvenile's chances of going to the GAB (rather than going elsewhere or dying young);

3. sampling coverage of the spawning grounds (in time and space) is substantially uneven, and correlated with the fidelity patterns in (1). (In other words, if adults showed timing-fidelity but not spatial-fidelity, whereas coverage was even across the spawning season but not across the spawning grounds, then the uneven spatial coverage would not matter.)

There is no direct information on condition 1. With respect to condition 2, much the greatest part of SBT spawning occurs within the North Australian Basin ([**?**]), and particularly towards the east and south of the basin beyond the Australian shelf, where the Indonesian through-flows in summer would tend to push the larvae together into the Leeuwin current. These conditions seem unlikely to induce a strong location-of-spawning effect on most juvenile's subsequent propensity to go to the GAB[24], although a timing-of-spawning effect is possible. With respect to condition 3, the Benoa-based operations that we are sampling coincide well with this main spawning area ([**?**], Figure 4.3.1; note that the fishing range has expanded southwards since then, as per [**?**]). Approximate timing-of-effort information could be probably be obtained from the sampling program; spatial information has proved harder to get, but the data obviously do exist somewhere at the company level, and some insights may be obtainable through, for example, the observer program ([**?**]) or the Fishery High School program ([**?**]).

Fortunately, there is enough information in the project data to check the first two conditions. If the seasonal/spatial distribution of identified parents of GAB juveniles is substantially different to the seasonal/spatial distribution of all adult samples, then that is a clear signal that

[24] A small proportion of larvae are found to the north of the NAB and west of it. Different oceanographic conditions apply there, and those larvae could well end up somewhere different as juveniles. However, at least until 1981, this proportion was small.

the first two conditions do apply. Such evidence of population structure[25] would be of major qualitative importance to management, regardless of its impact on quantitative results.

If and only if the first two conditions do apply, then the third could be checked using timing (and perhaps location) information on Indonesian samples. And if all three conditions do apply, then it should be possible to adjust for the uneven adult sampling probabilities, again using sampling coverage information. That is very much a bridge to be crossed only if we come to it; but because the sampling coverage is at least fairly complete[26] even if not necessarily balanced, we would in principle be able to develop a correction if required.

[6.0.6 from CCSBT 2007 CK paper] Population substructure and sampling bias Suppose the entire adult population of N is made up of two sub-populations with proportions π and $1-\pi$, and that adults are sampled proportionally from their respective sub-population, so that the overall adult sample contains $m_A\pi$ fish from the first sub-population and $m_A(1-\pi)$ from the second. Juveniles, though, are not necessarily sampled in proportion to sub-population abundance; let m_{J1} and m_{J2} be the numbers sampled from each sub-population.

If the entire dataset is analysed without regard to sub-populations, then the expected number of POPs can be calculated by considering samples from each sub-population separately (since there will be no cross-POPs between juveniles from one sub-population and adults from the other):

$$\begin{aligned}\mathbb{E}[H] &= \frac{2m_{J1}(\pi m_A)}{\pi N} + \frac{2m_{J2}(1-\pi)m_A}{(1-\pi)N} \\ &= \frac{2m_{J1}m_A}{N} + \frac{2m_{J2}m_A}{N} \\ &= \frac{2m_J m_A}{N}\end{aligned}$$

just as in the case without sub-populations. In other words, the basic estimate is unbiassed provided at least one life-stage is sampled in proportion to sub-population abundance. If both are sampled disproportionately, though, there will be bias.

[25] "Population structure" is probably the wrong phrase, because the behaviour does not have to be heritable; adult spawning preference need not be related to earlier juvenile GABness, even if offspring's GABness is driven by adult spawning preference.

[26] Again: over the great majority of the spawning area.

CONTACT US

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

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

Wealth from Oceans Flagship
CSIRO Mathematics, Informatics and Statistics
Mark Bravington
t +61 3 6232 5118
e mark.bravington@csiro.au
w www.csiro.au/cmis

CSIRO Marine and Atmospheric Research
Peter Grewe
t +61 3 6232 5222
e peter.grewe@csiro.au
w www.csiro.au/cmar

CSIRO Marine and Atmospheric Research
Campbell Davies
t +61 3 6232 5222
e campbell.davies@csiro.au
w www.csiro.au/ cmar