SCIENTIFIC COMMITTEE
SEVENTEENTH REGULAR SESSION

Electronic Meeting
11-19 August 2021

South Pacific Albacore Close-Kin Mark-Recapture: update on design (Project 100b)

WCPFC-SC17-2021/SA-IP-14

M. Bravington[1], S. Nicol[2], G. Anderson[2], J. Farley[1], J. Hampton[2], C Castillo-Jordan[2], J Macdonald[2]

[1] CSIRO Marine Lab, Hobart, Australia
[2] Pacific Community (SPC), Ocean Fisheries Programme (OFP), Noumea, New Caledonia

South Pacific Albacore Close-Kin Mark-Recapture: update on design

Mark Bravington[*], Simon Nicol[†], Giulia Anderson[†], Jessica Farley[*], John Hampton[†], Claudio Castillo-Jordan[†], Jed Macdonald[†]: July 2021

Summary

This Information Paper summarises for SC17 the work to date on design of a Close-Kin Mark-Recapture (CKMR) program for South Pacific albacore in the WCPO to estimate absolute population size, natural mortality rate, and connectivity (SC16-SA-IP-15-Alb). Progress includes:

1. Replacing the very simple model used for preliminary feasibility checks with a more sophisticated age-based CKMR-driven analysis, and then checking what CVs might be obtained for various sampling strategies and total sample sizes.
2. Consideration of the basic implications of potential spatial substructure among juvenile albacore.
3. Field trials to develop "standard operating procedures" for the rapid and contamination-free collection of high-quality DNA samples, while minimizing logistical overheads.
4. Collation of samples to test and calibrate the epigenetic ageing of albacore; this new technique, which can in principle deliver cheap age estimates just using the tissue biopsies collected for CKMR, thereby removing/reducing the necessity for collecting and ageing of otoliths.

In terms of high-level design, our interim results support last year's indication that a total sample size around 25,000, split across say 3 years of sampling, should give robust estimates with acceptable CVs. That number may need adjustment based on results from the 2021 stock assessment (Castillo-Jordan et al., 2021), and further consideration of sampling logistics. During the coming 12 months, we will update the design calculations accordingly, and will present CKMR background and albacore-specific design results via online seminars, in the expectation of presenting a full albacore CKMR proposal to SC18.

[*]CSIRO Marine Lab, Hobart, Australia
[†]Pacific Community (SPC), Ocean Fisheries Programme (OFP), Noumea, New Caledonia

1 Introduction

Close-Kin Mark-Recapture takes advantage of modern genotyping methods to identify pairs of close relatives (e.g. parent-offspring, half-brother/sister) among large collections of tissue samples (i.e. biopsies). The number of kin-pairs found, and the way they are distributed in space and time, can be embedded into a population dynamics model and used to estimate important demographic parameters such as absolute adult abundance, mortality rates, and connectivity; the fundamental idea is that every animal was born with exactly one living mother and one living father, which it "marks" genetically. Unlike conventional mark-recapture, CKMR biopsies can be taken just from dead animals, e.g. fishery catches; and unlike conventional fisheries data, CKMR can estimate absolute abundance directly, without needing to rely on catch rates (or, in extreme cases, even without reported catch totals). CKMR data can nevertheless be incorporated readily into integrated assessments, as with Southern Bluefin Tuna [SBT]; see e.g. Hillary et al., 2020.

For south Pacific Albacore, the inclusion of CKMR data has the potential to substantially improve the robustness of the stock assessment process. SC16-SA-IP-15-Alb (Bravington, Farley, et al., 2020) presented an initial scoping of feasibility (e.g. sample size requirements) and benefits. CSIRO and SPC are subsequently furthering that scoping analysis, so that the SC will have all necessary information to determine the merit and feasibility of implementation. Sample collection, genotyping, and modelling CKMR for SP albacore is a substantial task that will take several years. This Information Paper elaborates for SC17 a few issues that are specifically important for SP albacore, presents an interim update of design calculations, and summarizes briefly the remaining steps. The Appendix gives a general outline of how CKMR results might be taken up in assessment/management.

2 Age measurements: from length, epigenetics, and/or otoliths

CKMR requires information on the likely age of each juvenile and adult, so that kinship probabilities can be back-dated to juvenile birth and the likely fecundity/maturity of the potential parent at that time. Age estimates do not have to be perfect, but if the precision is poor then the model becomes unable to estimate abundance or other demographic parameters reliably (the parameters all become statistically confounded). Potential sources of information on an individual's age are its length, an age reading from otolith or other hard-part, and epigenetic age estimates from the same biopsy used for CKMR.

2.1 Length-based age estimates

For albacore, juveniles age 1–3 can be aged adequately just from length and time-of-year at capture, since the length-frequency modes are well separated (those few samples falling into ambiguous length ranges need not be genotyped). However, for adult albacore age 4yo and up, length is a poor predictor of age because there is rather little growth during adulthood, and so some kind of age measurement will be necessary, for at least a proportion of the genotyped adult samples.

2.2 Otolith ageing

Otoliths can be collected in quantity from *some* SP albacore fisheries via observer programs, and can give reliable annual age estimates for albacore (Farley, Williams, Clear, et al., 2013;Farley, Krusic-Golub, et al., 2021). However, logistics may prevent collecting otoliths in the quantities required for CKMR across the whole longitudinal range of adult SP albacore and any necessary fleet/flag configurations. The preparation/ageing of large numbers (potentially >10,000) of otoliths would also be a formidable challenge, and might require additional regional capacity.

2.3 Epigenetic ageing

Epigenetic age estimation just from biopsy tissue may provide an alternative to large-scale otolith reading. This technology[1] has emerged rapidly in the last couple of years; see Jarman et al. (2015), Horvath (2013), Mayne, Korbie, et al. (2020), and Mayne, Espinoza, et al. (2021). Mayne's approach in particular is highly automated and less expensive than otolith-reading, since the DNA already parepared and extracted for CKMR can be re-used directly. We have already seen encouraging preliminary results for another tuna species as well as other fish, and the technique is highly transferrable between species. The main necessity for each new species is a one-time calibration against known-age samples that have associated biopsy tissue, with a sample of 100–200 individuals.One key question about epigenetic age for albacore will be its precision; age estimates do not have to be perfect for CKMR, but precision and cost will affect sample size requirements.

2.4 Population age composition

Aside from ageing individuals, the adult age composition in the population is also important for CKMR, because the kinship probabilities involve a fecundity-weighted sum across population numbers-at-age. That information can be hard to obtain from fishery-independent[2] sources for many species where length-based selectivity (a fairly hard thing to estimate) affects the age-proportion in *samples.* For some of those species, including SBT, information on average adult age can still be gleaned from CKMR data itself based on the ratio of Half-Sibling Pairs (HSPs) to Parent-Offspring Pairs (POPs) (Davies et al., 2020); however, for SP albacore that approach would not be reliable because of possible confounding from spatial structure (see below). Fortunately, one advantage of limited adult growth within albacore is that population age composition of adults actually could be inferred from age-at-length samples. Although length-based selectivity may still operate across the adult size range, measurements of age *within* any adult length class should give a largely unbiased reflection of true age composition from 5yo or 6yo up. While 3yo and 4yos might still be subject to selection bias

[1] Specifically: age estimates based on DNA methylation proportions in particular gene-promoter sites. The methylation technology is completely different to the telomore-based approach used in several previous attempts— an approach which unfortunately does not seem to work.

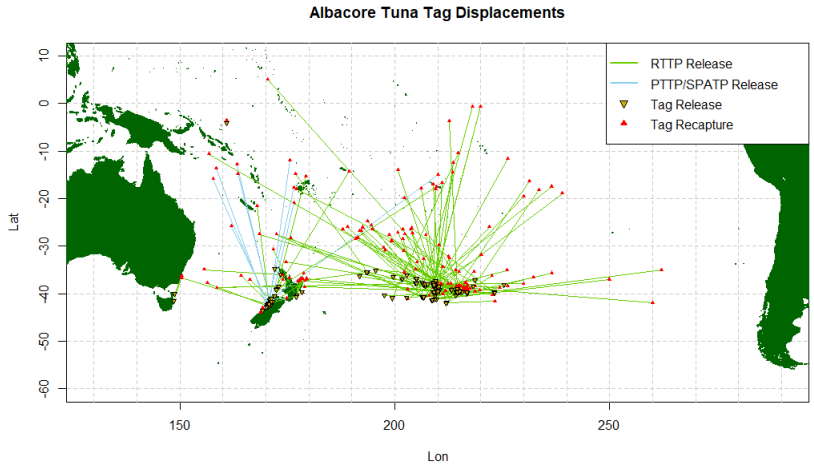
[2] I.E. not susceptible to *fishery-driven* effects, such as targeting, gear selectivity, or search behaviour.

(because they are smaller), the population age composition two or three years later gives retrospective information needed for CKMR.

In the framework we are using to test Designs (see section below), we assume that a modest age-composition sample equivalent to several hundred independent adults animals per year is available; that sample could be taken from any adult-oriented fishery using either otoliths or epigenetic age, and could of course simply be part of the CKMR sample at no extra effort. We also assume that reliable age composition data is available for catches.

3 Spatial considerations

Figure 3.1: Tagged albacore movements (from Castillo-Jordan et al., 2021)



CKMR needs some care when there may be persistent spatial structure within a "population". The details are complex, and there is no general-purpose rule, so case-by-case consideration is required. In the case of SP albacore, tagging data shows that the animals do move extensively East-West (Fig. 3.1), as well as ontogenetically from South to North as they grow to maturity. Albacore also spawn right across their wide longitudinal range, rather than in discrete "spawning grounds" (Farley, Williams, Hoyle, et al., 2013). Nevertheless, there remains the possibility (something which tagged-juvenile data alone can never resolve) that individual adults hold a specific *longitudinal* preference for spawning site— whether-or-not they stay near that longitude for the rest of the year, and whether-or-not that longitude is close to where they themselves were born. There is also some evidence from otolith microchemistry, and indirectly from length/age data, that lifetime movements may be somewhat restricted longitudinally (Macdonald et al., 2013). Population-genetic signals (Anderson et al., 2019) are somewhat equivocal (e.g. panmixia over evolutionary timescales at an ocean-basin level, but evidence of differences between years within the same location); but in any case "genetically detectable structure" is quite a different thing from "demographic separation within zero-to-two generations",

the latter being what matters for CKMR (and usually for management). Overall, the real situation is simply unknown: fine-scale persistent spatial structure is implausible given albacore's mobility, but equally it would be unwise to simply assume a total absence of spatial structure, i.e. that pairs of close-kin will be spatially independent[3].

3.1 Implications and detection of spatial structure for south Pacific Albacore

What might "spatial structure" actually mean for SP CKMR? First, some care is needed when designing the sampling; and second, the CKMR model itself *may* need to be adapted beyond a "naive" (completely non-spatial) formulation. Sampling is key, because a really-badly-designed program would be unable to deliver reliable estimates regardless of modelling efforts, and would not provide the signals required to determine whether a spatially-resolved analysis is necessary. The discussion below assumes that juvenile samples (1–3yo) come only from the NZ troll fishery, which at present seems the only source for the necessary large quantity of tissue samples for those ages; the issue is where the adult samples are taken.

As one example of why sampling design is important: supppose all the juveniles were collected near NZ, whereas all the adults were collected around French Polynesia and during the spawning season. Then, if adults do exhibit persistent longitudinal fidelity at spawning (PLF@S), there would be few POPs overall— many fewer than if similar samples had been drawn from "well-mixed" adults and "well-mixed" juveniles. But without any comparative data on POPs from adults sampled at other places, there would be no way to distinguish between "one huge well-mixed population" (i.e. no spatial structure) and "two much smaller sub-populations with limited mixing". A naive CKMR model that simply assumed the former would not be reliable— it would be potentially subject to an unknowable level of positive bias. Conversely, if the juvenile samples again came from near NZ but now all the adults were also collected only due North of the juvenile samples, then it would be impossible to decide whether "the" estimated adult abundance pertained only to an NZ-specific subpopulation, or to the entire WCP stock.

The way to avoid those pitfalls is obvious: don't take all the adult samples from one place, but rather spread them out spatially across the main areas fished. The total POPs found should not be very different to "ideal" sampling (from well-mixed adults) because the parents do, after all, have to be Somewhere. Thus a simple non-spatial POP-based absolute abundance estimate should provide reasonable results even in the presence of spatial structure. Formally, a simple CKMR POP model is unbiased as long as just *one* of the two sample-classes (either adults, or juveniles) be representative (i.e. taken in proportion to true spatial abundances)— having both representative is nice, but not essential. While it is probably impossible to ensure truly representative sampling, or to measure exactly the departures from representativeness (so that "reweighting" might be used to compensate), exactness may not be necessary to still get a useful aggregate estimate. Further, with well-designed

[3]There are many nuances around "spatial population structure", whether it be heritable, non-heritable, sex-linked, and/or across widely different timescales. Whether or not these matter to CKMR is quite case-specific, and depends on sampling as well as objectives. We do not try to define or separate them all here; the existence of "spatial structure" would just mean that there was some correlation between the spatial locations of kin-pair-members in the whole population.

sampling, the CKMR data itself will provide direct information on spatial structure, based on how much the POP rate (per comparison, i.e. allowing for different sample sizes in different places) varies in space and time. Table 1 shows some possibilities.

Table 1: Potential diagnostic patterns in POP rates against juveniles from NZ troll fishery (e.g. per million comparisons, standardised for age/sex/year). PLF means "Partial Longitudinal Fidelity" of adults, either throughout the year, or only during spawning ("@S"). Note that adults caught during a spawning season would only be compared to juveniles born in *previous* spawning seasons.

Type of structure	NZ adults during spawning	NZ adults outside spawning	Other adults in spawning	Other adults outside spawning
PLF	+	+	-	-
PLF@S	+	=	-	=
None	=	=	=	=

If a summary like Table 1 shows no evidence of spatial structure, or if it is evident but only during the spawning season, then unbiased estimation of total abundance is straightforward (in the latter case, by just using non-spawning-season adults). If there is a clear signal of year-round spatial structure (i.e. almost all parents found in the W rather than the E), then there would be various options: to make an estimate of just-Western abundance using just-Western adults if the POPs split cleanly enough (e.g. almost all in the W); to decide that adult sampling is sufficiently representative (perhaps after reweighting in line with catches or catch rates) to make a defensible aggregate estimate; to also incorporate spatial locations of Adult-Adult POPs, leading to an explicitly spatial CKMR model that estimates single-generation demographic exchange; to set up a deliberate sampling program of eastern juveniles with the same aim; etc. In any case, a strong spatial structure result would be important enough for management (in terms of whether single-stock management is tenable) to be well worth discovering in its own right.

Clearly, the total number of POPs actually found sets a limit on the ability to detect/rule out subtle spatial structure. However, weak spatial structure would not matter much for overall bias (given reasonable spread of sampling) and strong spatial structure would be apparent in a good design even with moderate numbers of POPs. For example, if equal numbers of adults were sampled from NZ/ western SPO as from French Polynesia/east-central SPO, and a total of 60 POPs were found overall, then a split as strong as, say, 40W/20E would be exceedingly unlikely by chance (well below $p = 0.01$) if there was really no spatial structure. Thus, if spatial structure is important enough to matter, it should be detectable given reasonably-well-spread adult samples.

What would the implications of spatial structure be for HSPs? Clearly, if juveniles are only sampled in one spot, then there is no possibility of looking at cross-site HSP rates (which would, if available, inform on adult "spawning site fidelity"). Also, the *number* of XHSPs (cross-cohort half-sibling pairs; pairs within the same cohort are not useful for CKMR) in the sampled juveniles only provides information about the abundance of whichever adult "subpopulation" spawns them. If all NZ juveniles are offspring of just half the adults— and, crucially, if it is the *same* half from year to year— then we would see about twice as many HSPs as if the juvenile sample was taken across the longitudinal range. Thus, in the context of estimating abundance for the entire SP Albacore population, the HSP count

is not directly informative about abundance or average adult size/age[4]. Nevertheless, XHSPs among NZ juveniles would still provide reliable information on (western/NZ) adult Z; and the *total* number (compared to the number of POPs) provides some further information on the extent of spawning-season spatial structure.

Until there is some reasonably-informative CKMR data, there is no point in building explicitly-spatial CKMR frameworks to cover a huge range of possibilities— a task that would vastly complicate the design process without adding much clarity. Instead, since the above qualitative arguments suggest that useful estimates could be made with *reasonably* representative adult sampling plus spatially-restricted juvenile sampling (and that the number of POPs would be fairly insensitive), a sensible approach is: to build a non-spatial model (except allowing for "spatial" HSP effects as just discussed) aimed at estimating aggregate abundance; to ensure an adequate spatial range of adult samples; and to confirm that the expected overall number of POPs is enough to allow strong effects to be detected.

3.2 Sampling opportunities

SPC has summarized WCPFC-held information for SP Albacore to characterize the quantity, location, and fleets of catches and landings. Preliminary indications are encouraging in terms of potential sample sizes. After completion of the experimental field studies described next, the summaries will be used to evaluate feasibility of potential port and at-sea sampling strategies, and to propose some cost-efficient and practicable sampling schemes for SP albacore CKMR.

4 Field studies

4.1 Calibrating Epigenetic Age

The SC's previous albacore project (Project 106; Farley, Krusic-Golub, et al., 2021; SC17-SA-IP-10) estimated the decimal age of 660 albacore in the range 43-115 cm fork length from collected otoliths. Corresponding tissue biopsies for these samples are held in the SC's Pacific Marine Specimen Bank and in CSIRO archives. CSIRO and SPC will use these samples for the epigenetic age calibration. SC17 is advised that a request will be forthcoming to withdraw these samples from the Pacific Marine Specimen Bank.

4.2 Standard Operating Procedures for genetic sampling

A crucial part of any CKMR study is the efficient collection and storage of high-quality, uncontaminated tissue. This is important for all types of marine genetic studies, and with albacore there are

[4]The expected number of HSPs is affected not just by adult "SSB" but also average adult size, in that a population of fewer larger adults will lead to a lot more HSPs than would be found in a larger population of smaller adults with the same "SSB". Thus, in SBT for example, the main information provided by the *number* of HSPs is on average adult age, rather than on "SSB" per se. Details are beyond the scope of this report.

particular concerns about potential for DNA degradation, depending on freshness and preservation of tissue. We have thus taken the opportunity to develop best sampling practices for SP albacore, which can be generalised into Standard Operating Procedures for fisheries-based genetic sampling. Specifically, CSIRO and SPC have implemented a trial with the assistance of New Caledonia to evaluate tissue quality and degradation rate using various tissue extraction methods (including biopsy punches and CSIRO's gene-tagging tool) and different types of sample: frozen on capture, fresh on capture, delayed sampling of fresh tuna up to 10 days after capture, etc. The field sampling has been completed, with genetic analyses expected in the second half 2021. The derived SOPs will be presented as part of the feasibility analyses for consideration at SC18. The SOPs are expected to have broader application for genetic sampling beyond the CKMR application.

5 CKMR model and interim results

Feasibility calculations in SC-16-SA-IP-15 were based purely on trying to find sample sizes/strategies that might yield respectable number of POPs and XHSPs (e.g. at least 50 of each). Respectable totals are a necessary but not a sufficient condition for CKMR; a badly-enough-designed project (say with just a single cohort of juveniles, or only sampling young adults) might yield a decent number of POPs but still give disappointingly high CVs on quantities-of-interest.

To address this, we coded a simplified mostly-standalone CKMR analysis (taking POP, HSP, and some age composition data as input, and estimating abundance and mortality rates) and experimented with different sampling strategies in terms of age breakdown (i.e. size stratification), duration, and total sample size (as per the section "*Realistic consideration of achievable precision in a stock assessment context*" in SC-16-SA-IP-15). So far we have considered 3-year and 4-year sampling schemes (there was little difference between these in overall CV for the same *total* sample size); we will also explore compressing the sampling timeframe further (thus getting results sooner), but the logistics may become overwhelming if trying to collect too many samples too quickly. We again took "truth" to be the 2018 stock assessment outputs, and assumed steady-state F and average recruitment for the subsequent period through the simulated CKMR study (although the new model could accommodate non-equilibrium dynamics). To calculate likely CVs, we applied the statistical theory in section 4 of Bravington, Skaug, et al. (2016).

The model makes minimal external assumptions (the only non-CKMR inputs are catch-at-age, weight-at-age and proportion-mature, plus some adult age composition data). The main differences from last year's rough feasibility calculation are listed here and in Table_2:

- HSPs are no longer assumed to reflect entire population, but only an unknown proportion of it. That requires introducing an extra parameter to be estimated; but at the same time, if spatial structure is present then it will *increase* the number the number of HSPs actually found (because each sampled juvenile from a limited area now has fewer potential parents to "choose from").

- Explicit parameter estimation for all population-dynamic parameters and derived quantities (e.g. Spawner-Per-Recruit-Ratio at current F-levels).
- Estimated parameter describing the fecundity/size relationship. Note that external estimates of the *female* fecundity/size relationship are available for SP albacore, in Farley, Williams, Hoyle, et al. (2013). We have not considered those here, assuming instead that fecundity will be estimated from CKMR data inside the model itself, but the use of external estimates could be considered in further work.
- Sex-disaggregated estimation (except for M); abundances and fecundity parameters are estimated separately for males and females.

	What gets estimated (omitted if obvious)	"Truth"	Notes
Adult age composition in population (4yo+)			equivalent sample size 250 p.a.
XHSPs	Scaling factor to absolute abundance (ie *not* assumed representative of all offspring, unlike e.g. SBT)	"NZ spawning adults" constitute 50% of entire population	
M	One value, assumed equal for 3+yo	0.3	
Sex ratio		50/50 ratio	catches assumed to split in proportion
Fecundity/size relationship	α the exponent in $\text{fec}_{\text{age}} \propto w_{\text{age}}^{\alpha} \times \text{ppnmat}_{\text{age}}$	$\alpha = 1$	no dimorphism; $w_{\text{age}}$and $\text{ppnmat}_{\text{age}}$as per 2018 assessment
Age and length		Age assumed known exactly for all samples; length assumed to have no effect per se once age is known	A CKMR analysis for real data would use both age *and* length; however, age-only models are simpler, and can be useful approximations at the design stage.

Table 2: Some structural features of our current CKMR model

Preliminary investigation of a number of possible designs suggests that respectable precision can be achieved by the end of a 3-year sampling program with about 25,000 total samples— assuming the 2018 assessment is accurate (see Discussion). Table 3 presents one example, together with a few quantities of likely interest. The most precisely-estimated quantity is adult biomass, and this has a good CV of 12% (at the start of the study, but estimated retrospectively at the end). TRO (Total Reprodutive Output)— a preferred alternative to SSB, that allows more faithfully for fecundity effects that are not directly proportional to weight— is less precisely estimated, presumably because (i) it is based here on females only, thus on only about half the number of POPs compared to adult biomass, and (ii) it will depend on the estimated fecundity/size relationship at 3yo and 4yo. While M and SPRR are not very precise, their uncertainties are *quantifiable* (unlike at present) and can in principle be allowed for in management. Continued CKMR sampling in subsequent years could improve those precisions to any desired extent. Of course, the *achieved* precision for all quantities

will differ in practice from the Table: partly because allowance *may* have to be made for spatial structure, depending on what the kin-pairs reveal; but mostly because of inevitable inaccuracy in the asssessment outputs that this analysis is generated from, abundance in particular. However, the general rule for CKMR in that situation is that there will either be vague good news, or precise bad news. That is: if true abundance is higher than assumed, there will be fewer kin-pairs and worse precision, but it will be clear that stock status is better than was thought; but if true abundance is lower, there will be plenty of kin-pairs to demonstrate very clearly that the status is actually worse.

Type\Size	Annual	Total 2023–2025	Notes
Adults 4yo+	3990	11970	Proportion-at-age matches catch for 6yo+; heavy undersampling for 4yo/5yo; no 3yo
Juveniles 1–3yo	4250	12750	62% 1yo; 31% 2yo; 7% 3yo
Adult age composition	(250)	(750)	(included in adult samples)

Expected kin-pairs	
POP	54
XHSP	49

Precision	Biom4up_23	TROeq6_21	TROeq6_23	TROeq6_25	Log trend (annual)	M	SPRR
"Truth"	44,000T	1,300,000	1,300,000	1,300,000	0.0038	0.30	0.40
SE					0.10	0.12	0.11
CV%	12	29	26	39			

Table 3: One possible design, and example precisions predicted if 2018 assessment is accurate. "Biom" is adult biomass; "TROeq6" is Total Reproductive Output in units of "equivalent 6yo females". Figures after underscore indicate year, so that "_23" means "in 2023". SPRR is Spawner-Per-Recruit-Ratio at estimated current F-at-age in 2018.

The sampling rate in Table 3 amounts to non-destructively sampling about 1% of the catch (1–3yo) from the NZ troll fishery, and about 0.3% of all adult (4yo+) catches, annually over a 3-year period (based on recent catches approaching 2018).

An important assumption behind Table 3 is that the NZ juveniles are linked to a persistently-Western-breeding subset of adults, assumed here to be 50% of the whole adult population. That doubles the number of XHSPs compared to an ideal scenario of "representative juvenile sampling", and thus might be seen as unfairly *good* for precision. On the other hand, if the number of XHSPs turns out to be consistent with "not much spatial structure", then there is no need for the extra XHSP scaling parameter (i.e. it could be set to 1), and it turns out that CVs would in fact be *better* in that case.

One conclusion from investigations so far, is that it will be worth heavily undersampling smaller adults, since these have much lower lifetime-fecundity-prior-to-capture and yield many fewer POPs per genotype (i.e. per sampling dollar). Although it is very important in CKMR to have data for estimating fecundity for all adult size/ages (including 3yo and 4yo for albacore; and for males as well as females), that will happen automatically even if only somewhat older adults get sampled, because of back-dating; a 6yo adult at time of sampling was 4yo two years previously, when some of the juvenile sample was born, so it provides information on fecundity at younger ages than 6 as well.

The high mortality on adult albacore means there is a preponderance of young adults in the catch, so that unstratified sampling of adults would be very inefficient. Since any stratification decisions will ultimately have to be length-based, further work is needed to decide those details. Table 3 assumes there is enough information about age contained in length data to allow heavy undersampling of 4yo and 5yo adults relative to 6yo-and-up, and to avoid sampling 3yo adults altogether.

We have deliberately not yet tried to explore design space thoroughly, nor to optimize the design (noting that different objectives might suggest different optima anyway— for example, estimates of trend may benefit from wider coverage of juvenile cohorts, whereas estimates of absolute average N may be most precise when sampling is concentrated onto fewer juvenile cohorts, since that tends to provide the largest number of POPs). There are several reasons to defer full calculations until 2022:

1. Wait for updated assessment results.
2. Need to know the precision of epigenetic ageing (if precision turns out to be poor, a more elaborate model also involving length may be required).
3. Need more detailed understanding of fleetwise sampling feasibility (linked to results on epigenetic age and DNA quality).

6 Discussion

This year we have developed a much more sophisticated CKMR model (like a simplified stock assessment) for SP albacore that can be used for quantitative investigation of possible designs. Overall, our interim results from this model confirm last year's feasibility check, suggesting that good precision on quantities-of-interest could come from a sample size of 25,000 spread over 3 years— *if* the 2018 assessment outputs are roughly correct (as per Tremblay-Boyer et al., 2018). (If not, the design will still yield either "vague good news" or "precise bad news".) Although the possibility of spatial population structure does need to be considered for SP albacore CKMR, its extent can be evaluated after-the-fact from CKMR data; and for SP albacore it seems unlikely to cause a major problem (certainly not an *undetectable* bias of the type that can occur with, say, CPUE when assumptions are violated) provided that the adult samples can be well-spread longitudinally, even if the juvenile samples are drawn exclusively from the NZ troll fishery. CKMR sampling is non-destructive, requiring only a tiny amount of tissue, and the proposed sampling rate amounts to about 1% of catches. Otoliths would be useful, though it might be difficult to collect/process sufficient quantities; however, epigenetic ageing— a new technique which, in preliminary results, has worked well for other tuna— is likely to be key in achieving that longitudinal coverage because only tissue samples are required and they can be taken at port-of-landing from some fisheries. Thus, calibration of epigenetic ageing for albacore, and determining its precision, is a key task over the coming year prior to finalizing a realistic design. So too is ensuring protocols for DNA quality and efficient biopsy collection, which is the subject of ongoing experimental work. We also need to consider in detail the fleet-specific sampling options and stratification issues, and to incorporate results from the updated assessment this year. There is no need,

however, to investigate genetic markers for kinship at this stage, however, because that has become a routine and reliable process at CSIRO through CKMR projects on other species (about 10 so far). We expect to complete all these tasks in time to present a complete set of design options for SPC18 in 2022.

The Scientific Committee has recognized that it is just as important to build the understanding of WCPFC around CKMR applications, as it is to do the modelling work around checking feasibility and designing efficient CKMR projects. This communication process was originally envisaged for the margins of WCPFC meetings. Given that regional travel restrictions are unlikely to change in the immediate future due to Covid-19, we propose instead a series of online seminars. The first, on CKMR in general and what it can and cannot address, will be held in September 2021 (contact Simon Nicol at SPC for details and registration). A second seminar will be held in May 2022, to present the outcomes of the feasibility study for South Pacific albacore.

Appendix: using CKMR results

The details of CKMR modelling are well beyond the scope of this document, but it may be helpful to see the following general notes on the nature of "the CKMR model" itself, the relationship to "stock assessment", and the way the results might be used.

1. There is some qualitative information that comes straightaway from CKMR data, primarily on connectivity/spatial structure based on the locations of close-kin pairs. This can have a direct impact on how assessment and management is approached, e.g. if clearly separate stocks are identified.

2. Despite the simplicity of the underlying idea that an offspring "marks" its mother and father at birth, most of the information contained in CKMR data (e.g. about absolute abundance) is linked to multiple parameters in complex ways, and can only be extracted by fitting a proper statistical model; it is not possible, for example, to "just come up with an abundance index" that could be reliably plugged into a separate stock assessment. Thus, CKMR analysis entails coding from scratch a statistical model that is able to directly handle CKMR data. The basis of that model is an age-structured population-dynamics framework as used in many stock assessments worldwide, covering the period of juvenile cohorts that were sampled, and the model can estimate most of the familiar biological/demographic parameters seen in stock assessments, although generally *without* having to include "non-biological" phenomena such as selectivity. The model might be referred to as a "CKMR-based stock assessment model" and/or as a "CKMR analysis". For certain species and sampling arrangements, it might be necessary to explicitly incorporate some spatial aspects; the background biology, and CKMR data themselves, should indicate whether that is necessary (which remains to be seen for SP albacore).
   It is beneficial though not completely essential to include catch data[5] in the CKMR analysis too,

[5]Without total catch data, it is still possible to estimate absolute abundance from CKMR, but not M— only average

but beyond that it is a matter of choice about how much to include of other conventional fisheries data such as CPUE. Starting with just CKMR and catch data and as little else as possible is a good idea, because it is reasonably simple to code such a model and it is safe from potential difficulies associated with CPUE etc. If there is an existing stock assessment process, then estimates from a CKMR standalone analysis can quickly serve as a ground-truthing constraint on plausibility, and on the reliability of other datasets[6]. That is the cue for development of an integrated stock assessment model, incorporating CKMR data alongside whichever other data are required. It is again a matter of choice whether to start from the CKMR analysis and add in other data, or to start from an existing stock assessment model and add in the CKMR data component. The "design" discussed in this document is aimed at a standalone CKMR-based analysis, with minimal use of other data apart from total catches.

3. The nature of CKMR data, where basically every sample gets compared to every other sample including ones collected in previous years, is that most of the pairs will only be found towards the end of the study. For example, after 2 years of a 3-year collection, we would expect less than half ($\frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$) of the ultimate number of pairs to be available, so interim precision would be poor (and might not even allow a model to be built). Thus, powerful quantitative results only come at the end. However, if early indications are that the rate of kin-pair-finding is substantially lower than expected (presumably implying that the assessment used in designing CKMR underestimated the true abundance), then sampling can be increased in remaining years to ensure that enough kin-pairs will be available for precise inferences.

4. Once a CKMR analysis has established a baseline for abundance and other demographic parameters, it may be desirable to continue sampling, in order to refine parameter estimates (e.g. M) and to have an ongoing fishery-independent absolute abundance estimate for management. That decision can be left until baseline CKMR results are available.

References

Anderson, G, Hampton, J, Smith, N, and Rico, C (2019). "Indications of strong adaptive population genetic structure in albacore tuna (*Thunnus alalunga*) in the southwest and central Pacific Ocean". *Ecology and Evolution* 9.18, pp. 10354–10364. DOI: https://doi.org/10.1002/ece3.5554.

Bravington, M, Farley, J, Hampton, J, and Nicol, S (2020). *Preliminary analyses for a Close Kin Mark Recapture feasibility study in WCPO*. Tech. rep. WCPFC-SC16-2020/SA-IP-15. WCPFC.

Bravington, MV, Skaug, HJ, and Anderson, EC (2016). "Close-Kin Mark-Recapture". *Statistical Science* 31.2, pp. 259–274. DOI: 10.1214/16-STS552.

Z. But if catches (and their breakdown by size) are not reasonably well-known, then there are significant management problems going well beyond stock assessment.

[6]That is roughly the path taken with SBT, where standalone CKMR results (in 2013 and updated in 2018) led to refinement/rejection of scenarios within the "Operating Model grid". Nowadays the CKMR data coexists happily with other datasets in assessments/operating-model-conditionings, where it has provided powerful information on absolute abundance, adult mortality rate, and the age/fecundity relationship. Ongoing CKMR data is also incorporated directly in the Management Procedure used by CCSBT for triennial TAC-setting.

Castillo-Jordan, C, Xu, H, Hampton, J, Ducharme-Barth, N, and Hamer, P (2021). *Stock assessment of South Pacific albacore tuna*. Tech. rep. WCPFC-SC17-2021/SA-WP-XX. WCPFC.

Davies, C, Bravington, M, Eveson, J, Lansdell, M, Aulich, J, and Grewe, P (2020). *Next-generation Close-kin Mark Recapture: using SNPs to identify half-sibling pairs in Southern Bluefin Tuna and estimate abundance, mortality and selectivity*. Tech. rep. CSIRO. URL: https://www.frdc.com.au/Archived-Reports/FRDC%20Projects/2016-044-DLD.pdf.

Farley, J, Williams, A, Clear, N, Davies, C, and Nicol, S (2013). "Age estimation and validation for South Pacific albacore *Thunnus alalunga*". *Journal of Fish Biology* 82.5, pp. 1523–1544. ISSN: 0022-1112. DOI: 10.1111/jfb.12077.

Farley, J, Krusic-Golub, K, and Eveson, P (2021). *Ageing of South Pacific Albacore*. Tech. rep. WCPFC-SC Project 106. WCPFC.

Farley, JH, Williams, AJ, Hoyle, SD, Davies, CR, and Nicol, SJ (2013). "Reproductive dynamics and potential annual fecundity of South Pacific albacore tuna (*Thunnus alalunga*)". *PLoS One* 8.4, e60577.

Hillary, R, Preece, A, and Davies, C (2020). Tech. rep. CCSBT-ESC/2008/BGD07. Commission for the Conservation of Southern Bluefin Tuna.

Horvath, S (2013). "DNA methylation age of human tissues and cell types". *Genome biology* 14.10, R115.

Jarman, SN, Polanowski, AM, Faux, CE, Robbins, J, De Paoli-Iseppi, R, Bravington, M, and Deagle, BE (2015). "Molecular biomarkers for chronological age in animal ecology". *Molecular Ecology* 24.19, pp. 4826–4847. ISSN: 1365-294X. DOI: 10.1111/mec.13357.

Macdonald, JI, Farley, JH, Clear, NP, Williams, AJ, Carter, TI, Davies, CR, and Nicol, SJ (2013). "Insights into mixing and movement of South Pacific albacore *Thunnus alalunga* derived from trace elements in otoliths". *Fisheries Research* 148, pp. 56–63. ISSN: 0165-7836. DOI: https://doi.org/10.1016/j.fishres.2013.08.004.

Mayne, B, Espinoza, T, Roberts, D, Butler, GL, Brooks, S, Korbie, D, and Jarman, S (2021). "Non-lethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod". *Molecular Ecology Resources* 00, pp. 1–9. DOI: https://doi.org/10.1111/1755-0998.13440.

Mayne, B, Korbie, D, Kenchington, L, Ezzy, B, Berry, O, and Jarman, S (2020). "A DNA methylation age predictor for zebrafish". *Aging* 12.24, pp. 24817–24835. ISSN: 1945-4589. URL: https://pubmed.ncbi.nlm.nih.gov/33353889.

Tremblay-Boyer, L, Hampton, J, McKechnie, S, and Pilling, G (2018). *Stock assessment of South Pacific albacore tuna*. Tech. rep. WCPFC-SC14-2018/ SA-WP-05. WCPFC.