Advice on Close-Kin Mark-Recapture for estimating abundance of eastern Atlantic bluefin tuna: a scoping study

Campbell Davies, Mark Bravington and Robin Thomson

GBYP 07c/2015 - ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (ICCAT GBYP – PHASE 5). Updated.

CSIRO Marine Laboratories, Castray Esplanade, Hobart, Tasmania, AUSTRALIA

campbell.davies@csiro.au

1 Executive Summary

Close-kin Mark Recapture (CKMR) is a new approach to estimating abundance and other important population parameters with demonstrated applicability to the highly migratory southern bluefin tuna fishery. This project was commissioned by ICCAT - GBYP to scope the potential application of the CKMR approach to the eastern stock (EBFT) of Atlantic bluefin tuna (ABFT), including a brief review of the concepts, its application to others species, salient technical considerations, current sampling programs under the ICCAT GBYP, and initial recommendations on the design of a pilot study to determine the feasibility of the approach for EBFT.

Close-Kin Mark-Recapture uses information on the frequency, and distribution in space and time, of closely related individuals in samples of tissue from live or dead animals. The first large-scale application was for southern bluefin tuna (SBT), where it was developed as an absolute abundance estimator independent of commercial catch per unit effort (CPUE) and total catch data. The SBT application was relatively simple, in that: SBT is a single population with one know spawning ground; much of the population biology is well documented;, and existing monitoring systems were in place that facilitated the provision of high quality length, age and tissue samples of known spawning adults and juveniles. An application to EBFT poses a number of challenges, including: east-west population structure across the Atlantic and possible structure within the Mediterranean, which require more complex sampling designs and estimation models; less biological background knowledge; and substantially more complex logistics/operational environment.

Our review of previous applications of CKMR highlights two central considerations for EBFT. First, It has been extended and generalised beyond the Parent-Offspring-Pairs (POP) used in the SBT case, to include more distant kin (e.g. Half-Sibling-Pairs), which reduce the sample size requirement (because for a given sample size, the total number of kin-pairs found will be larger), and reduce the need for untes

assumptions and/or extra biological information e.g. about fecundity-at-age. Second, a "naive" carbon-copy of the SBT approach to a species that (unlike SBT) may have substantial within-population structure (i.e. spawning-ground fidelity of some kind), could lead to badly biased estimates. However, in this report we demonstrate that a more sophisticated version of CKMR, using POPs and HSPs and sampling in multiple locations, can solve the problem. Specifically, from CKMR it is possible in principle to identify "management relevant" structure in populations, and to estimate the relative contribution of "spawning units" to effective reproductive output of the population as a whole (i.e. the quantity of primary concern to fisheries management). The latter does not require the existence of a genetic marker, in the conventional population genetics sense; rather, the nature of structure and the extent of mixing can, in principle, be estimated from the distribution of POP and HSP among spawning and juvenile grounds.

Based on a review of the relevant literature, the GBYP sampling programs and communications with ICCAT ABFT scientists, we consider that CKMR should be feasible for EBFT, assuming it is possible to: (i) increase the annual sample size of tissue, otolith and length samples obtained from within Mediterranean and eastern/central Atlantic sampling programs; (ii) distinguish between individuals of eastern and western origin with a high probability; and (iii) implement high quality sample, processing and data management programs to minimize the likelihood of genotyping errors. To

demonstrate statistical feasibility and to broadly investigate sample size requirements, we developed an age-structured, multiple-population CKMR model and used current estimates of EBFT population parameters consistent with the most recent ICCAT stock assessment for a simple 2-spawning ground and 2-juvenile ground example. We used the model to explore a range of sampling designs, covering factors such as total sample size, split of samples between adults and juveniles, assumption about age-structure of the adult samples, and length of sampling program in years. Assuming a primary design criterion of a CV of around 15% on the estimated 2014 spawning biomass, it appears that the desired CV might be obtainable for total sample sizes (i.e. adult and juveniles) in the order of ~30,000-40,000 individuals. The total number required should not depend too much on the actual number of spawning and juvenile grounds, but will depend somewhat on the duration of the study (we considered 3, 4, and 5 year designs) and other design details (e.g. what size of adults to concentrate on genotyping). More importantly, though, the *actual* number of samples required may well turn out to be considerably different, because the true stock size and other true biological parameters (including the nature of any population structure) could well be quite different from (i) the current stock assessment results that we based the calculations on, and from (ii) other assumptions (e.g. about mixing proportions) that we had to make in order to explore possible designs. Sample sizes can be adjusted as the study goes on and knowledge accumulates (just as happened for SBT), especially if extra samples are collected (usually cheap) but not genotyped (usually less cheap) in the first pass, but are available subsequently for genotyping if sample sizes need to be increased (in order to find enough kin-pairs to make a reliable estimate).

Because of the many uncertainties, it is not possible to provide specific costings for a CKMR study at this stage. However, based on these sample size calculations, the cost (excluding the cost of obtaining biological samples) of the orignal SBT application and reductions in the cost of marker development and large-scale genotyping since then, we would expect the annual cost to be in the order of Euro 250-300k per annum for the period required to provide a first estimate. For specific mathematical reasons (and unlike an annual trawl survey), CKMR is most efficient when used not as a "one-off" estimator, but rather as part of a time series, whereby abundance estimates are updated (e.g. as is now planned for SBT). If a CKMR program for EABT were to continue after the first few years, it is entirely reasonable to expect sample size requirements to decrease and the ongoing annual cost to decline further.

Given this, we conclude that there is scope for CKMR to significantly improve the data and understanding available to effectively assess the status of EBFT, and EBFT in particular. Assuming there are sufficient resources and institutional commitment to modify and expand the current level of biological sampling completed under the GBYP to the level required to obtain an informative number of close-kin (POPs and HSP) and associated ancillary data, then we recommend the following activities in order of priority:

1. Determine the most cost-effective form of genotyping that can demonstrably identify HSPs. By cost-effective, we mean the GBS (Genotyping-By-Sequencing) method that can provide the required level of genotyping reliability required to consistently identify HSP for the lowest cost per fish (Note: if the method can do this for HSP, it can necessarily do it for POPs.)

2. Consideration should be given to doing 1 in conjunction with a workshop that includes expertise from a range of other areas that are active in large-scale, high through-put genotyping for applied fisheries and/or natural resource management purposes (e.g. Pacific Salmon, the FishPopTrace Consortium, GBYP Biological Program Consortium, CSIRO) to learn from their experience and share the cost involved in evaluating alternative GBS platforms in a very rapidly developing and technically complex field.

3. In consultation with GBYP Biological Program and SCRS BFT WG, select juvenile and adult sampling locations for an "initial round of CKMR sampling", which are consistent with the current understanding of spawning units and juvenile grounds, and initiate sample collection as soon as possible. These samples can, in the short-term, be archived and, or, used to develop genotyping and data processing work-flows and quality control procedures for identifying kin; genotyping itself can happen later.

4. Commence an inclusive, expertise-based process to review and identify candidate markers (genetic and/or micro-chemical) for assigning samples to eastern and western populations. While it may be appealing to include "within Med" markers as part of this exercise, it is not necessary for the purposes of CKMR, and there is no virtue in waiting for the (unlikely) outcome of a within-Med marker search before starting CKMR. As noted in section 6, the CKMR data will reveal any possible population structure in the Mediterranean, as long as the sampling of spawning grounds and juvenile areas is sufficiently comprehensive. The final E-W candidate(s) markers, including assignment probabilities, should be decided based on a validation study conducted using known origin fish of sufficient sample sizes to provide statistically reliable estimates of assignment probabilities.

5. Finally, it is important to recognise that design and implementation of CKMR requires a combination of both broad (fisheries biology, field and laboratory logistics, statistics, mark-recapture theory, population dynamics, population

genetics and genomics, applied stock assessment) and deep knowledge and expertise (in this case, in ABFT population biology and fisheries, CKMR design and implementation). CKMR data will not fit into a VPA. Hence it will be important to establish close linkages with the development of new assessment methods and the MSE work program of the GBYP and broader ICCAT assessment process to ensure the greatest benefit is obtained from the data and information that would be provided by such a program. There is a very substantial process of statistical and programming development required for both the stand alone CKMR assessment model and the incorporation of the CKMR results into an integrated stock assessment (see Hillary et al 2012, 2013). CKMR itself is quite new, and the extension to population-structured settings like EBFT is completely new; in these (relatively) early stages of development and implementation, it will be important to consider the best mechanism (contracting and institutional) to establish and maintain a suitable experienced and qualified team for design and implementation to deliver high quality and robust results in the short-term and, if successful, the development of the necessary capability to maintain an ongoing program into the future.

2 Introduction

Abundance estimation is a fundamental challenge in ecology and the management of harvested populations. This is particularly the case for highly migratory species, such as tunas. Conventional stock assessment methods based on commercial Catch Per Unit Effort, are plagued by well-documented issues associated with changes in the spatial and temporal effort distributions of international fleets and fishing and reporting practices that make it extremely difficult, if not impossible, to derive unbiased abundance indices from these forms of data (Polacheck and Davies, 2008; Polacheck, 2012a; Maunder et al., 2006). Alternatives to CPUE-based methods, such as conventional tagging experiments, suffer from issues related to spatial and temporal coverage of releases and of recapture effort, and can be rendered worthless by high levels of non-reporting and/or inadequate estimates of tag reporting rate (e.g. Polacheck and Eveson, 2007; Davies et al., 2007). These examples, which are grounded in decades of wider experience in international tuna stock assessment and management, clearly demonstrate the need for new approaches to abundance estimation that are independent of the biases associated with commercial CPUE and conventional tagging data.

Close-kin Mark Recapture (CKMR) is an approach with demonstrated ability to meet this need for an internationally managed tuna fishery (Bravington et al., 2012; Bravington et al., 2014; Bravington et al., 2015; CCSBT ESC, 2013; CCSBT ESC, 2015). It has delivered cost-effective, direct estimates of adult abundance and spawning potential. CKMR has been adopted by the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) to provide a regular time-series of the spawning potential of the SBT population. (CCSBT, 2015).

The ICCAT Grand Bluefin Year Program (GBYP) aims to increase the understanding of Atlantic bluefin tuna (ABFT), improve the data available for stock assessment and provide modelling tools to conduct improved stock assessments and management strategy evaluation (MSE) of Harvest Control Rules and/or Management Procedures. In light of the issues noted above, the GBYP Steering Committee recommended that ICCAT invite proposals to review the application of CKMR and scope its potential application to EBFT, with a specific focus on the eastern stock (ICCAT, 2015). The terms of reference for the project (TAGGING PROGRAMME – ADVICE ON CLOSE-KIN GENETIC TAGGING STUDY ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (ICCAT GBYP – PHASE 5) are provided in Appendix 1.

This report is structured into five main sections:

1. A brief overview of CKMR studies on other species.
2. A summary of the salient technical considerations associated with designing CKMR projects and, in particular, how the complications associated with stock structure can be addressed.
3. A review of tissue sampling programs conducted through the GBYP and their suitability for CKMR.
4. A preliminary experimental design for a pilot study to establish the feasibility of CKMR for EBFT.
5. Summary and recommendations.

3 Overview of applications of Close-Kin Mark Recapture

3.1 SBT: Stand-alone abundance estimation

Close-Kin Mark-Recapture is a suite of methods to estimate abundance of adults and other important demographic parameters (Skaug, 2001;Bravington et al., 2016) using information on the frequency of closely related individuals in samples. The first large-scale application was for southern bluefin tuna (Bravington et al., 2012; Bravington et al., 2014), where it was developed as an abundance estimator independent of commercial catch per unit effort (CPUE) data (and indeed independent of total catch too). The impetus to do so was three-fold:

1. There was no direct index of abundance for the spawning stock, i.e. the mature component of the population. Instead, it was extrapolated from sub-adult abundance estimates via the stock assessment model;

2. There were (and are) unresolved issues associated with statistical methods and interpretation of longline CPUE as an index of immature abundance (CCSBT OMMP, 2014);

3. There were revelations of large, long-term, unreported catches from the longline fisheries which generated unquantifiable uncertainty (CCSBT ESC, 2006; Polacheck, 2012a; Polacheck, 2012b) to the extent that the Extended Scientific Committee (ESC) of the CCSBT could no longer conduct a stock assessment in the conventional sense (CCSBT ESC, 2006). The last point, in particular, increased the urgency for developing more reliable sources of abundance information for the spawning stock, which is the primary focus of the CCSBT rebuilding plan.

The SBT application used specifically designed microsatellite loci (Bravington et al 2014) to identify Parent-Offspring-Pairs (POPs) in about 14,000 samples of known spawning adults (Indonesia) and known-age juveniles (Great Australian Bight). These were embedded in a statistical mark-recapture framework, and combined into a stand-alone mini-assessment of adults that used length and age composition data from Indonesian longline catches on the spawning ground, plus histological information on relative *daily* fecundity-at-size. The model was able to estimate a time-series of absolute spawning stock biomass, effective *annual* fecundity-at-size[1], and total mortality rate of the mature component of the population. Full details of the sampling design, marker development, genotyping, quality control, procedures for identifying POPs, estimation model, and independent review process are provided in Bravington et al. (2014). The approach and the final results were reviewed by the CCSBT ESC in 2012 and 2013 and accepted as: (i) a valid fishery-independent[2] estimate of spawning stock abundance and spawning potential for SBT, and (ii) as valid input data (the POP information) for the CCSBT Operating Model (Hillary et al., 2012; Hillary et al., 2013; CCSBT ESC, 2013).

Although CKMR for SBT was technically challenging because it was so novel and had to be developed entirely from scratch, with hindsight it was relatively straightforward, for several reasons: a solid understanding of the underlying population biology; a single fishery on spawning adults, which covers the whole spawning season and region; an existing infrastructure for collecting length/age/sex/genotype samples from that fishery; no stock structure complications; and straightforward sampling of juveniles from readily identifiable cohorts. Not all of these apply to other species, and if not then the particular way in which CKMR might be used needs case-specific consideration. As shown later in this report, naive mis-application of the SBT CKMR model to other species could simply give the wrong numbers, especially in the presence of stock structure.

3.2 SBT: Close-kin Mark Recapture in Operating Models

The CCSBT Operating Model (OM) is a set of integrated statistical-catch-at-age models used for development and testing of Management Procedures (MP) (Hillary et al., 2015and periodic assessments of stock status (e.g. Preece et al., 2014; CCSBT ESC, 2014). The unquantifiable uncertainty resulting from the unreported catches means that a variety of historical-catch scenarios, provided by the CCSBT, are used to scale the standardised CPUE from the reported catch and effort data from the primary longline fleet (CCSBT, 2009; CCSBT ESC, 2014); this CPUE of largely juvenile and sub-adult age classes was (and is) the primary abundance index in the OM. The reason for having a set of models, rather than just one, is to accommodate different scenarios about historical catch and other structural uncertainties. Other sources of abundance information include conventional tagging data from the 1990s, and a relative abundance index of juveniles from a scientific aerial survey from 1993-2014 (e.g. Eveson et al., 2012).

[1] IE relative annual production of surviving juveniles, from adults of different size (by sex).

[2] IE independent of the vexed CPUE and total-catch data, though still reliant on adult age- and length-composition data, which are not contentious.

The close-kin data (i.e. the outcome of each juvenile-adult POP comparison) are incorporated into the OM directly as mark-recapture data, with a corresponding mark-recapture component in the likelihood (Hillary et al 2012, 2013). Two substantive adjustments to the OM were required to make it structurally compatible with the CKMR data: first to deal with the lack of sex- and length-substructure in the OM (since CKMR fundamentally requires that both be accounted for somehow); and, second, to change the form of the maturity ogive from knife-edge to logistic, consistent with the results on fecundity-at-size from the stand-alone CKMR analysis (Bravington et al., 2012;Hillary et al., 2012; CCSBT, 2013b).

The CKMR data were very informative when incorporated into the OM. This in part reflected the new absolute abundance information on the spawning component of the population they provided, where there previously was no information; however, it was also because some of the adult cohorts in the close-kin data were also observed as juveniles in the 1990s tagging data. Since both data sets contain information on abundance and mortality, the combination of the two data sets constrain the plausible fits and parameter space considerably. This resulted in the exclusion of the more pessimistic OM scenarios and a revision to the OM "grid" (Hillary et al., 2013; CCSBT, 2013b; CCSBT ESC, 2013).

3.3 SBT: Beyond Parent-Offspring-Pairs and microsatellites

The potential of CKMR for directly estimating absolute abundance and other key demographic parameters of natural resource management, has led to substantial investments in the theory and practice subsequent to the first tranche of SBT work. This has included:

1. Development of demographic CKMR models that can use Half-Sibling Pairs (HSPs: where two animals have one shared parent) as well as POPs.
2. Reviewing and testing the suitability and cost-effectiveness of different Next Generation Sequencing platforms (e.g. DArT, RadSeq, Sequenom, GBS) for large-scale close-kin genotyping to find HSPs and POPs (Bravington et al., 2015).
3. Development of general statistical/demographical theory for CKMR (Bravington et al., 2016)
4. Design and implementation of CKMR studies for other species (especially sharks) with very different sampling and demography (e.g. where only juveniles can be sampled).
5. Design work for CKMR as a long-term monitoring tool for SBT, using HSPs as well as POPs (Bravington and Davies, 2013; Bravington, 2014; Bravington et al., 2015). The long-term use is now endorsed and funded by CCSBT (CCSBT ESC, 2015; CCSBT, 2015).

3.4 Sharks

Many shark and ray species have been problematic for conservation and for commercial management, because of low productivity and because data of the "traditional fisheries" variety can be particularly dubious for by-catch and/or discard species. CKMR is particularly attractive for sharks because there is no need to rely on dubious catch-rate (or even catch) data, and for some species because CKMR can be combined with live biopsies as well as with samples from dead animals, to generate much more information than from conventional mark-recapture alone. Sharks have quite different reproductive biology to teleosts (much lower litter sizes, and often little lifetime change in fecundity after maturity). For some species, this makes it feasible just to work with HSPs among juveniles, rather than with POPs (see 4.4.2); this is useful when, as is often the case, juveniles are easy to sample but adults cannot be sampled in useful numbers (with teleosts, though, it is almost always necessary to have some adult samples for POP comparisons too, to disentangle the effects of increasing fecundity in adult life). The other benefit of CKMR for sharks has been in unambiguously revealing stock structure, or its absence (e.g. section 3.4.1). Since 2012, a number of CKMR shark projects have been started in order to provide baseline management information, especially on abundance, for which there are no credible alternatives.

3.4.1 Northern Australian River Sharks

Several euryhaline elasmobranchs in Northern Australia spend their juvenile years within a river system, before moving to (and between) estuaries and the open sea as adults, returning to rivers to breed: Freshwater/largetooth sawfish (*Pristis pristis*), Speartooth shark (*Glyphis glyphis*), Northern river sharks (*Glyphis garricki*). Historically, all have been subject to some degree of fishing, in some cases leading to substantial declines, but no credible quantitative estimates are available, and adults are now difficult or impossible to catch. Sampling juveniles by live biopsy for CKMR began in 2012 (with

fairly small samples, i.e. 100s of animals), and preliminary results are now available for one species. Aside from adult abundance estimates, the data clearly show breeding fidelity of females to particular groups of rivers (because the two members of each maternally-linked HSP are almost always found in the same river).

3.4.2 School Shark

School shark (*Galeorhinus galeus*) is a long-lived slow-breeding species which used to be the mainstay of commercial shark fisheries in southern Australia, before over-exploitation led to its collapse and listing under conservation legislation. The formal recovery plan, which has been in place for some years, should in theory have allowed some level of recovery by now. However, the reductions in TACs associated with the recovery plan has meant the conventional catch-rate monitoring used as an index of abundance in the assessment is no longer available/reliable; hence it is not a useful way to tell whether a recovery has really occurred. Gear selectivity means that adults are hard to catch, but juveniles are caught in some numbers; a large-scale CKMR project (1000s of juvenile samples from catches in the fishery) began in 2015, and is expected to deliver preliminary adult abundance estimates in 2017.

3.4.3 White shark

The eastern Australian / New Zealand population of white sharks (*Carcharodon carcharias*) was exposed to appreciable human-induced mortality in the middle-to-late last century. There is some public perception that numbers may have increased recently, but no suitable data is available for constructing an abundance estimate, nor for reliably monitoring trends in the future. Adult white sharks are too rarely encountered to be useful in any abundance-estimation method, but juveniles can be reliably sampled along the eastern Australian coast. A CKMR study began in 2011, using tissue samples from living and dead juveniles to identify half- (and full-) siblings, with age estimated from body length or vertebral ring counts. So far, sufficient sib-pairs have been found from samples around Australia and New Zealand to provide clear information on stock structure, and to fit a CKMR model and provide a preliminary estimate adult abundance and survival.

3.5 Other tuna species

3.5.1 Pacific Bluefin Tuna

The highly depleted status of Pacific Bluefin Tuna (PBF), and the uncertainty in the current assessment, prompted consideration of CKMR's suitability for PBF. The situation is more complicated than for SBT. The main spawning grounds of PBF are covered by multiple fisheries with different selectivity patterns, and juveniles spawned in the different spawning areas *may* head to different juvenile destinations for the first few years of life. A review and design workshop was held in La Jolla in May 2015 (NOAA et al., 2015). The workshop resulted in a proposal for a multi-national project, using POP and HSP CKMR, to the International Science Committee (ISC) in July 2015, which was subsequently supported by the Northern Committee at their 2015 meeting (WCPFC, 2015). Members of the ISC are initiating sample collection, and it was reported at the 2016 Monterey "Bluefin Futures" meeting that some progress has been made on genetic planning and demographic modelling.

3.5.2 Western Atlantic Bluefin Tuna

From a CKMR perspective, the biggest difference between SBT and ABFT is the strong large-scale population structure (West vs East Atlantic) in the latter. This would not matter if the two sides never mixed (because each could be sampled and modeled separately); however many EBFT fisheries catch both WBFT (i.e. spawned in or near Gulf of Mexico) and EBFT (i.e. spawned in the Mediterranean), so mixing must be explicitly considered. As explained in section 5, a naive misuse of CKMR that relies on uneven sampling[3] but fails to account for population structure will result in biased abundance estimates. Fortunately, there is good reason to believe that genetic markers can be found that discriminate with reasonable accuracy between WBFT and EBFT (e.g. Arrizabalaga et al., 2014; Fraile et al., 2014; Rooker et al., 2014). As long as the assignment accuracy of the markers is quantifiable and reasonably high (it does not have to be 99%), then tissue samples can, in effect, be assigned post hoc to W/E populations, so that CKMR can in effect be applied separately to WBFT and EBFT. For design purposes, we have assumed this will apply for both WBFT and EBFT.

[3] IE when per capita sampling probability varies between stocks.

NOAA's South-East Fisheries Centre hosted a workshop in February 2014 to explore the potential of CKMR for the western population of WBFT. There is limited complexity of population structure within Western EBFT (though see recent results from Richardson et al., 2016), and it was concluded that CKMR along similar lines to SBT (i.e. POP-only) should be feasible; based on the current ICCAT stock assessment and preliminary design calculations, sample sizes could be lower than for SBT. The workshop identified that the main complications of WBFT compared to SBT for CKMR are: the need to discriminate WBFT from EBFT fish; that the current fishery for adult WBFT is not a spawning-ground fishery; and the logistics of obtaining adequate juvenile samples of WBFT. While none are considered insurmountable at this stage, a number of questions need to be addressed before a full design study would be warranted, including: the extent of "sibship" in collections of larvae from surveys from the Gulf of Mexico; the suitability of historical samples, especially larvae, for genotyping for CKMR; and the availability of a genetic or chemical marker to discriminate between WBFT and EBFT with sufficient accuracy. A NOAA-CSIRO-VIMS project is underway to address these questions and will be complete in the second half of 2016; contact John Walter at NOAA or Campbell Davies at CSIRO for further details.

3.5.3 Eastern Atlantic Bluefin Tuna: what is different?

Eastern BFT is more complicated than any other CKMR case we have considered to date, because of the potential for stock/population/subpopulation[4] structure *within* the eastern (i.e. Mediterranean-spawned) population, without any guarantee that genetic markers exist at all[5]. As demonstrated in section 5, failure to allow for such structure, i.e. naively mis-applying CKMR to haphazardly-sampled EBFT, could easily lead to biased results. Fortunately, provided that adequate sampling can be arranged (see section 5), we show that CKMR can be used to tell whether structure really does exist within EBFT, regardless of whether markers for within-Mediterranean structure exist. Aside from the general CKMR issues such as "how many to sample", in this report we focus on these key issues for EBFT :

- possible structure (spawning-ground fidelity) within EBFT;
- no guarantee of "balanced" sampling across EBFT spawning grounds;
- multiple juvenile aggregations of predominantly eastern fish, none of which necessarily receive proportional contributions from different spawning populations.

3.5.4 Feasibility and benefits of combined EBFT-WBFT Close-kin Mark Recapture study

While the concept of a combined E-W EBFT CKMR study is intuitively appealing, it is not necessary, and is likely to be more complicated to implement and manage than separate individual studies in the first instance. The potential benefits of a combined study (see below) are unlikely to be large enough to out-weight the additional logistical and institutional difficulties that would be associated with development and implementation of a trans-Atlantic project.

At a practical level, the different nature of the fisheries, the migratory paths and the sampling opportunities between the east and west stocks means that there are unlikely to be substantial efficiencies in sample collection and logistics. Also, the more advanced stage of scoping and planning for the western stock, relative to the substantial uncertainties that remain to be addressed to design and implement an eastern project, would mean a number of years delay for the western stock. Given the immediate issues for the western stock, it would not seem appropriate to delay progress there for the sake of a combined program, in the short-term.

Notwithstanding this, CKMR is very likely to provide additional benefits and cost-efficiencies when considered as a long-term monitoring tool (Bravington and Davies 2013, Bravington 2015). So while we do not consider it desirable to seek to implement a combined program in the short-term, should individual close-kin mark recapture programs proceed (and prove successful over the coming 3-5 years and the issues to each application have been resolved), then it would make good sense to revisit the proposition of combined program with a view to long-term monitoring of stock rebuilding plans and an input data series for stock assessments and/or multi-stock management procedures.

There is one area in which there would be immediate benefits of continuing close collaboration, that of genetic markers east-west stock identification. As noted above, there is the need for stock identification markers for close-kin mark recapture. There has been a substantial investment in the development of genetic markers for stock identification through a range of programs, including the GBYP. There is still some refinement required to settle on a final set of genetic markers

[4]There are no universally agreed operational definitions for these terms. We have tried to avoid using any of them, just saying "structure" where possible, but in other places we have used them interchangeably.

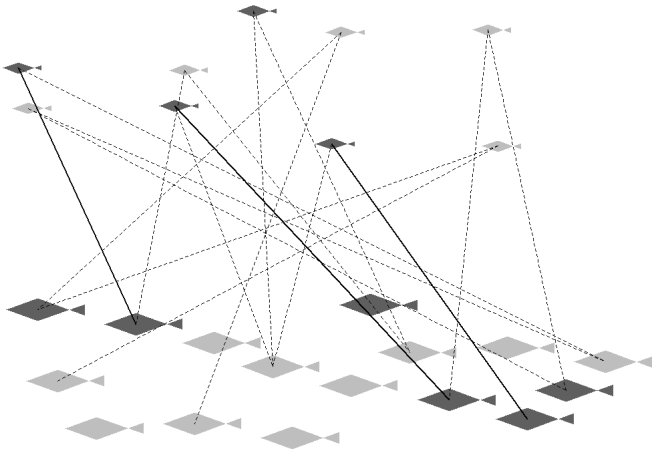
[5]The relevant aspect of "structure" here is fidelity to particular spawning ground, which does not have to be heritable to affect CKMR. Thus no marker might ever exist; and even if fidelity is heritable, low rates of "migration" can erase any genetic signal.

and these will then need to be validated using sufficiently by large numbers of known source samples. We consider this a priority for continued collaboration, which could effectively be addressed as part of the international workshop on GBS methods for CKMR and stock discrimination (See recommendations section).

4 General considerations for design of Close-kin Mark Recapture

To provide a conceptual starting point, the simplest version of CKMR is shown in Figure 1. Each juvenile is an offspring which "marks" its two adult parents, via its DNA. In the cartoon, there are $n_J = 4$ juvenile samples, and $n_A = 6$ adult samples. We compare the genotype of each of the n_J juvenile samples to each of the n_A adult samples, to check if a "mark" is recaptured. In each comparison, the probability that the adult happens to be one of the juvenile's two parents is $2/N_A$, where N_A is adult population size. Hence, if the entire set of $n_J \times n_A$ comparisons yields H Parent-Offspring Pairs, then adult abundance can be estimated as $\hat{N}_A = 2n_J n_A/H$. In this carefully-contrived example, $H = 3$ and $\hat{N}_A = 16$, which happens to be exactly right.

Figure 1: The simplest form of CKMR. Juveniles are small, adults are big; parents and offspring are linked by lines; dark means sampled, light means unsampled.



Real applications to open populations— even "simple" ones like SBT— are more complicated due *inter alia* to adult mortality in the interval between birth and sampling, non-random sampling, reproductive variability, different types of "mark" (i.e. different kinships), and uncertainty in genotyping, all of which can affect the probability of recapture. EBFT is yet more complicated because of the potential for structure.

In the rest of Section 4, which is fairly technical until section 4.4, we explain the basic principles of CKMR in realistic settings. Population structure and ABFT-specific considerations are largely deferred until Sections 5 and 6. Throughout, we treat the genetics as a "given", i.e. making the assumption that genotyping is done well enough to find kin-pairs accurately; our experience with SBT and other species has shown that this is entirely possible if— but only if— the right approach is chosen, and implemented with careful attention to quality control procedures to detect and minimise errors.

4.1 Beyond the cartoon: demographic probability of kinship

The CKMR cartoon, where "each fish tags its two parents", is great for conveying that light-bulb-moment of insight, but not actually much use for implementing CKMR. In reality, sampling will generally occur over more than one spawning season, the chance of being sampled will not be equal for all adults, and it will not be possible for all adults in each sample to be potential parents for all juveniles in each sample (e.g. by virtue of being already dead). Hence it is necessary to construct a more complicated statistical likelihood to account for factors that affect the probability of each pairwise comparison yielding a POP (Bravington et al 2014). For example, those adults that produce a greater number of fertile eggs (because they are bigger, say) end up with more "tags", and so are more likely to be found in POPs. Rather than the cartoon, it is more useful to think in terms of Expected Relative Reproductive Output (ERRO), and then how to calculate

it in terms of parameters and variables in a population dynamics model. For a complete explanation, see section 3 of Bravington et al. (2016).

Suppose we have a "juvenile" called j (note: j doesn't have to be young, it just has to be a potential offspring) and an "adult" (just a potential parent) called i. For simplicity, assume i is female, so we are specifically considering whether i is likely to be j's Mother. It is in one sense obvious that the probability *is* the ERRO:

$$\mathbb{P}\left[i \text{ was } j\text{'s Mother}\right] = \frac{\text{Expected number of } i\text{'s offspring that are } j\text{-like}}{\text{Expected } \textit{total} \text{ number of } j\text{-like animals}} \tag{1}$$

What does "j-like" mean? For a start, it has to be the right age. That is: if j is age 2 when sampled ("now"), then we are only interested in "j-like" animals that were born 2 years ago. Also, if there is a possible correlation between birthplace and sampling-place of juveniles (as there could be for EBFT between spawning-grounds and juvenile sampling-grounds) then j-like also means "born in the same place as j". So i must have been on the right spawning ground 2 years ago (we may not know which spawning ground that was, but for now that's a separate issue).

When it comes to i, then one has to express mathematically what her fecundity might have been 2 years ago, based on how big/old she is now, and whether she could have been in the right place. Implicitly, (1) should be conditioned on any covariates measured for i and j (e.g. time and place of sampling, age, and— at least for i— size). Unmeasured but biologically important covariates need to be integrated over, not ignored. In fact, (1) can be quite a deep and subtle equation, and the key to CKMR is to *carefully* think through what it should mean, given the biology, sampling process and the things that are measured.

In a cartoon-like example, where nobody dies and there are no time-lags and everyone is much the same except for being juvenile or adult, then all juvenile-adult comparisons are equivalent and the denominator of (1) is N times larger than the numerator, where N is the number of adults of i's sex; so the equation is just $1/N$. For teleost fish, though, things like adult age (an individual covariate) and fecundity (which depends on size and/or age, and involves parameters as well as covariates), and residency/selectivity on the spawning ground affect the probability.

The HSP case is similar. Here we are comparing two individuals, j to k, to see whether they have the same mother. Since we don't know who j's mother was, we have to sum over all possible females alive at the right place and time to give birth to j, for each one considering her subsequent ERRO of k-like animals as per (2.1). The exercise is then repeated to see if j and k have the same father (note that it is possible to distinguish, at least probabilistically, between maternal and paternal HSPs). See sections 3.2 and 5 of Bravington et al. (2016).

Because of the possibility of strong "litter-to-litter" variation in larval survival, HSPs in the same cohort should be avoided (i.e.: for demographic modelling, one should only use the results of *between*-cohort comparisons). It is also probably best to confine HSP tests to immature fish, since HSPs cannot in practice be genetically distinguished from grandparent-grandoffspring pairs, which become fairly common in long-term studies and could contaminate or complicate the demographic probability formulae. This cannot become an issue provided that only immature fish are checked for HSP status, and that sampling is "with removal" (i.e. lethal).

Aside from increasing the number of close-kin pairs and thus the general statistical power of a CKMR study, HSPs provide qualitatively different information which, at least for teleost fish, is important to ensure statistical estimability; in particular, to cleanly distinguish natural mortality from selectivity (section 4.4). In the POP-only application to SBT, we were able to bypass HSPs through the good fortune of having detailed reproductive biology and other ancillary data from the spawning ground; even then, we had to make an assumption that would ideally be avoided (and that can be if HSPs are also used).

The quantities required to calculate (1) and its HSP analogue, are typically adult numbers-at-age-and-year (by population, if necessary), and fecundity-at-size. (Of course, these are unknown, and instead are manifested algebraically through parameters which need to be estimated.) When faced with real data, it's important to do calculations that explicitly include adult body-size, not just age, otherwise the results can be biased. For design purposes, though, it is probably adequate to work as if age was always measured and all important phenomena were driven by age rather than length.

4.2 Parameter estimation and requirements for other data

Given a set of parameters for which to compute the CKMR log-likelihood, we first compute, as above, the prior probabilities of POP-hood and HSP-hood for each pairwise comparison (of which there may be millions, although they can be grouped into a much smaller number of categories), then calculate the Bernoulli (single-trial binomial) log-probability for the actual outcome, then add them all together.

The direct evidence about N_{adult} from CKMR is always back-dated to year of juvenile birth. Since several juvenile cohorts will have to be involved (see below), and a breakdown of adults by age is essential for constructing the probabilities, it's clear that for statistical identifiability CKMR requires a "mini-assessment model"[6] which tracks adult age composition for several years (as was done for SBT). In particular, it is necessary to have age-composition data from one or more fisheries that catch adults[7], and to incorporate a log-likelihood component for those data. This entails introducing parameters for selectivity and (adult) mortality rate. In other words, CKMR cannot work in complete absence of other "informative" data, at least not for teleost fish; the underlying issue is that the same total ERRO could come from lots of small adult animals, or a few big ones.

Importantly, however, CPUE-type data is certainly not needed, and in fact nor is total catch; the CKMR stand-alone assessment is naturally structured around total mortality rate z, which might be modeled as time- and age-dependent. However, if the total catch data is deemed trustworthy, then it could be useful statistically in refining the dependence of z_t on time t, for example. A reasonably accurate breakdown of the age composition of the total catch (as is available for most large-scale developed-world fisheries) would be needed to make this worthwhile; in our calculations for this project, we have assumed such a breakdown will be available, although at present data of sufficient accuracy might not be available. The method used for ICCAT assessments is to apply cohort slicing to length data, which is relatively inaccurate, particularly for all but the youngest most rapidly growing age classes.

4.3 Design

The construction of the log-likelihood can also be used for experimental design calculations. The design reveals how many comparisons of each "type" (i.e. with given values for the covariates) there will be, and given guesstimates for the parameters (e.g. from an existing assessment), one can then work out the expected Fisher information from a comparison of a given type. In addition, an approximation to the Fisher information from non-CKMR sources is required, in particular age- or length-compositions for adults from the spawning ground. Expected precision for any quantity of interest can be computed by standard statistical techniques, such as, the "delta method" and asymptotic variance formulae.

4.4 What is Close-kin Mark Recapture really telling you? A heuristic explanation

In thinking carefully about a CKMR design study, or the design of an estimation model, one should endeavour to develop some intuitive understanding of the relationship between the variables, the observations and the factors that affect them. The following is a simplified interpretation, pretending that only age (and not length) matters, and that things are in quasi-equilibrium (i.e. stable age composition, though not necessarily stable abundance). In practice, there is no way around actually building a model to jointly estimate the things considered below (time lags make it too complicated to consider separate "direct" estimation) but the principles should be clear. The following omits consideration of population structure, even though it is central to EBFT and this project. The reason is that structure (given adequate sampling) does not affect total numbers or patterns of kin-pairs expected, just the proportion of cross-over kin-pairs between populations; as such, inferences about structure are "orthogonal" to the things below. One of many attractive properties of CKMR is that the structure relationships will be evident in the final data themselves (section 6); e.g. in whether kin-pairs are always found close together, or are evenly spread.

4.4.1 POPs

- The sample age composition of adults is affected by selectivity-at-age and by total mortality (z), and so is the age composition of actual parents, but the latter is also weighted by fecundity-at-age. So, by "dividing" the two distributions, we can immediately estimate relative fecundity-at-age.
- With anything else that can be measured, adult selectivity-at-age and mortality are always entwined. This applies both to the "catch curve" (sample age composition) and also to average time-lags between offspring birth and parental "recapture". So there is no easy and model-free way to separate selectivity from adult total mortality just with POPs and age-composition data (although it may be technically possible if sufficiently strong assumptions are made about functional forms, as is common in conventional stock assessments).

[6] At least for teleost fish it does. For some sharks and marine mammals, where dynamics are slow and changing fecundity is not a big deal, things can be simplified considerably.

[7] Length-composition alone *might* be sufficient if there is enough information about length-at-age or vice versa, but it is by no means clear whether length-alone would work. We assume that some age data from spawning-ground fisheries will be available, as should be expected for any high-value, developed-world fishery.

- The total relative reproductive (RRO)— which affects how many POPs are found— depends not just on total adult numbers, N, but also on fecundity-at-age (estimable) and true population age composition (not directly estimable, because the sample age composition is affected by selectivity). So there is partial confounding between N and selectivity, unless there is some additional source of data or, as for SBT, a lucky sampling setup.

4.4.2 HSPs

- The adults are never seen, so selectivity on adults doesn't matter. If j is born in year y_j and its maternal half sibling (HS) k is born later in y_k, then we know that the mother survived the intervening $y_k - y_j$ years. The average birth-gap between the two members of a HSP is the inverse of the total adult mortality (which, if z differs by age/size, is in some sense weighted towards the main breeders). So, from the HSPs you can deduce z; then you can deduce what age-specific selectivity must be by looking at the sample age-composition of adults; then you can work out the population age composition, weight it by fecundity-at-age from the POP data, and get an unbiased, unconfounded estimate of N.

- It is also possible to use HSPs to estimate absolute adult abundance (the chance of j and k sharing a mother is inverse to the number of adult females), though this requires additional assumptions. Note that for teleost fish, it is not possible to use HSP-only CKMR to estimate abundance; POP information is still necessary to establish fecundity-at-age. In this study, we take the conservative option of assuming that HSP data will be used only to help estimate mortality etc., and to inform on trends in *relative* adult abundance.

5 Why a naive approach to Close-kin Mark Recapture won't work for EABT

The naive version is just to follow what was done for SBT: that is, comparing each juvenile sample from (say) the Bay of Biscay with each adult sample from the Mediterranean. Then the probability that the adult is the juvenile's parent, is naively assumed proportional to the adult's fecundity relative to the entire fecundity of all adults in the Mediterranean in the year the juvenile was born. Then one "simply" computes (1) and applies the machinery of section 4.3.

There are three points which, taken together, mean this approach is too simple (and will give wrong answers) for EBFT:

1. Juveniles in a sampling-area can originate from any of the spawning grounds, and there is no reason to believe that a larva from spawning ground A has equal probability of making it to the juvenile sampling-area as a larva from spawning ground B; there could be big differences in early-life-stage mortality as well as in dispersal and movement patterns. It can't be assumed there's any way to tell whether the juvenile comes from A or B.

2. Adults are caught on different spawning-grounds, and there is no reason to assume that total catch (or catch rate) by ground is proportional to true abundance by ground (i.e. there is a "spawning-ground catchability" which we cannot reliably predict).

3. Adults may tend to return to the same spawning-ground year after year.

To illustrate the problem: suppose there are just two spawning-grounds A and B, and for convenience ignore year and assume that fecundity (larval production) is equal for all adults. If a juvenile happens to come from A, then the probability that a female adult from A is its mother, is $1/N_{\text{♀A}}$ where $N_{\text{♀A}}$ is the number of female adults from the "A stock"; if the female adult is from the B-stock, the probability is zero. And if the juvenile is a B-stocker, then the probabilities are 0 and $1/N_{\text{♀B}}$ respectively. The ratio of juveniles that are A-stockers is, say, $\phi_{\text{A}} N_{\text{♀A}} :: \phi_{\text{B}} N_{\text{♀B}}$ where ϕ is the stock-origin-dependent probability of surviving until sampling *and* of moving to this particular juvenile sampling-ground. Dropping the ♀-symbol for brevity, the probability that any juvenile matches an A-stock female is a weighted average over the two possible stock-origins of the juvenile, i.e.

$$\begin{aligned} &\frac{1}{\phi_A N_A + \phi_B N_B}\left(\phi_A N_A \times \frac{1}{N_A} + \phi_B N_B \times 0\right) \\ &= \frac{\phi_A}{\phi_A N_A + \phi_B N_B} \end{aligned} \tag{2}$$

and for matching a B-stock female, it's

$$\frac{\phi_B}{\phi_A N_A + \phi_B N_B} \tag{3}$$

If we happen to have m_A A-stock females and m_B B-stock females, and we naively lumped all comparisons, then the overall probability per-comparison of a POP would be

$$\frac{m_A\phi_A + m_B\phi_B}{m_A + m_B} \times \frac{1}{\phi_A N_A + \phi_B N_B} \tag{4}$$

which is, as noted above, not the same as $1/(N_A + N_B)$. If we were lucky enough that $m \propto N$ (i.e. if point #2 didn't apply) or that $\phi_A = \phi_B$ (i.e. if point #1 didn't apply), then we'd be OK; but they aren't, so we're not.

A more sophisticated attempt would be to split the comparisons depending on whether they're with A-stock or B-stock females, and look separately at (2) and (3). Unfortunately, we have no idea what the relative ϕ's should be, so we are still left with 2 equations in 3 unknowns. There are a few notes to make:

- Only the ratio $\phi_A :: \phi_B$ matters, not the absolute value of the ϕ's; it is therefore OK to define $\phi_A \equiv 1$ and to consider only ϕ_B as an unknown (which is why there are only 3 unknowns).
- There is an important insight for POP CKMR: for it to provide an unbiased estimate of adult abundance, offspring-capture should be statistically independent of parent-capture, given the variables included in the model. If not, one is faced with what is known in the mark-recapture literature as Unmodelled-Heterogeneity-of-Capture-Probability, which is a serious source of bias. A key point in design is to ensure that sampling is adequate, and the estimation model is flexible enough, to avoid this problem.
- In point #3, persistence is crucial; there would be no problem if adults moved randomly from one spawning ground to another during their reproductive lifespans. This is because we deliberately avoid comparing a juvenile to an adult that was caught in the juvenile's birth-year anyway (see Bravington, 2014); and if the adult is caught in a subsequent year, then the chance of its being caught would be independent of whether its offspring has been caught.
- Point #3 does not require any *heritable* stock-structure— simply that each adult tends to stick to the same spawning ground after reaching maturity, regardless of which spawning ground it was originally born into. The problem still remains that juvenile probability-of-capture is correlated with adult probability-of-capture, via the unknown ϕ-ratio.

6 What would work for EBFT?

The possible complication of stock structure within EBFT entails a more intricate version of CKMR than for SBT, both in the sampling and in the modelling. The key to avoiding the problems described in section 5, is to collect samples over several years from:

1. one fishery on adults in every main spawning ground, and
2. known-age juveniles from at least as many juvenile "sites" (fisheries) as there are main spawning grounds— ideally from one more.

Then the general idea is to do site-specific comparisons of:

1. juveniles with adults (JA: POPs);
2. juveniles with juveniles (JJ: HSPs);
3. adults with adults (AA: POPs).[8]

[8] AA is feasible for EBFT only because they mature much younger than SBT; there is no point (yet) for SBT because AA POPs are too rare to be informative. It takes >10 years for an SBT to be an effective breeder, and in our fairly short (by SBT lifetime standards) studies to date, we do not have enough comparisons between old (20yo+) adults and young (<10yo) ones; though eventually this will change, as sampling intervals lengthen. For EBFT, the much shorter maturation implies that a much higher proportion of AA POPs will be found quickly.

By "site-specific", we mean that comparisons are tabulated by sampling site (i.e. fishery) as well as year, age, length, etc.; and that kinship probabilities are computed taking sampling site into account, under a specific demographic model described shortly. (Because the results will not be dependent on the number of sampling sites chosen, we assume, for this report, that there are 2 spawning grounds, and 2 distinct juvenile sampling sites even though more are known to exist.) As previously noted, we assume throughout that western (W) (Gulf of Mexico origin) fish can be genetically separated from eastern (E) (Mediterranean-origin) fish, subject to a known error rate. The W samples are then "discarded" for EBFT purposes, and the remaining discussion pertains only to E samples; for example, "spawning ground" always means "some place within the Mediterranean", never "Gulf of Mexico".

In the rest of this section, we try to explain how CKMR data collected as above can reveal the nature of any stock structure among EBFT, and then be used to estimate abundance and other demographic parameters. We have tried to keep this as non-mathematical as possible, but the concepts involved are very subtle, and considerably more complex than any other CKMR application we have considered, so some mathematics is unavoidable.

First, it is important to understand the conceptual difference between three stock-structure scenarios:

Heritable structure: where adult fish tend to return to the spawning-ground they originally came from.

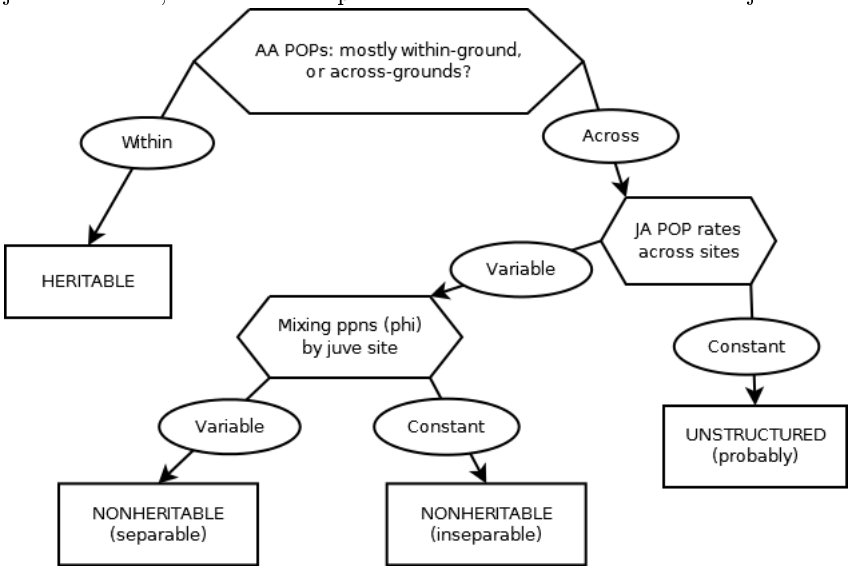
Nonheritable structure: where each newly-mature fish celebrates adulthood by selecting the spawning ground it will use for the rest of its life, *independently of where it was originally spawned*.

Unstructured: where each adult randomly chooses its spawning ground anew every year, independently of what it did in the past.

These three scenarios (H,N,U) will manifest themselves differently in CKMR data (provided the sampling is good enough), and will require different estimation models. They also clearly have different implications for assessment *sensu lato*, and maybe for management. Of course, the scenarios really exist on a continuum, and the demographic model that is ultimately used for a CKMR-based assessment may need to be a hybrid, allow for partial fidelity via "migration rates" between spawning grounds. Note that, even if H applies, there is no guarantee that any "classical" genetic marker exists, and it is guaranteed that no such marker exists if N applies. We therefore consider only such information as can be guaranteed from CKMR data alone.

The first task is to use the CKMR data qualitatively to decide which scenario (or hybrid) applies, so that an appropriate demographic model can be developed. This can be done by following the flowchart in Figure 2. The rationale for deciding "Heritable" is obvious: do adults tend to spawn in the same ground as their parents? If not, then the alternatives are "Nonheritable" and "Unstructured"; the rationale for deciding between these two is more subtle.

Figure 2: Stock structure decision tree. This must exclude POP comparisons where the adult is caught in the year of juvenile birth, and HSP comparisons within the same cohort of juveniles. See text for explanations of terms.



If EBFT are truly Unstructured, then comparison probabilities will be unaffected by where the samples are taken (after conditioning on covariates such as year and age). The reason is that, in an Unstructured population, a parent's spawning-ground at the time of its sampling is independent of the parent's location at its offspring's birth (guaranteed to be in a different year, as per the figure caption). We will have empirical estimates of each probability after doing the comparisons of each type, so this condition can be easily checked from the data.

It is unlikely that the comparison probabilities will be similar if there is Nonheritable structure. To explain why, algebra seems unavoidable. Suppose we sample juveniles from sites c and d (lower-case), and adults from the (only two, as assumed for this report) spawning-grounds E and F (upper-case). Then the probabilities of finding a POP[9], depending on place of sampling, can be calculated as per section 5. Again, in the interests of clarity, we show here a single-sex version omitting all covariates except capture-site:

$$\begin{aligned} p_{cE} &= \frac{1}{N_E + \phi_c N_F};\ p_{cF} = \frac{\phi_c}{N_E + \phi_c N_F} \\ p_{dE} &= \frac{1}{N_E + \phi_d N_F};\ p_{dF} = \frac{\phi_d}{N_E + \phi_d N_F} \end{aligned} \tag{5}$$

The interpretation of ϕ_c, say, is as a "mixing proportion", describing the relative per capita contribution of adults from spawning ground F to the juvenile pool in c, compared to the per capita contribution of adults from spawning ground E.

These equations are valid generally regardless of what type of stock structure applies, if any. The point under discussion is what they will reveal empirically and qualitatively about stock structure, and which parameters will be statistically estimable. If EBFT are Unstructured, for example, then $\phi_c = \phi_d = 1$ and this will be evident because the empirical $\hat{p}$'s will all be similar. If not, we have 4 equations in 4 unknowns (two N's and two ϕ's), and can in principle solve to estimate $\hat{N}_E$ and $\hat{N}_F$ by "method of moments" (e.g. first estimate $\hat{\phi}_c = \hat{p}_{cF}/\hat{p}_{cE}$ and similarly for $\hat{\phi}_d$, then invert both LHS equations to get linear simultaneous equations in $\left(\hat{N}_E, \hat{N}_F\right)$). This will only go wrong if $\hat{\phi}_c = \hat{\phi}_d$, i.e. if the mixing-proportions are the same on the two juvenile sites. That cannot occur if the juveniles are young enough and the sites far enough apart; in the extreme case, larvae from the Balearics have to come mostly from the Balearic spawning ground. By the time juveniles reach a more promising age for sampling, we might reasonably hope that 2yo juveniles in the Bay of Biscay will have a higher proportion of W Med-spawnees than 2yo juveniles in the Levantine Sea will have (i.e. different ϕ). However, this issue of "sufficient natural contrast" in ϕ can only be checked by actually doing a CKMR study.

If it turns out that there is not much natural contrast in ϕ across juvenile sampling sites, i.e. if $\hat{\phi}_c \approx \hat{\phi}_d$, then we can still *detect* Nonheritable (as opposed to U) structure, based on differences between $\hat{p}_{.E}$ and $\hat{p}_{.F}$. However, we could not then separately estimate N_E from N_F, nor form an unbiased estimate of the total $N_E + N_F$. (This is the reason for the "separable" and "inseparable" boxes in Figure 2.) Presumably, ongoing CKMR would still provide a reliable *relative* rather than absolute abundance time series (much better than CPUE), along with all the other CKMR benefits: direct estimates of fecundity-at-age, mortality rates, etc. In the context of a full assessment, the loss of absolute abundance might not be critical, since catch data (if reliable) can fill the gap just as it does with CPUE-based assessments. Nevertheless, absolute abundance is an important part of CKMR's appeal, so it is worth making the effort to obtain widely-spaced-enough and young-enough juvenile samples to have good natural contrast in ϕ.

6.1 Further observations

As we forewarned, the issues around stock structure in CKMR are complex. For any readers still keen for further details, we proffer a few more observations based on the biology and mathematics behind (5). Other readers may wish to skip to section 6.2 on a first reading.

- Mistaking Nonheritable for Unstructured is theoretically possible, but it would require very "unlucky" biology (or bad sampling), whereby larvae from different spawning grounds experience similar cumulative mortality prior to being sampled as juveniles, leading to $\phi \approx 1$. (As an example of bad sampling: if no juveniles at all were sampled, we would be dependent entirely on AA comparisons. By definition of Nonheritable structure, $\phi = 1$ among adults, so we would be unable to distinguish N from U.)

[9] Recall that comparisons are only "valid" between a juvenile and an adult that are caught *after* (not during) that juvenile's birth-season. Obviously, samples of early-0-group fish near spawning ground X would only yield POPs *in that year* against adults spawning in X, regardless of whether we have U- or N-structure. The interesting question is whether they would yield POPs when compared against adults caught spawning in *other* years away from X.

- There may also be partial stock structure (H or N), whereby each adult has an individual tendency to use a particular spawning ground, but in some years "decides" to go somewhere else. This would shrink the estimated $\hat{\phi}$'s towards 1, so that the corresponding abundance estimates by stock would not reflect the real numbers found on that spawning ground each year. In such cases it is not obvious what the "right" stock-level answer actually would be, since it is not clear exactly what "stock" should mean. However, the good news is that (5) is still valid; the difficulties relate only to interpretation of the *split* of N across spawning stocks, not to the total adult abundance estimate.

- Somewhat similar considerations apply to JJ comparisons for HSPs as to AJ comparisons for POPs. For example, if there is Heritable or Nonheritable structure, combined with variation in mixing-proportions ϕ across juvenile sites, then there will tend to be more within-site HSPs than across-site HSPs. However, the HSPs cannot be linked back to particular spawning grounds, so the information content is more limited than for AJ comparisons.

Parameter estimation would of course be done not via 5), but inside a proper statistical model constructed around a site-specific version (1) and as described in section 4. Exactly how that model is constructed, would depend on the qualitative conclusions about stock structure gleaned from the POPs. However, the basic structure would be fairly similar regardless of whether H, N, or U structure (or some hybrid allowing quantified "migration rates") is being described; parameters such as fecundity-at-age will be estimated alongside time-series of numbers-at-(adult)-age. Some non-abundance parameters, such as mortality rate, might be allowed to vary by "stock"; these are detailed modelling decisions that cannot be made until some close kin data are collected. Section 8 describes one specific version that we have implemented for design purposes.

Heritable stocks: the AA comparisons lead directly to unbiased stock-specific adult abundance estimates. As always with CKMR, though, the estimates are back-dated to average birth-year of the potential offspring. With AA comparisons, the "potential offspring" are actually the younger adults sampled, so that the average back-dating may be substantial. The AJ comparisons, as in (5), will provide more up-to-date estimates, since they are only back-dated by juvenile age. We would also expect that AJ comparisons will yield more POPs. HSPs from JJ comparisons are used to separate the effects of mortality and fecundity.

Nonheritable stocks: the AA comparisons lead to a combined (and back-dated) adult abundance estimate. Provided that there is enough natural contrast in mixing-proportions (ϕ) between juvenile stocks, then AJ comparisons will give up-to-date spawning-ground-specific estimates of adult abundance. The spawning-ground-specific interpretation becomes somewhat fraught if spawning-ground-fidelity is partial rather than complete, but the estimate of aggregate adult abundance remains unbiased. Use of HSPs is basically as for H.

Unstructured stock: all abundance estimates are combined across adult spawning grounds (regardless of whether all grounds are sampled or not). The statistical model is simpler because there are fewer parameters to estimate. There is no way to estimate separate abundances by spawning ground from CKMR, but by the same token it probably doesn't matter for management; with an Unstructured stock, all EBFT are eventually exposed to the same mortality rate because they move around so much. Use of HSPs is as for N and U.

6.1.1 Cryptic stocks and diagnostics

The ability of (5) to produce separate abundance estimates by stock relies on there being enough natural contrast in the ϕ's amongst juvenile sampling sites. It is interesting to note that, if there are *more* juvenile sites sampled than (putative) adult spawning grounds, and provided that the natural contrast exists, then it becomes possible to detect— and even, to some extent, estimate— an additional "cryptic" spawning stock that has not actually been sampled. This becomes mathematically apparent in an extended version of (5) where there are more rows (sampled juvenile sites) than columns (spawning grounds), but where one of the columns is missing (not sampled).

A further internal check on the consistency of the whole CKMR-based assessment, including but not limited to stock structure, comes from cross-checking the somewhat-retrospective abundance estimates from the AA comparisons against the more up-to-date estimates from the AJ comparisons. This is done implicitly in the estimation model used for section 8, but as part of estimation under the "right" model, not as a diagnostic. The reason for mentioning it here is simply to point out that some diagnostics are available under the sampling program we propose, so that whichever demographic assumptions end up being adopted for a CKMR-based assessment need not go untested for all eternity.

6.2 Summary

This section has concentrated on stock structure and abundance in an expanded version of the "cartoon" (Figure 1). It should be evident that (i) stock structure makes things complicated, but (ii) a well-designed CKMR sampling programme for EBFT will yield a great deal of qualitative information about any stock structure, and (iii) that there should also be enough data to yield unbiased abundance estimates, using a model that will depend on what stock structure scenario turns out to apply; in some scenarios but not all, estimates can be made down to the level of individual spawning stocks.

Given that no-one knows what the real stock structure of EBFT is yet, there is little point in testing a vast and elaborate range of scenarios. For the numerical results in section 8, we opted (for this report) to consider one fairly challenging scenario (Heritable separate stocks on two spawning grounds, with two juvenile sampling sites), to give an indication of what sample sizes might be needed. The vagaries of stock structure, and indeed the substantial vagaries of the entire EBFT assessment which inevitably underpins any design process, mean that the numbers are sure to be refined in future as data accumulates.

However, the basic necessity remains of genotype-sampling all major known adult spawning grounds, and at least as many juvenile sites— though it is *not* necessary to sample *all* EBFT fisheries (e.g. adults outside the spawning season are not useful). The other sampling requirements (length, age, sex, thoroughness of genotyping) remain as per section 4; this is essential so that other demographic parameters such as fecundity-at-size, selectivity, and natural mortality can be estimated, and to avoid abundance being statistically confounded with these other parameters.

A dot-point version of the requirements is as follows:

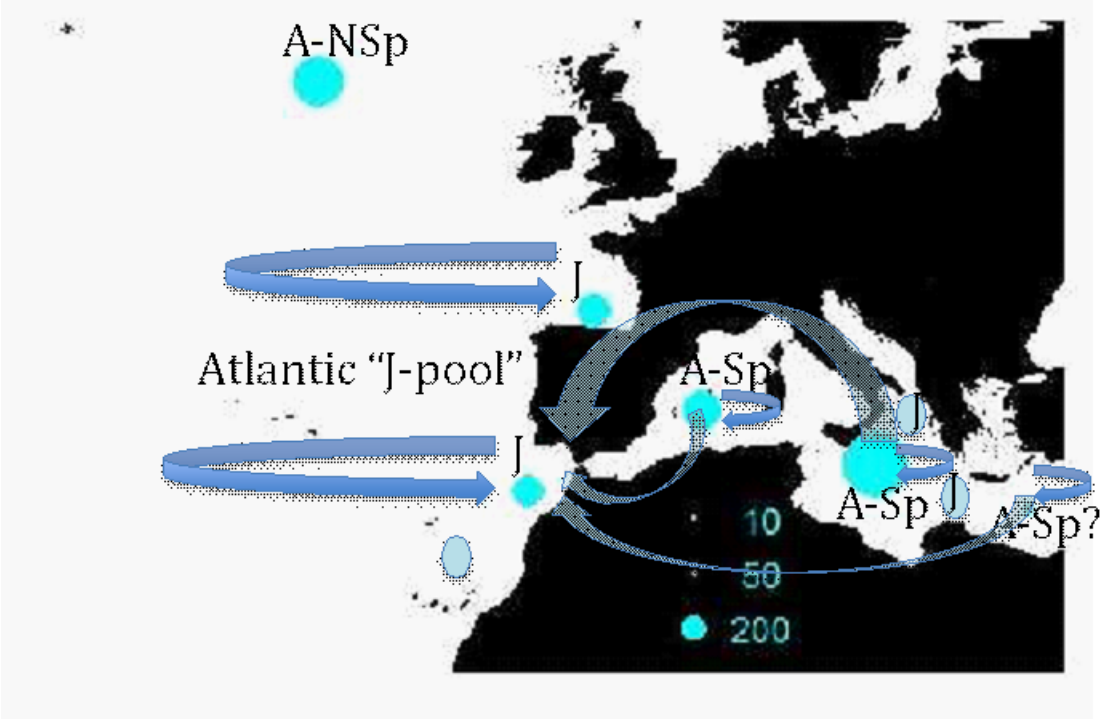
- DO need samples of adults from one fishery in each main spawning-ground. If not, it is impossible to resolve stock structure. Abundance estimates may then be substantially biased; the bias will be undetectable.
  - It may be hard to know whether all the "main" spawning-grounds have been covered. Some degree of safeguard is provided by the diagnostics in section 6.1.1, provided juvenile sampling is adequate.
- DO need samples of young juveniles from at least as many well-separated juvenile sites as there are main spawning-grounds. If not, stock structure cannot be fully resolved, and only relative abundances may be possible.
  - It is important to have good natural contrast in mixing proportions (see text) between juvenile sites. This can't be checked in advance, but is more likely for younger juveniles.
  - It is essential that juvenile age can be inferred fairly reliably[10]. Length alone may well be adequate for this, provided that the juveniles are sampled young enough.
- DO need to genotype to HSP-finding quality (a fairly stringent requirement)— without HSPs, the ability to separate age-specific fecundity from overall mortality rate within CKMR is much weaker, and the complexities of EBFT stocks/fisheries/sampling do not permit the shortcut used previously for SBT.
- DO need other data from the sampled fisheries, as per section 4. Catch data from *unsampled* fisheries is also useful to establish total removals, but is not an absolute requirement— though breakdowns of total catch by fishery, by length, and if possible by age, are the cornerstone of any reliable stock assessment, CKMR or otherwise.
- DO NOT need samples from other fisheries, such as adults out of the spawning season where the "true stocks" (if any) would be mixed. The information for CKMR would be limited, because of extra layers of mixing parameters. To the best of our understanding, CKMR for EBFT will work if and only if samples of the right type can be collected; samples of the wrong type would be a distraction and not an adequate substitute.

7 Initial review of existing collections and tissue samples for EABT sample sizes and distribution

Figure 3 depicts a proposed population structure, connectivity and sampling concepts required to be considered in the design of a CKMR study for EBFT. It is not intended to reflect the "true" or only potential spawning structure for Mediterranean EBFT. An initial review of the samples available for known spawning adults and juveniles of EBFT, and discussions with experts involved in their collection indicates:

[10] Because it determines year-of-birth, which is used to decide which comparisons are "legitimate". While adult age is also important, accuracy is not so important.

Figure 3: Schematic of adult spawning areas and juvenile grounds to illustrate population structure, connectivity and life-history concepts of the kind that need to be considered for design of CKMR for eastern Atlantic bluefin tuna. Other scenarios have been put forth by other authors. This scenario depicts three self-recruiting spawning populations within the Mediterranean Sea. Juveniles migrate to other locations in the Mediterranean and the Atlantic, returning to their natal spawning ground once mature (assumed to be from 3 years on). Size of circles in the juvenile (J) and adult (A) areas reflect numbers of tissue samples collected in 2013 as part of GBYP biological sampling program. A-Sp notes adult spawning grounds; A-NSp adult winter feeding ground. Adapted from Arrizabalaga et al 2014. Similar sample sizes were collected in 2013 (Table 1).



- The current GBYP sampling strata for the GBYP program are likely to be suitable even if the truth is as complicated as Figure 3. The essential elements are the timing of sampling occasions (through-out the spawning season for adult fish) and the quantity and quality of the samples and associated data obtained (see below).
- The sample sizes available in the GBYP collections to date are insufficient for CKMR (see for example Table1 and Table 4); they would need to increase substantially and be continued for several years. This is also the case for the more recent 2015 sampling and current strata for the GBYP (Appendix 2).
- Initial discussions indicate it should be possible to collect the larger sample sizes required for CKMR in the future—both for adults on the spawning grounds in the Med, and for juveniles in the Med and the Atlantic.
- Unbiased abundance estimates require adult samples from all main spawning grounds (unless it turns out that there really is no within-Med structure; but we will not find out without the samples anyway). For the purposes of the total sample size calculations we have assumed there are two "known" (Western and Central Med) and a third "unconfirmed" (eastern Med) spawning ground. If adequate sample sizes can be obtained from all main spawning areas (i.e. Balearic Sea, Tyrrhenian Sea, southern-central Mediterranean Sea and Levantine Sea), and from at least five juvenile grounds (i.e. one more than the number of spawning areas), then it should be possible to diagnose from the CKMR data alone whether all the important spawning grounds really have been encompassed. For simplicity, for this initial scoping exercise we present numerical results (sample size requirements etc.) under a simpler scenario, with just two spawning grounds and two juvenile sampling sites, to demonstrate that the required parameters are estimable, in principle. These can be refined to reflect more specific scenarios, informed by more detailed data and information, should the project proceed to the detailed feasibility study.

- It is essential that samples from spawning adults can be assigned with high reliability to the correct spawning ground, and associated with the correct measurement of length (and, ideally, an otolith). Note, by "assignment" we are referring to direct assignment from observer records, catch documentation, harvesting records etc; not chemical or genetic markers. For EBFT, this may not be easy for some of the purse-seine fisheries, where long tows to farm operations are involved. An important issue for any proposed follow-up study, is to consider how serious this issue might be.

- Age data (otoliths) are required for the samples from each spawning-ground fishery, or at least for a large enough proportion of samples to provide a representative age-length key with low CV. As many as possible should be collected, even if not all are read; it is particularly useful to be able to ascertain age for identified parents, by going back to the otolith collection once a POP is identified. Length compositions from those sampled fisheries are also needed. Further consideration of subsampling (i.e. how many fish per length class to genotype) would be needed in a full implementation. Note that sampling and archiving tissue is expected to be cheaper than genotyping, so it makes sense to collect a lot of samples now, and decide later which ones (and how many) to genotype. Again, these points should receive attention in any follow-up study. Importantly, it is not necessary to obtain samples from all fisheries— only from fisheries on adults on the spawning grounds in the spawning season, and from enough fisheries on young juveniles (e.g. YOY or 1-2yo) to deal with mixing (section 6).

- While it is not necessary to sample juveniles from all fisheries, it is necessary to sample from at least as many juvenile regions as there are sampled adult spawning grounds, and ideally from more; the extra juvenile grounds give a diagnostic on whether all important adult spawning grounds have been sampled.

- Otoliths are not required for juveniles, provided that a precise-enough age-length key is available to allow selection of known-age (YOY, 1yo or 2y) juveniles for subsequent genotyping.

  Table 2 identifies the gaps in the currently GBYP biological sampling program, given the requirements for CKMR outlined above.

  On our current understanding of the most "complex" population structure considered plausible by ICCAT, sampling should involve three spawning grounds/populations in the Mediterranean and 4 juvenile feeding grounds spread between the Mediterranean and eastern Atlantic Ocean. We also consider 1yo or 2yo fish to be the best options for juvenile sampling; 1yos have the advantage of being less-mixed (i.e. better natural contrast among the ϕ, as per section 6;), but also more risk of high within-cohort sibship[11]. It is not possible to say which age would be better without trying both. Larvae are likely to be more troublesome, because of tissue quality and contamination concerns, because of increased within-cohort sibship risk, because they require dedicated surveys rather than fishery by-products, and because (paradoxically) they are so "pure" with respect to spawning ground that they carry no information about other spawning grounds.

[11] Within-cohort sibs are not useful for CKMR, and if their incidence is high, the effective sample size is reduced; it is only between-cohort sibs which are useful.

Table 1: Number of EBFT sampled by area/fishery and size class, in 2013, as an indication of the sampling intensity relative to that required for CKMR (compare total numbers of juveniles and adults with those in Table 4). Empty cells indicate that no sampling was planned or accomplished in that stratum. Green cells indicate strata where sampling took place despite not being planned. Taken from Arrizabalaga et al 2014; caption has been paraphrased. Review of more recent sampling strata and total samples of juveniles and adults numbers (GBYP 2015, Appendix 2) also indicate that substantially larger numbers of samples will be required to meet the preliminary estimates of sample size required for CKMR.

		Larvae	Age 0	Juveniles	Medium	Large	Total		
			<=3 kg	>3 & <=25 kg	>25 & <=100 kg	>100 kg		Target	%
Eastern Mediterranean	Levantine Sea		86				86	50	172%
Central Mediterranean	Adriatic Sea			60			60	60	100%
	Malta		50				50	50	100%
	South of Sicily and Ionian Sea	58	50	50			158	100	158%
	Gulf of Syrte				114	140	254	150	169%
Western Mediterranean	Balearic	42	134				176	110	160%
	Ligurian Sea		33				33	50	66%
	Sardinia				36	5	41	60	68%
	Algeria					0	0	50	0%
	Tyrrhenian		50		3	2	55	50	110%
Northeast Atlantic	Bay of Biscay			265	26	1	292	110	265%
	Canary Islands					89	89	50	178%
	Gibraltar			2	59		61	60	102%
	Morocco					110	110	110	100%
	Mauritania					23	23	0	>100%
	Norway					1	1	0	>100%
Central North Atlantic	Azores-Madeira					0	0	50	0%
	Central North Atlantic				34	380	414	50	828%
Northwest Atlantic	Gulf of Mexico	84				0	84	50	168%
	Gulf of Saint Lawrance					23	23	0	>100%
	Newfoundland-Labrador					9	9	0	>100%
	Nova Scotia					17	17	0	>100%
	Total	184	403	377	272	800	2036	1210	168%

Table 2: Summary of sampling options for CKMR relative to current GBYP Biological sampling strata. √ indicates substantial sampling within current GBYP program, albeit not at the annual sample sizes required for CKMR.

Region	Area/Fishery	0+	1-2 yo	Spawning Adults
Eastern Mediterranean	Levantine Sea	√		√
Central Mediterranean	Southern-Central Mediterranean Sea	√	√	
	Adriatic Sea		√	
	Gulf of Syrte			√
Western Mediterranean	Balearic Sea	√		
	Sardinia			√
	Ligurian Sea	√		
	Gibraltar			√
	Tyrrhenian	√		
North-east Atlantic	Bay of Biscay		√	
Central North Atlantic	Central North Atlantic			√

7.1 Genotyping to identify kin

If this project proceeds beyond scoping into a full design and implementation, a number of technical issues will need to be considered in depth. These include sampling protocols, lab handling and management of tissue samples, genotyping method and quality control, and management of genotype data. We don't consider these in any detail here, but rather note a number of points which should be considered if the project proceeds.

The Genotyping by Sequencing (GBS) methods used to identify more distant kin than POPs (e.g. DArT, which we are now using for SBT) require high-quality tissue samples to provide reliable genotypes. Although we prefer DArT to microsatellites, we have noticed that it is more sensitive to contamination (and this is likely true for any GBS approach). The potential for cross-contamination of samples is real, and it is essential to pay attention to quality control procedures in the collection of samples and diagnostic analyses for detecting incidences of contamination. Similarly, there is the need for investment in relational database systems to effectively manage and curate the very large data volumes generated.

Our brief review of the genetic components of the GBYP Biological Program to date indicate that considerable thought has been given to these issues and it has been possible to collect high quality samples (tissue and fin clips). We understand, however, that here have been some issues with contamination and/or incorrect identification of larvae, which are being addressed as part of current and future work. An additional consideration, which would form part of the proposed second stage of this project, is the genotyping approach used to identify kin. As recently demonstrated ((Bravington, 2014; Bravington et al., 2015), with the right type of GBS method it is possible to identify HSPs, as well as POPs, which significantly increases the information content of CKMR. This additional information content means it would not be sensible to contemplate a new CKMR study without HSPs. The "focused" DArT approach tested by Bravington et al. (2015) delivers very high read-depths and genotyping accuracy— a necessity for reliably identifying HSPs— at very low unit cost. The consortium delivering the GBYP Biological Program currently uses an alternative form of GBS, which may or may not be adequate for determination of kinship to HSP level. Part of any second phase of this project should be a detailed evaluation of candidate genotyping techniques (already done for DArT, but not for others) to see whether HSPs can be found accurately and affordably with that technique. Given the interest in this rapidly developing area of technology, it would seem prudent to conduct a review and/or expert workshop with others who are already implementing similar methods or are considering initiating large-scale, high through-put genotyping for fisheries/natural resource management applications.

8 Preliminary design for EBFT: model structure and results

8.1 Model structure and assumptions

The model mimics a stand-alone CKMR assessment that uses fishery age compositions, total catch-at-age and the CKMR results (i.e. frequency, distribution and nature of close-kin pairs). It is fairly similar to the stand-alone model used for SBT, but differs structurally in terms of:

- stock structure;
- assumed existence of catch-at-age data (adult catches are a small proportion of the total catch for SBT, particularly relative to EBFT);
- use of HSPs and of AA-POPs, as well as AJ-POPs (AA-POPs are rare for SBT, due the substantially older age at maturity).
- various simplifications appropriate for a design study as opposed to a real post-hoc data analysis, e.g. concerning length data.
- Information on age of adult samples: a real-data CKMR analysis would need to use both length and age information (for adults). In practice this is not necessary for the design stage. For design purposes, we pretend that length does not matter, and instead that age does matter and is directly measured for all genotyped adults. This captures the general point, which is to allow for (and to be prepared to estimate) substantial somatic growth and change in fecundity after reaching adulthood. In the actual estimation model for a CKMR project, the age and length data are important and the length at age and associated variability are accounted for explicitly.
- Stock structure: for this first part of the feasibility study, we applied a model that corresponds to Heritable stock structure with 2 adult spawning grounds, and 2 juvenile sampling sites with mixing parameters ϕ of 0.3 & 0.7 (i.e. partial mixing, differing by juvenile site). In practice, choice of model would be data-dependent, as explained in section 6. The focus here is to demonstrate that it is statistically possible to do CKMR in a stock-structure-affected setting.
- Uncertainty was computed via the expected Hessian of the log-likelihood, as per standard statistical theory.

8.1.1 Some details of model input and assumptions

Our demographic model starts in 2009. We assume that total catch-at-age exists, or can be accurately hind cast, back to 2009, and into the future.

To seed the model, we need some idea of true numbers-at-age. As the current ICCAT VPA fails to converge when the plus-group is set much above 10yo it was necessary to develop an appropriate approach to construct numbers-at-age out to a sensible plus group, in this case 35+ years. This is necessary for CKMR design as fecundity-at-age is a crucial determinant of ERRO (Expected Relative Reproductive Output); the average bodyweight of a 10yo is only about 1/3 of the asymptotic average weight, and the effective fecundity could vary at least as much. Hence, CKMR design requires numbers-at-age out to a more biologically "stable" plus-group, so we had to somehow split up the 10yo plus-group in the VPA out to age 35. While we tried to do this in a mathematically consistent way, the numbers inferred can be no more than informed guesswork.

Estimation also requires simulated age compositions for adults, which we assumed would come from the same spawning-ground fisheries from which genotype samples are taken. We assumed an *equivalent* sample size of 1000 fish per year. In practice this would not necessarily mean 1000 otoliths; length compositions also provide some information on age, though probably not enough on their own for CKMR, hence the priority on obtaining at least enough otoliths to develop precise length at age relationships for each spawning area sampled.

Juvenile age: for this initial trial, all juveniles were assumed to be age 2, without error (assumed to be inferred from length, via an annual age-length key, rather than specifically from reading otoliths).

Sex: this is important for CKMR. For SBT, adult size-at-age and effective-fecundity-at-age both differ by sex (Farley et al., 2014). This life history trait has also been demonstrated for other tuna, such as albacore (Farley et al., 2013). In the absence of information on EBFT, we assumed no sex-specific differences in growth or size at maturity, and equal sex ratios in abundance. We did not estimate extra parameters for males, whereas in a real analysis one would do so; this means the results presented are likely to *slightly* overstate the precision.

Natural mortality m: we assumed an asymptotic relationship to age, with lower and upper limits as estimable parameters.

E/W assignment: we assumed 100% reliable discrimination of EBFT vs WBFT, for simplicity at this stage of the design exercise. The impact of this can be explored in more detail as part of the more detailed feasibility study.

HSP detection rate: we assumed 25% of HSPs will be missed. In the case of HSP identification, a cut-off level needs to be set to avoid false-positives, which may unavoidably lead to some proportion of HSPs being missed as false-negative. That proportion cannot be predicted, but it can be estimated after the event from the genetics; independent of the demographic model. Having done this, it can be treated as a known parameter in the demographic results. The real percentage of HSP lost is highly likely to be between 0% and 50%, so we have assumed 25%. Sensitivity to this assumption is likely to be less than for other assumptions we have had to make to this stage of the design exercise.

Recruitment variability: it is essential to use some prior on recruitment to the adult population (remembering that CKMR only provides information on the adult component of the population, which for CKMR purposes is defined as age-of-first-maturity, i.e. 3 for EBFT); otherwise there is almost no information to infer abundances towards the end of a study. Based on VPA results for 1950-2010, estimates, we took $\mathbb{V}[\log R_t] = 0.3$; estimates since then show more variability, presumably due to a well-known artefact of VPA behaviour.

Future average recruitment (of adults, i.e. 3yo): we used the average VPA estimate over 2009-2013. This results in a high figure compared to longer-term averages and we note there is a measure of skepticism about the reliability of recent estimates of recruitment. If it is the case that the recent estimates of recruitment are upwardly biased, then our results may be "pessimistic" from a design perspective. That is: if there are fewer young fish than we have projected, then fewer genotyped samples will be needed to get the same level of precision in practice.

8.2 Results

The results are summarized first for this "base case":

- Abundance measured in terms of SSB(2014), i.e. biomass of 3yo and up in 2014.

- Genotyping deliberately "over-sampling" old (large) adults[12]

Table 3 shows the expected CV on the spawning biomass for different combinations of annual adult and juvenile sample sizes and study durations (3-5 years) for the base case. These suggest it is possible to achieve a CV of about 15% for under 30,000 samples, even for the shortest duration considered. Aiming for a "looser" (larger) CV would be a false economy, especially in a study that might be a one-off (unlike, say, an annual abundance index). Longer studies actually give slightly better CV-per-sample, and are expected to provide better estimates of other demographic parameters (see below). As to proportion of adult to juvenile samples: a reasonably even mix is preferable in terms of precision-per-sample, if logistically feasible.

Table 3: Base case SSB

B2014 CV%; 3-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	29	20	16	13	11
4000	23	16	13	11	9
6000	20	14	11	10	9
8000	18	13	10	9	8
10000	16	12	10	8	7

B2014 CV%; 4-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	23	16	12	10	9
4000	18	13	10	9	8
6000	16	11	9	8	7
8000	14	10	8	7	6
10000	13	10	8	7	6

B2014 CV%; 5-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	20	13	10	8	7
4000	16	11	9	7	6
6000	13	10	8	7	6
8000	12	9	7	6	5
10000	11	8	7	6	5

Table 4: Base case kin

PO_AA; 3-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
any	2	9	21	37	58
PO_AJ; 3-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	9	18	26	35	44
4000	18	35	53	70	88
6000	26	53	79	105	131
8000	35	70	105	140	175
10000	44	88	131	175	219

PO_AA; 4-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
any	5	18	42	74	116
PO_AJ; 4-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	14	29	43	57	71
4000	29	57	86	114	143
6000	43	86	128	171	214
8000	57	114	171	228	285
10000	71	143	214	285	357

PO_AA; 5-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
any	8	31	70	124	194
PO_AJ; 5-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	21	42	63	84	104
4000	42	84	125	167	209
6000	63	125	188	251	313
8000	84	167	251	334	418
10000	104	209	313	418	522

HS_JJ; 3-yr; Geno N1	
nJ\nA	any
2000	6
4000	25
6000	56
8000	99
10000	155

HS_JJ; 4-yr; Geno N1	
nJ\nA	any
2000	12
4000	50
6000	112
8000	199
10000	311

HS_JJ; 5-yr; Geno N1	
nJ\nA	any
2000	21
4000	83
6000	186
8000	331
10000	516

Table 4 summarizes the result for the number of expected close-kin pairs associated with the different designs considered. The duration and samples-per-year also affect the number of close kin-pairs expected. These results are an important common-sense indicator of the likely viability of alternative designs in practice, because "count data" is intrinsically rather noisy when expected values are low. While there are some designs in Table 3 which may appear to yield satisfactory CVs for B2014 based on fewer than, say, 20 kin pairs of one type, such designs are likely undesirable in practice, because there may not be enough kin-pairs of every type to allow the appropriate model to be chosen reliably.

In this respect, the limiting factor is likely to be achieving sufficient AA Parent-Offspring, as these are essential to estimating the form of stock structure (H/N/U). For these preliminary calculations, we have assumed that the appropriate stock structure is "known", so that we are applying the "right" model. This will not be the case in reality, so it is important

[12]Specifically, we assumed that number-at-age-genotyped is proportional to age multiplied by number-at-age-in-population. In practice one would of course need to specify a subsampling scheme based on fish size, using some guess as to likely numbers-at-age; the details don't matter, but the general point of biassing somewhat towards older/bigger adults is important.

to aim to have a reasonable chance of reaching 20-30 AA POPs, in order to distinguish between different population structures and thus fit the right model for abundance estimation.

Juveniles are also important, because HSPs are central to being able to separate selectivity, mortality and spawning abundance. "Skimping" on HSPs (which would mean sampling few juveniles but many adults) would impair model checking and general reliability, whatever the nominal CV from any *single* model (as in Table 3) might be. There are substantial gains in the expected number of HSPs from longer studies, in part because within-cohort comparisons have to be excluded for HSPs, so that one unavoidably "loses one year" in HSP terms from any study; losing 1 out of 3 is worse than losing 1 out of 5.

Table 5: Sampling younger adults

B2014 CV%; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	13	10	9
6000	11	9	8
8000	10	8	7

B2014 CV%; 4-yr; Geno N0			
nJ\nA	4000	6000	8000
4000	13	10	9
6000	11	9	8
8000	10	9	8

B2014 CV%; 4-yr; Geno C			
nJ\nA	4000	6000	8000
4000	17	14	12
6000	15	12	10
8000	13	11	9

PO_AA; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
any	18	42	74

PO_AA; 4-yr; Geno N0			
nJ\nA	4000	6000	8000
any	19	43	76

PO_AA; 4-yr; Geno C			
nJ\nA	4000	6000	8000
any	3	7	12

PO_AJ; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	57	86	114
6000	86	128	171
8000	114	171	228

PO_AJ; 4-yr; Geno N0			
nJ\nA	4000	6000	8000
4000	34	51	68
6000	51	76	101
8000	68	101	135

PO_AJ; 4-yr; Geno C			
nJ\nA	4000	6000	8000
4000	82	123	164
6000	123	185	247
8000	164	247	329

Table 5 demonstrates the impact of emphasizing, or not, the sampling of older fish. For CKMR on teleosts, it is *essential* to sample the full age range (otherwise fecundity-at-age cannot be estimated, and becomes confounded with abundance); however, one can choose whether to shift the main sampling effort towards older or younger fish. The results indicate it is likely to be more effective to "bias" the genotyping of adults towards larger/older adults. The "N1" base-case was presented in Table 3. "N0" also shifts towards older adults, but less so "genotype in proportion to N-at-age" (not further multiplied by age, so less focused on older adults). "C" means genotyping random samples from the adult catch, and leads to samples dominated by 3yo and 4yo fish. The results indicate that C is definitely likely to be worse in terms of the CV on abundance, even though it pushes up the number of AJ POs. The problem with C is that it returns much less information on fecundity-at-age because there are fewer AA PO pairs. This additional uncertainty in fecundity-at-age is propagated though to the abundance estimates; in addition, the AA POs are essential for stock structure inference.

Table 6: Alternate measures of adult abundance

B2012 CV%; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	16	14	12
6000	15	13	11
8000	13	12	11

B2014 CV%; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	13	10	9
6000	11	9	8
8000	10	8	7

B2019 CV%; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	17	14	13
6000	15	13	11
8000	14	12	10

N2019 CV%; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	21	20	19
6000	18	17	17
8000	17	16	15

Table 6 compares alternative ways of measuring adult abundance, and shows that B2014 has a slightly better CV than the other abundance measures considered. The message is really that the CV is not very sensitive to which measure is chosen, except perhaps that N2019 is noticeably less precise. The N-measures are more dominated by younger adults e.g. 3yo, and there is unlikely to be much information about the newest cohorts at the end of the study periods considered.

Table 7: Other things of interest

logTrend10yr SE; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.32	0.29	0.26
6000	0.29	0.27	0.24
8000	0.28	0.25	0.23

log_fec_ratio SE; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.38	0.30	0.24
6000	0.33	0.26	0.22
8000	0.29	0.23	0.19

logM_10 SE; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.32	0.29	0.27
6000	0.28	0.25	0.23
8000	0.25	0.22	0.21

Table 7 summarises the impact of sample size on the SE of three non-pure-abundance parameters: estimated *trend* in spawning biomass over a ten-year period; fecundity-at-age; and natural mortality of adults. This illustrates that these important stock assessment parameters are in-principle estimable, even if not very precisely in a 4yr study. As an example, for `logTrend10yr` : 0.32 (for the 4000Jx400A) means, more or less, that a change of 88%, or more, ($= \exp(1.96 * 0.32)$) in spawning biomass over the 10yr period 2009-2019 would be expected to be detectable in a CKMR analysis as statistically significant at the 95% level. The `log_fec_ratio` is the estimated difference in *per-capita-per-kg-of-bodyweight* effective fecundity between a 15yo and 5yo adult; while an SE of 0.3, for example, is hardly precise, it is enough to distinguish between instant maturity (i.e. "knife-edge" at 3yo) and a gradual increase in effective relative reproductive output with age.

Table 8: Effect of duration on non-abundance parameter

log_fec_ratio SE; 3-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.51	0.40	0.33
6000	0.43	0.34	0.29
8000	0.39	0.31	0.26

log_fec_ratio SE; 4-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.38	0.30	0.24
6000	0.33	0.26	0.22
8000	0.29	0.23	0.19

log_fec_ratio SE; 5-yr; Geno N1			
nJ\nA	4000	6000	8000
4000	0.30	0.24	0.19
6000	0.26	0.21	0.17
8000	0.23	0.19	0.16

Table 8 demonstrates the benefits of a longer study in terms of estimating non-abundance parameters (here, the age-specific fecundity effect also shown in Table 7). This is evident from the substantial reduction in the estimated SE from the 3yr (left) to 5yr (right) study.

Table 9: Sensitivity to assumed equivalent age-composition sample (500 LHS, 100 RHS per spawning ground per year)

B2014 CV%; 4-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	23	16	12	10	9
4000	18	13	10	9	8
6000	16	11	9	8	7
8000	14	10	8	7	6
10000	13	10	8	7	6

B2014 CV%; 4-yr; Geno N1					
nJ\nA	2000	4000	6000	8000	10000
2000	24	16	13	10	9
4000	19	13	11	9	8
6000	16	12	9	8	7
8000	15	11	9	7	7
10000	13	10	8	7	6

Finally, the results so far have assumed that the age-composition-data from the adult fisheries (*nothing to do* with the sub-sampling schedule for genotyping; this is standard fishery age-composition data, which is essential for CKMR) is equivalent to knowing the exact ages of 500 adults per spawning ground (of which there are assumed to be 2 in these simulations). That equivalence could come entirely from otoliths, or from some combination of otoliths, length compositions, and reliable up-to-date statistical analysis (which excludes "cohort slicing", for example). There is really no reliable way to guess what the appropriate "equivalent sample size" might be; this is something to think about hard in a detailed design. For now, though, just to check whether the assumption of 500 has much effect, Table 9 shows how the CVs change if the equivalent sample size turned out considerably smaller. Fortunately, there is little effect, *at least down to this level*.

8.3 Summary

This is the very first attempt to allow quantitatively for stock structure in a CKMR design or analysis (not just for EBFT). The keys to this project have been to develop an appropriate statistical framework for CKMR in the presence of stock-structure, to show that certain stock-related parameters are estimable in principle (section 6), and to demonstrate that a complete stock-ready model can actually be constructed and estimated based on reasonable data. And that does turn out to be true: the statistical model (including estimation of stock-specific abundances, and of mixing parameters ϕ) is fully estimable (i.e. a positive-definite Hessian showing no ill-conditioning), and the sample sizes do not seem exorbitant (e.g. compared to SBT, allowing for the greater general abundance of EBFT).

Any estimates of sample sizes for a design study for EBFT CKMR will be inevitably imprecise, primarily because of:

- great uncertainty about true numbers-at-age from the assessment over the next few years;
- uncertainty about fecundity-at-age (since young adults are so numerically dominant in the current VPA outputs);
- having no prior idea of the extent of mixing from different spawning grounds in the juvenile fisheries (we assumed "partial mixing", but the reality may differ);
- not knowing what other data would be available in practice (e.g. total catch, total catch-at-age breakdown, extent of age-composition data from spawning ground fisheries).

Specific reasons why our suggestions might be "pessimistic" (i.e. liable to indicate higher sample sizes than really will be needed) include:

- assuming that recruitment over the next few years remains at the high levels of the most recent (and least certain) VPA *estimates;*
- assuming that effective fecundity is directly proportional to bodyweight as soon as EBFT reach age 3; if this is not the case— e.g. because only some young fish are actually mature, or because younger fish cannot spend as long on the spawning grounds— then the "equivalent breeding population" is numerically smaller, and kin-pairs will be more common.

Specific reasons why our suggestions might be "optimistic" (i.e. liable to indicate lower sample sizes than really will be needed) include:

- assuming that total catch, and its age breakdown, is reasonably well known. (The current cohort-sliced VPA inputs would not qualify.) If not, it is still possible to apply CKMR, by estimating total z and overall selectivity— the model can be made statistically estimable, provided that age composition data from *some* adult fishery can be obtained, even if total catch across *all* fisheries cannot be. But it seems likely that more parameters would be needed (though we have already allowed for estimating m-at-age), and that the information content on abundance might be appreciably reduced. This whole aspect of what other data could be available— on the implicit assumption that the effort was already being made to do a CKMR study— would require more detailed consideration in a follow-up study.

Although not too much credence should be placed in these results, for the reasons noted above (they are a starting point for considering whether CKMR has merit for EBFT, and not a final sampling design) they do provide a reasonable indication of the minimum sample size required to generate a reasonably precise result. It is better to collect as many as possible, and then decide later *not* to genotype them all (that being the most expensive step) if the number of kin-pairs turns out much higher than expected from processing an initial proportion. The worst possible scientific study is the one that spends a substantial sum, but not enough to answer the question.

8.3.1 Longer versus shorter duration

Total sample size required to achieve a given target CV on abundance seems roughly similar for a 3–, 4–, or 5–year study. Nevertheless, our experience from SBT (where the study had to be prolonged because the abundance turned out to be much higher than expected, so that more time was required to collect enough samples to get a precise result) is that a longer study is better in terms of ironing out difficulties and providing better information on model selection, mortality, fecundity, and other not-strictly-abundance parameters, all of which are in some sense confounded with abundance. In general, we consider that CKMR studies for fisheries are best thought of as longer-term monitoring tools, rather than one-off "anchoring" exercises; there are great efficiency gains if one already has several years of samples "in the bank", so that ongoing sample sizes can be kept considerably lower than is required to get an initial estimate quickly.

8.3.2 Sampling of larger/older adults

Concentrating genotyping efforts on larger/older adults is helpful both qualitatively and quantitatively. Most of our simulated adult samples are younger animals because of recent estimates of high recruitment and projections that continue at that level. The benefits of deliberately oversampling bigger adults (at least if the age structure is close to what we have assumed) are that:

- It will increase the number of AA PO pairs, because younger adults are unlikely to be parents yet. These are the least-common kin-pair otherwise, but yield the most information about Heritability of stock structure. In other words, having plenty of AA-PO pairs will be very useful for *choosing* the right estimation model in the first place, if the project proceeds to implementation.
- It will yield more precise estimates of fecundity-at-age, because the age coverage will be better.

Notwithstanding all that, it is also important to sample plenty of juveniles, since that is where HSPs come from, and HSPs are essential to disentangling the effects of mortality, fecundity, and abundance per se.

8.3.3 Stock structure

Stock structure (number of spawning units, and number of places where samples are taken) will not make much difference to overall sample size required for a given level of precision; we give qualitative recommendations on that elsewhere. In quantitative terms it would be reasonable to pro-rate the annual sample sizes of adults/juveniles across the final selection of fisheries of appropriate type to source the highest quality adult and juvenile samples. The key point is the same total number of adults are breeding whatever the underlying structure, so the expected number of kin-pairs is unaffected by the distribution of sampling effort, provided sampling is spread out evenly; it is the *pattern* among the kin-pairs that informs on stock structure, stock-specific abundances, and mixing.

8.3.4 Design framework

A computational framework now exists that can be used to consider alternative designs, should the project proceed to the second stage. The extent of further design refinements should be tempered by the availability of "CKMR informative" data and information. There is no point in investing in additional detailed quantitative design work if there is no additional information available to address the substantial unknowns identified in this initial scoping study. Provided that sampling is adequate quantitatively (as in this section) and qualitatively (as elsewhere in the report), then in a few years' time (i.e. before the end of a 5-year study, for example), there should be sufficient data (POPs, HSP and associated age, length, sex data) and experience to refine the design and re-assess the sample sizes and study duration.

9 Summary and recommendations for design of Close-kin Mark Recapture for EBFT

9.1 Summary

The more complex population structure of ABFT, and EBFT in particular, means the design of an appropriate sampling program and estimation framework for CKMR for ABFT is considerably more complex that the application to SBT or PBT. Nevertheless, an appropriately-designed and carefully-implemented sampling program, that samples multiple spawning and juvenile grounds, and that genotypes thoroughly enough to find HSPs, as well as POPs, should provide the close-kin data required to disentangle stock structure, selectivity, fecundity-at-age, mortality, and adult abundance. There seems little doubt that the information gleaned would be sufficiently valuable to justify taking the next step, namely a more detailed review and design exercise.

9.1.1 General strategy for CKMR

CKMR is much more efficient (from an information content gained for investment made sense) as a long-term program than as a one-off abundance estimate. The information content of a CKMR study is determined by the number of POPs/HSPs found; this depends mostly on abundance, which is of course unknown when doing the design exercise. Because each new sample gets compared to every pre-existing sample, the effective "sample size" (number of comparisons) grows quadratically

as the study continues— in contrast, say, to the linear accumulation of information from an annual trawl survey. Hence, in the long term, annual sample sizes for CKMR can be kept quite low, just enough to keep the "information tank topped up". Conversely, in the short term at the start of a project, it is necessary to sample rather more heavily in order to get enough POPs and HSPs to provide an "informative" first estimate, build an appropriate CKMR estimation model, and refine the design for the longer term (e.g. for SBT Bravington and Davies, 2013, Bravington et al., 2015). Since tissue-sampling is usually cheaper than genotyping, it makes very good sense to collect a lot of samples (i.e. tissue, body length, age and sex of adults) early in the sampling program, even if not all tissue samples ultimately require genotyping. In the case of EBFT, at least 3 years of sampling of sufficient intensity will be required to provide an initial abundance estimate (useful juvenile comparisons can only be made among separate cohorts, and the potential of skip-spawning means that three years may be needed to see two cohorts of HSPs from the younger parents), although with some luck there may be some information on structure earlier on.

9.1.2 Importance age and length data

Along with tissue samples, it is essential that representative age- and length-composition data is collected from at least one fishery on each spawning ground. Without that, it becomes impossible in CKMR to disentangle growth and fecundity from abundance (and it is also impossible to do a reliable conventional stock assessment without such data). It is not necessary to collect age information for the juvenile samples, as long as age can be accurately inferred from precise age-length keys; and assuming they can be clearly identified as 2 year-olds by the modal length of their year class. If there is evidence of difference in size-at-age among juvenile feeding grounds, then separate age-length keys will be required for each juvenile sampling area.

9.1.3 Sample first process later

Our review of the sampling protocols and tissues preservation indicates that they should provide high quality samples suitable for CKMR. CKMR genotyping requires high-quality tissue; particular attention should be paid to field and lab preservation/archive methods, hence this should be one focus of any follow-up project. The primary issues identified with existing data, are that the sample sizes per strata so far are too low to provide an abundance estimate, and it is also not clear whether the required age and length composition data are currently available for the adult samples from the spawning grounds. However, these are not fundamental problems, and could presumably be addressed in future. Continued attention should be paid to potential for cross-contamination when collecting samples in the field (clean/change dissection equipment between individuals) and QC of the data collection and management processes in the field, when archiving and genotyping— experience has taught us that the chance of mix ups is high, and kin-finding is not a fault-tolerant process.

9.1.4 Sample sizes and cost

It is not possible to provide definitive costs for project management, sample collection, genotyping, age determination, model development, analysis, and reporting, at this stage of project design. This reflects (i) the uncertainty about the true state of knowledge of EBFT (abundance, extent and nature of any population structure, fecundity-at-age, etc.), which would be updated during the course of a real CKMR program; (ii) the many logistic decisions and trade-offs on where and what to sample that would need to be considered before moving to full-scale implementation (identified as stage 2 of the current project), and (iii) the genetic and data-analytical costs. We have demonstrated that there are many different sampling designs that could achieve a respectable precision (more adults, less juveniles; more on one site than another; etc.) and a final choice between these designs requires the more detailed follow-up, planning and costing identified for stage 2 of the design study, should the project proceed.

Nevertheless, given the samples size calculations provided in section 8, it is possible to provide a qualitative indication of the likely scale of the project budget. The total budget for the CKMR project for southern bluefin tuna was in the order of $AUS 1.5M. This included sample collection, marker development, processing and genotyping of ~14,000 individual SBT, development of the estimation model and reporting. Based on the initial sample sizes required for a CV of 15-20% reported here, about 1.5–2 times as many samples are likely to be required for EBFT. This is larger than for SBT because the EBFT abundance is (thought to be) substantially higher. However, it is not as much bigger as one might expect based purely on abundance estimates, for two main reasons: first that HSPs (not part of the original SBT analysis) would be an essential part of EBFT CKMR, so each juvenile sample "generates" many more kin-pairs; and second that EBFT are considered to reach maturity much younger than SBT, which means more information among adult-adult comparisons. Although allowance for possible population structure is a major complication for EBFT, in terms of imposing requirements

on sampling and statistical analysis, it should not actually affect total sample sizes very much (as long as the precision criterion is on *aggregate* abundance, rather than *per-stock* abundance).

Aside from sample size per se, an important difference from the earlier SBT application is the subsequent development of GBS and the associated reduction in cost of large-scale genotyping. This will result in substantial reductions in both the development costs and the per unit genotyping cost. For example, in the case of SBT, the development of the specific micro-satellites alone was ~$AUS250k (almost eliminated with GBS); and the current cost of genotyping-per-individual has more than halved, and will continue to decline. The specific costs associated with GBS genotyping will vary depending on the specific method selected and the read-depth and quality control procedures employed by the lab providing the service. Notwithstanding this, on the basis of the sample size calculations presented here and our understanding of the per unit cost of suitable GBS genotyping approaches, the sample processing and genotyping for a 4-5 year CKMR study might cost a similar total amount to that for SBT project, i.e. ~Euro 200-250k/year, not including the cost of field sampling. Sample size requirements might need to be changed considerably in the light of interim results from the project, so this figure is indicative and can never be precise at the outset. It also does not include the cost/resources required for sample collection and data analysis and development of the bespoke estimation model. A more detailed estimate of the likely cost of each component and overall project cost is identified as an output of the second stage of the current project, should it proceed.

9.1.5 Population marker

A population marker that can accurately discriminate between western and eastern populations of ABFT is essential for effective implementation of CKMR (and encouraging results have been found). However, it is not necessary to have a population genetic marker to discriminate among populations within the Mediterranean. Provided that the sampling program covers all of the potential spawning entities well enough (i.e. consistent coverage over years, with sufficiently large sample sizes and ancillary data), then the POP and HSP data will reveal the actual underlying population structure relationships and provide quantitative estimates of exchange between spawning grounds and juvenile areas.

9.1.6 Close-kin Mark Recapture data in Assessment & MSE

The properties of the close-kin data (i.e. POP, HSP; and associated covariates such as age, length, sex) are such that they can be incorporated directly (via a modified mark-recapture model) into the likelihood of statistical catch-at-age/length models for stock assessment (e.g. Hillary et al., 2012; Hillary et al., 2013) and operating models for evaluation of harvest strategies/management procedures via MSE (e.g. CCSBT ESC, 2015). In the ICCAT context, where a VPA is used as the primary assessment model, it will not be possible to incorporate the CKMR data directly into the current assessment model.

The operating model being developed for MSE, as part of the GBYP modelling and MSE work program (Carruthers et al 2014; Butterworth et al 2016), would be able to accommodate the CKMR data. It is likely that these data would be extremely informative and valuable for this purpose, if and when they become available; and, particularly, if they can be obtained for both the east and western populations.

A stand-alone CKMR "mini-assessment", constructed along the lines of SBT to estimate adult abundance, mortality and population structure, and using CKMR data along with age- and length-compositions from the sampled spawning ground fisheries, can be used to monitor the adult population, quite independent of the primary stock assessment used for TAC setting and or MSE. However, CKMR itself is intrinsically limited to providing estimates of quantities associated with adult fish. That is, it does not provide estimates of recruitment to the juvenile component of the population, nor harvest rates on the juvenile and sub-adult components, which constitute a substantial proportion of the catch for EBFT (though much less than for SBT, at least for the years prior to the enforcement of a minimum size and a quota.). Even though some EBFT do mature as young as age 3, CKMR is unlikely to be able to provide cohort-specific estimates until the cohort has grown through many years (in effect, until there are enough POPs per adult cohort to make meaningful inferences). Depending on the quality of age-composition data (which do contain some information about cohort strength), it may be that specific abundance information on younger cohorts is also needed for effective management. The next subsection describes one possible way to get that young-cohort information

9.1.7 Combining Close-kin Mark Recapture and gene-tagging

Gene-tagging, in principle, is simply conventional mark-release-recapture where the function of the plastic "spaghetti tag" is replaced by a tissue sample from the fish at release (which "marks" the fish) and a large number of "recovery" tissue

samples at some point along the post-capture supply chain. The substantive advantages are: i) if done properly, taking of the tissue sample leaves no visible "mark", hence there is no issue with reporting rates; ii) the "marks" are permanent so there is no need to estimate, or account for, tag loss; and iii) the expected number of "recaptures" is determined by both number of "marks" released and the proportion of the catch sampled for "recaptures" (and obviously the size of the population of interest). Hence, from a design perspective it is possible to consider different permutations of "release" and "recovery" effort to optimise the precision of the quantity of interest (e.g. abundance of recruits) (see Preece et al. (2013) and Preece et al. (2015)).

The CCSBT has recently funded an operational trial of gene-tagging for juvenile SBT as an alternative method of recruitment monitoring to the scientific aerial survey (CCSBT ESC, 2014; CCSBT ESC, 2015). The recruitment index from the scientific aerial survey (relative abundance of 2-4 year olds, see Eveson et al., 2015) is the only long-term fisheries independent abundance index for the population and is one of the two input series for the current CCSBT Management Procedure (the other is standardised Japanese longline CPUE , as per Itoh et al. (2011), CCSBT ESC (2013), and Hillary et al. (2015)). The cost and logistic vulnerability of the scientific aerial survey have been of concern to the CCSBT Commission and ESC and gene-tagging is seen as the only viable alternative for a quantitative index of recruitment in the short to medium term (CCSBT ESC, 2013). If the large-scale field trial (2016-2018) is successful (see Preece et al. (2015)and CCSBT ESC (2015) for specific details of the trial design), the CCSBT intention is for annual gene-tagging of 2 year olds to replace the scientific aerial survey from 2018. If this happens, absolute estimates of abundance of 2 year olds would replace the relative index of abundance of 2–4 year olds from the scientific aerial survey as part of the transition to a new MP for recommending the global TAC (CCSBT ESC, 2015; CCSBT, 2015.

In addition to the investment in the large-scale gene-tagging trial, CCSBT and CSIRO have invested in development and design work to determine the cost-effectiveness of using CKMR as a long-term monitoring series for the spawning component of the population, that is independent of the catch and effort of the main targeted fisheries (Bravington and Davies, 2013, CCSBT, 2013a, Bravington et al., 2015). As a result, the CCSBT has committed to the ongoing collection and genotyping of samples of adults and juveniles to provide the basis for periodic CKMR estimates of adult abundance based on the POP+HSP designs outlined in Bravington et al. (2015) and an estimation framework being developed by CSIRO. In this context, the combination of gene-tagging, for the juvenile component of the population, and CKMR, for the spawning component of the population, provide the necessary monitoring series for the development of a long-term MP that is largely fisheries independent and wholly CPUE independent.

Again, in principal, there are parallels in the potential application of gene-tagging to ABFT; however, the population, fishery and geo-political complexities of design and implementation are not insignificant and may well be prohibitive. A detailed examination of the issues is beyond the scope of this CKMR scoping exercise. However, key considerations include:

- What is known and/or can reasonably be assumed about mixing of juveniles, sub-adults and adults from both eastern and western populations?
- What is known or can reasonably be assumed about missing juveniles spawned on different grounds within the Mediterranean?
- From which fisheries is it possible to "mark and release" large numbers of "known-age" fish (i.e. take a tissue sample and measure length)?
- From which fisheries, or points along the processing chain, is it practically possible to obtain large numbers of "recovery" samples (1,000s-10,000s) for juveniles and adults?

From considering the above, and reflecting on section 5, it should be evident that proceeding with a number of years of "preliminary sampling" and "exploratory genotyping" for CKMR study to provide adult abundance, population structure and connectivity for the eastern population alone, should provide valuable information for assessing whether or not gene-tagging is likely to be logistically feasible for ABFT and essential information on population structure and connectivity for design of a pilot gene-tagging study, should it be considered feasible.

An additional consideration, is the design and implementation of the genotyping for gene-tagging and the interaction between this and selecting a GBS platform for CKMR. Identification of POPs with the extremely low level of error required for CKMR is a substantially more difficult task (genetically and statistically) than matching an individual to themselves for gene-tagging: there are few markers required and, hence, the cost per fish is substantially less for gene-tagging than for CKMR (see Preece et al. (2015) and Bravington et al. (2015) for specific comparisons for SBT). The specific costs do depend, however, on the particular GBS platform employed and the extent to which the process is designed to "double dip"

for both gene-tagging and/or CKMR at the same time. In short, you can "double dip" on the tissue sampling and the DNA extractions for both gene-tagging and close-kin, but the extent to which you can do so for the genotyping step depends on the GBS approach employed. This is a substantive technical issue (from both a statistical and genetics perspective) that we would recommend be considered in some depth (see recommendation 1 and 2 below) as it has substantial cost and "capability" implications for long-term consistency and quality of the required data streams.

9.2 Recommendations

Given the view expressed above that: i) there is scope for CKMR to significantly improve the data and understanding available to effectively assess the status of ABFT, and EBFT in particular; and ii) assuming there are sufficient resources and institutional commitment to modify and expand the current level of biological sampling completed under the GBYP to the level required to obtain an informative number of close-kin (POPs and HSP) and associated ancillary data, we recommend the following activities in order of priority:

1. Determine the most cost-effective form of genotyping that can demonstrably identify HSPs. By cost-effective, we mean the GBS method that can provide the required level of genotyping reliability required to consistently identify HSPs for the lowest cost per fish (note that if the method can do this for HSPs, it can necessarily do it for POPs).

2. Consideration should be given to doing 1 in conjunction with a workshop that includes expertise from a range of other areas that are active in large-scale, high through-put genotyping for applied fisheries and/or natural resource management purposes (e.g. Pacific salmon, the FishPopTrace Consortium, and CSIRO) to learn from their experience and share the cost involved in evaluating alternative GBS platforms in a very rapidly developing and technically complex field.

3. In consultation with GBYP Coordination, select juvenile and adult sampling locations for an "initial round of CKMR sampling" that are consistent with preliminary design options (i.e. Table 2), and initiate sample collection as soon as possible. These samples can, in the short-term, be archived and/or, used to develop genotyping and data processing work-flows and quality control procedures for identifying kin and validating genetic stock discrimination markers.

4. Commence an expertise-based process to review and identify candidate markers (genetic and microchemical) for assigning samples to eastern and western populations. While it may be appealing to include "within Med" markers as part of this exercise, this is not necessary for the purposes of CKMR, and there is no virtue in waiting for the (uncertain) outcome of a within Mediterranean marker search before starting CKMR. As noted in section, the CKMR data will reveal the population structure in the Mediterranean, as long as the sampling of spawning grounds and juvenile areas is sufficiently comprehensive. The final E-W candidate(s) markers, including assignment probabilities, should be decided based on a validation study conducted with known origin fish of sufficient sample sizes to provide statistically reliable estimates of assignment probabilities.

Finally, it is important to recognise that design and implementation of CKMR requires a combination of both broad (fisheries biology, field and laboratory logistics, statistics, mark-recapture theory, population dynamics, population genetics and genomics, applied stock assessment) and deep knowledge and expertise (in this case, in ABFT population biology and fisheries, CKMR design and implementation). In the (relatively) early stages of the development and implementation of this new approach (i.e. CKMR) it will be important to consider the best mechanism (contracting and institutional) to establish and maintain a suitable experienced and qualified team for design and implementation to deliver high quality and robust results in the short-term and, if successful, the development of the necessary capability to maintain an ongoing program into the future.

10 Acknowledgements

We would like to acknowledge the contributions of the ICCAT BFT scientists who provided their expert input on to this scoping exercise for EBFT, in particular, Haritz Arrizabalaga, Sylvain Bonhommeau, Tom Carruthers, Doug Butterworth and Antonio Di Natale, and Peter Grewe and Pierre Feutry for their input on details of GBS approaches.

11 References

References

Arrizabalaga, H et al. (2014). *Short term contract for the biological and genetic sampling and analysis (ICCAT-GBYP 02/2013) within the GBYP (Phase 4)*. Tech. rep. ICCAT.

Bravington, M and C Davies (2013). *Close-kin for SBT: where to now?* Tech. rep. CCSBT-ESC/1309/17. Commission for the Conservation of Southern Bluefin Tuna.

Bravington, M (2014). *Close-Kin Mark-Recapture for SBT: options for the longer term*. Report for Scientific Committee CCSBT-ESC/1409/44. Commission for the Conservation of Southern Bluefin Tuna.

Bravington, MV, HJ Skaug, and EC Anderson (2016). "Close-Kin Mark-Recapture". In: *Statistical Science*.

Bravington, M, P Grewe, and C Davies (2012). *Report of the Close-Kin Project: estimating the absolute spawning stock size of SBT using genetics*. Scientific Committee Report CCSBT-SC/1208/19. Commission for the Conservation of Southern Bluefin Tuna.

Bravington, M, J Eveson, P Grewe, and C Davies (2015). *SBT Close-Kin Mark-Recapture: options for the medium term*. Tech. rep. CCSBT-ESC/1509/19. Commission for the Conservation of Southern Bluefin Tuna.

Bravington, M, P Grewe, and C Davies (2014). *Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers*. FRDC Report 2007/034. CSIRO, Australia.

CCSBT (2009). *CCSBT Report of the Fourteenth Meeting of the Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

— (2013a). *CCSBT Report of the Eighteenth Meeting of the Extended Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

— (2013b). *Report of the Fourth Operating Model and Management Procedure Technical Meeting*. Workshop Report. Commission for the Conservation of Southern Bluefin Tuna.

— (2015). *Report of the Twenty Second Annual Meeting of the Commission*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

CCSBT ESC (2006). *Report of the Eleventh Meeting of the Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

— (2013). *Report of the Eighteenth Meeting of the Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

— (2014). *Report of the Nineteenth Meeting of the Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

— (2015). *Report of the Twentieth Meeting of the Scientific Committee*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

CCSBT OMMP (2014). *Report of the Fifth Operating Model and Management Procedure Technical Meeting*. Tech. rep. Commission for the Conservation of Southern Bluefin Tuna.

Davies, C, A Preece, T Polacheck, and M Basson (2007). *A review of the Commission's Scientific Research Program, and considerations of current priorities and ways forward*. Tech. rep. CCSBT-ESC/0709/16. Commission for the Conservation of Southern Bluefin Tuna.

Eveson, P, J Farley, and M Bravington (2012). *The aerial survey index of abundance: updated analysis methods and results for the 2011/12 fishing season*. Tech. rep. CCSBT-ESC/1208/16. Commission for the Conservation of Southern Bluefin Tuna.

Farley, JH, JP Eveson, TL Davis, R Andamari, CH Proctor, B Nugraha, and CR Davies (2014). "Demographic structure, sex ratio and growth rates of Southern Bluefin Tuna (Thunnus maccoyii) on the spawning ground". In: *PloS one* 9.5, e96392.

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

Fraile, I, H Arrizabalaga, and JR Rooker (2014). "Origin of Atlantic bluefin tuna (*Thunnus thynnus*) in the Bay of Biscay". In: *ICES Journal of Marine Science: Journal du Conseil*, fsu156.

Hillary, R, A Preece, and C Davies (2013). *Updates to the CCSBT operating model including new data sources, data weighting and re-sampling of the grid*. Tech. rep. CCSBT-ESC/1309/15. Commission for the Conservation of Southern Bluefin Tuna.

Hillary, R, A Preece, C Davies, M Bravington, P Eveson, and M Basson (2012). *Initial exploration of options for inclusion of the close-kin data into the SBT operating model*. Scientific Committee report (OMMP workshop) CCSBT-ESC/1208/21. Commission for the Conservation of Southern Bluefin Tuna.

Hillary, RM, AL Preece, CR Davies, H Kurota, O Sakai, T Itoh, AM Parma, DS Butterworth, J Ianelli, and TA Branch (2015). "A scientific alternative to moratoria for rebuilding depleted international tuna stocks". In: *Fish and Fisheries*.

ICCAT, GSC (2015). *Time to plan for the future of GBYP*. Tech. rep. Collective Volume of Scientific Papers, SCRS/2014/194. ICCAT.

Itoh, T, O Sakai, and N Takahashi (2011). *Standardized CPUE for Management Procedure in 2011*. Tech. rep. CCSBT-ESC/1107/30. Commission for the Conservation of Southern Bluefin Tuna.

Maunder, MN, JR Sibert, A Fonteneau, J Hampton, P Kleiber, and SJ Harley (2006). "Interpreting catch per unit effort data to assess the status of individual stocks and communities". In: *ICES Journal of Marine Science: Journal du Conseil* 63.8, pp. 1373–1385.

NOAA, N et al. (2015). *Outcomes from a Workshop (27-29 May 2015) on Developing Close-Kin Mark Recapture techniques for Pacific Bluefin*. Tech. rep. NOAA National Marine Fisheries Service, Southwest Fisheries Science Center, La Jolla, CA, USA.

Polacheck, T and C Davies (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*. Research paper 023. CSIRO Marine and Atmospheric Research.

Polacheck, T and P Eveson (2007). *Updated analyses of tag return data from the CCSBT SRP tagging program*. Tech. rep. CCSBT-ESC/0709/19. Commission for the Conservation of Southern Bluefin Tuna.

Polacheck, T (2012a). "Assessment of IUU fishing for Southern Bluefin Tuna". In: *Marine Policy* 36.5, pp. 1150–1165.

— (2012b). "Politics and independent scientific advice in RFMO processes: A case study of crossing boundaries". In: *Marine Policy* 36.1, pp. 132–141.

Preece, AL, CR Davies, and R Hillary (2014). *Southern Bluefin Tuna inter-sessional science 2014–15*. Tech. rep.

Preece, AL, C Davies, M Bravington, R Hillary, P Eveson, and P Grewe (2013). *Preliminary cost and precision estimates of sampling designs for gene-tagging for SBT*. Tech. rep. CCSBT-ESC/1309/18. Commission for the Conservation of Southern Bluefin Tuna.

Preece, AL, P Eveson, CR Davies, P Grewe, R Hillary, and M Bravington (2015). *Report on gene-tagging design study*. Tech. rep. CCSBT-ESC/1509/18. Commission for the Conservation of Southern Bluefin Tuna.

Richardson, DE, KE Marancik, JR Guyon, ME Lutcavage, B Galuardi, CH Lam, HJ Walsh, S Wildes, DA Yates, and JA Hare (2016). "Discovery of a spawning ground reveals diverse migration strategies in Atlantic bluefin tuna (/emphThunnus thynnus)". In: *Proceedings of the National Academy of Sciences* 113.12, pp. 3299–3304.

Rooker, JR, H Arrizabalaga, I Fraile, DH Secor, DL Dettman, N Abid, P Addis, S Deguara, et al. (2014). "Crossing the line: migratory and homing behaviors of Atlantic bluefin tuna". In: *Marine Ecology Progress Series* 504, pp. 265–276.

Skaug, H (2001). "Allele-sharing methods for estimation of population size". In: *Biometrics* 57, pp. 750–756.

WCPFC, NC (2015). *Report of the Fifteenth Meeting of the International Scientific Committee for tuna and tuna-like species in the North Pacific Ocean*. Tech. rep. WCPFC-NC11-2015/IP-01. WCPFC.

12 Appendices

12.1 Appendix 1: Terms of Reference for GBYP Project: GBYP 07c/2015

The Contractor shall provide a comprehensive report, including the following points:

1. Describe in a clear and synthetic way the close-kin genetic tagging and its uses for assessment purposes, including the MSE;

2. Overview of the close-kin genetic tagging activities carried out on tuna species in various areas;

3. An evaluation of the potential to apply close-kin genetic tagging method for obtaining estimates of the size of the spawning population for eastern Atlantic bluefin including sample size for various level of precision ranging from cv's of 10-30%;

4. A detailed experimental design including the steps and timeframe for the implementation such a program including realistic sampling options and strategies;

5. A comprehensive consideration of the assumption involved and how they might be tested and dealt with to ensure that robust estimates are obtained (e.g. stock structure; skipped spawning and relative spawning potential;
6. The feasibility and benefits of combining a close-kin genetic tagging for eastern and western Atlantic bluefin;
7. Potential risk and strategies for minimizing them;

12.2 Appendix 2: Sampling strata and sample size for 2015 GBYP biological sampling program.

Table 10: Sampling strata and sample size for 2015 GBYP biological sampling program.

GBYP BIOLOGICAL SAMPLING 2015					
		Otolith	Spine	Muscle/Fin	Sampler
Eastern Mediterranean	Levantine Sea	62	71	71	AZTI (Oray)
Central Mediterranean	Adriatic Sea	50	50	50	UNIBO
	Malta	10		10	FMAP
	South Sicily, Strait of Sicily	50	50	50	UNIBO
	East Sicily and Ionian Sea		100	100	Necton
	Gulf of Syrta	27	27	27	FMAP
Western Mediterranean	Tyrrhenian Sea	44	124	124	Necton
	Balearic	38	38	118	IEO
	Gibraltar	15	15	15	
	Ligurian Sea	25	25	25	UNIGE
	Sardinia	43	78	78	UNICA
Northeast Atlantic	Bay of Biscay	4		4	AZTI
	Madeira, Canarias	23		23	IEO
	Morocco	50		50	INRH
	Norway		1	24	IMR
	Portugal	40	36	44	IPMA
Central North Atlantic	Central North Atlantic	402		408	NRIFSF
Northwest Atlantic	Gulf of Mexico			182	NOAA/AZTI
	Gulf of Saint Lawrance	30			DFO
TOTAL		**913**	**615**	**1403**	
			2931		