THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BFT (GBYP Phase 12)

SHORT-TERM CONTRACT (ICCAT GBYP 02/2023)

PILOT STUDY ON EPIGENETIC AGEING TECHNIQUE FOR AGE ESTIMATION OF ATLANTIC BLUEFIN TUNA – GBYP 2/2023.

Final Report

July 2023

THE COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION (CSIRO)

Campbell Davies, Ben Mayne, Peter Grewe, and Jessica Farley

INSTITUTO ESPAÑOL DE OCENOGRAFIA

Enrique Rodriguez-Marin

This project is co-funded by the European Union

PILOT STUDY ON EPIGENETIC AGEING TECHNIQUE FOR AGE ESTIMATION OF ATLANTIC BLUEFIN TUNA

SUMMARY

The ICCAT GBYP program is investigating the feasibility of applying Close-kin Mark Recapture (CKMR) to the eastern stock of Atlantic bluefin tuna. A key consideration in this investigation is the ability to obtain accurate age information for use in the construction of the kin probabilities. This paper reports on progress of a pilot study to assess the suitability of epigenetic ageing - estimating the age of individuals from the level of DNA methylation from analysis of tissue samples - for this purpose. The aims of the pilot study were to: i) calibrate an epigenetic age model using the approach developed by Mayne et al 2021 and tissue samples from eastern and western stocks with age data from standardized ICCAT otolith reading methods, ii) examine the influence of stock and sex on epigenetic age calibration, and iii) assess the relative cost-effectiveness with the use of otolith-based age estimation for use in CKMR for eastern BFT. This report summarises the study design and activities completed within GBYP Phase 12, as well as the schedule for completion of outstanding activities that were not completed before the close of Phase 12 (23 July 2023). The aim is to complete the outstanding activities in time to report the results and conclusions of the full pilot study to the Bluefin Species Group meeting at the SCRS in September 2023.

KEYWORDS

Epigenetic ageing, calibration Close-kin mark–recapture (CKMR), Atlantic Bluefin tuna, genetic methodologies, stock structure, genetic sex determination.

1. Introduction

The ICCAT GBYP program is investigating the feasibility of applying Close-kin Mark Recapture (CKMR) to the eastern stock of Atlantic bluefin tuna (BFT) (Anon., 2023). A key consideration in this investigation is the ability to obtain accurate age information for use in the construction of the kin probabilities (Anon., 2023). This paper reports on progress to date of a pilot study to assess the suitability of epigenetic ageing - estimating the age of individuals by analysing the degree of DNA methylation in their tissue - for this purpose.

Development of an epigenetic clock for BFT, will follow the approach used on other fish species (Mayne et al 2021 and 2022). In short, an epigenetic clock works by measuring DNA methylation at multiple cytosine-phosphate-guanine (CpG) sites and uses machine learning methods to translate DNA methylation to age. We have previously identified 1,311 CpG sites, known to be age associated in zebrafish (*Danio rerio*), from a total of several hundred thousand markers from genome-wide scans (Mayne et al., 2020). These sites are conserved over a wide range of teleost taxa and have been used to successfully calibrate epigenetic age models for a range of teleost species, including tuna (Davies et al, In prep; Mayne et al, in Anon.2023). Age associated sites from this set that are conserved in BFT genome will be targeted for amplification by multiplex PCR. The DNA is then sequenced using an Illumina MiSeq sequencer with high coverage to measure the degree of methylation at each site. This data is used in combination with the "known age" data, estimated from otolith readings, to construct a statistically calibrated epigenetic age model that can be used to independently predict age directly from the level of DNA methylation.

The aims of the pilot study are to: i) calibrate an epigenetic age model, using the approach developed by Mayne et al 2021, and tissue samples from eastern and western stocks and age data from standardized ICCAT otolith reading methods; ii) examine the influence of stock and sex on epigenetic age calibration; and iii) assess the relative cost-effectiveness with the use of otolith-based age estimation for use in CKMR for eastern BFT. In addition to the use in CKMR, if successful, epigenetic age would also be valuable for providing catch-at age data, which would be a significant improvement on current methods used in the regular stock assessment.

This report summarises the study design and activities completed within GBYP Phase 12, as well as the schedule for completion of outstanding activities that were not able to be completed before the close of Phase 12 (23 July 2023). The aim is to complete the outstanding activities in time to report the results and conclusions of the full pilot study to the Bluefin Species Group meeting at the SCRS meeting in September 2023.

2. Methods

Source of samples

Tissue samples and age data for the eastern stock were provided by AZTI from the GBYP tissue bank, while samples and age data for the western stock were provided through the existing NOAA-CSIRO close-kin project for western BFT (Grewe et al, 2023 and Lauretta et al 2023 in CKMR workshop report) via the sampling conducted by Department of Fisheries and Oceans, Canada, the University of Maine and NOAA South-east Fisheries Centre, along the east coast of the USA. Table 1 provides a summary of the samples available to this pilot study by sampling region and age class.

The western samples were available as: i) previously extracted DNA stored in archive plates at -80 degrees celcius, which was used for the sex determination and epigenetic age analyses, or as existing DArTseq files, in the case of the provenance determination.

The same samples from each sampling source have been used for provenance and sex determination and the epigenetic age calibration.

DNA extraction and QC

DNA extractions were prepared from approximately 15mg of tissue subsampled from individual biopsies. Samples were extracted on an Eppedorf EP motion 5057 liquid robotic handler using a modification of the QIAamp® 96 DNA QIAcube HT Kit (QIAGEN, Hilden, Germany). This extraction includes a lysis step in the presence of Proteinase K followed by bind-wash-elute QIAGEN technology. Low quality/degraded samples were re-extracted using the modified CTAB method following Grewe et al. (1993).

Provenance determination

DArT-CAP sequence data was examined to confirm individual provenance by examination of stock specific markers obtained from training set data sampled from Gulf of Mexico and Mediterranean larvae (Grewe et al., 2021). Only samples with high degree of confidence for stock of origin were chosen to represent Western and Eastern Stocks used for stock specific age analysis. *(Best to get Shane to write some more detail here if required)*

For provenance analysis, DNA aliquots of the 250 eastern samples were shipped to Diversity Array Technologies (DArT) in Canberra for DNA complexity reduction and library construction prior to sequencing to generate genotype data for each individual. Identical markers and protocols to those used to generate the existing genotype data for the western samples were used for the eastern samples.

Individual DArT-SEQ library preps were sequenced on an Illumina platform that produced raw FASTQ sequence data files for each individual. The DArT-Soft14 (DS14) was used to generate multi-locus SNP genotype profiles from the raw FASTQ files. Individual genotypes were QC filtered using two separate pipelines, Kinference (Baylis et al, In prep) and RADIATOR (Gosselin, 2020) to delete poor quality (e.g., paralogous) loci and individuals exhibiting overall poor-quality DNA profiles (*e.g.* DNA cross-contamination and sequencing dropout). Data-cleaning and quality control workflows were completed prior to provenance analysis (see Davies et al., 2020, Grewe et al., 2021). DArT-CAP sequence data was examined to confirm individual provenance by examination of stock specific markers obtained from training set data sampled from Gulf of Mexico and Mediterranean larvae (Grewe et al., 2021). Only samples with high degree of confidence for stock of origin were chosen to represent Western and Eastern Stocks used for stock specific age analysis.

Sex determination

Genetic sex determination was completed using a set of 22 fluorescently labelled primer pairs in an 11-plex PCR assay that targets three male specific amplicons plus 8 autosomal DNA microsatellite loci. The assay produces sex specific fragments that are separated and visualised on an ABI 3730 DNA sequencer. DNA microsatellite loci provide confirmation of a positive PCR resulting from good quality DNA and are also used to assess the degree of DNA cross contamination for each individual tissue sample. Files for each individual are then genotyped using GeneMapper 4.0 software package (Applied Biosystems Inc.) following protocols developed by CSIRO (Grewe et al, unpublished).

Epigenetic age calibration and prediction

Age-associated CpG sites were identified through genome pairwise alignment with known age-associated sites in zebrafish. Briefly, by targeting conserved and age-associated sites from another species, this reduces the expense in biomarker development. This process has been carried out previously in other species, including fish (Polanowski, A.M., et al.,2014, Mayne et al., 2021). CpG sites that were both known to be age-associated in zebrafish and conserved between both species were targeted for primer design in the multiplex PCR assay specifically for BFT.

Genomic DNA was extracted using the QIAamp 96 DNA QIAcube HT Kit (QIAGEN) as instructed by the manufacturer's protocol. DNA was bisulfite treated using a modified version of a previous protocol (Lu, A., et al., 2021). Multiplex PCR was designed for one pool of primers using PrimerSuite (Li, H., et al.2019). Each amplicon was tested in singleplex with annealing temperatures between 55-60°C. Primer pairs that produced multiple or no bands were excluded from the multiplex PCR reaction. Barcoding was carried out using the Fluidigm 384 set of barcodes (Cat. 100-4876). Barcode reactions were pooled together in equal volumes and were sequenced using an Illumina MiSeq Reagent Kit v2 with custom sequencing primers (300 cycle; PN MS-102-2002) at the Australian Genome Research Facility (AGRF).

Sequencing data was aligned to a representative genome for each species based on the amplicon sequences. DNA methylation for each CpG site was called as a percentage, where 0% means the site is unmethylated and 100% the site is fully methylated. Samples that had no sequencing coverage for any amplicon were removed from the analysis. Samples were randomly assigned in either a training or testing data set with a 70/30 split respectively. An elastic net regression model and new approach developed by CSIRO (Lloyd-Jones et al., In

prep) that better accounts for the characteristics of the data will be used to generate calibration models using all targeted CpG sites (Engebretsen, S. and J. Bohlin, 2019). The performance of the model will be measured using Pearson correlations, median absolute and relative errors.

3. Results

Sample characteristics

There are more than 250 tissue samples available from both sources with otolith ages ranging from 2-27 years (Table 1). There were 15 year-classes in the eastern stock samples (2-16 years) with 17-20 samples per each age class between 2 and 15, providing good replication across this age range. There was a 50:50 sex ratio for the selected eastern samples based on the phenotypic sex provided with the samples.

The western samples ranged from 4-27 years, with 9-33 samples per year class between ages 6-17 and fewer than 2 per age class above 17 years old. Again this provides for sufficient replication out to 17 years, but low or no replication in the older year classes. Phenotypic sex was available for only a small subset of the western samples. Hence, it was not possible, *a priori,* to balance the selection the subset of samples for the western stock based on sex and, therefore, examination of the effect of sex on the epigenetic age estimation for the western samples will rely on the success of the genetic sex determination (see below).

The age distribution of the available samples should provide for a high-quality calibration for the 2-16 for the combined data set. The DNA extracted from the eastern samples was of high quality, with 246 (89%) of the 277 eastern samples passing the QC requirements for each of the downstream genetic analysis. Those that did not pass initial QC have been re-extracted and included in the subsequent processing.

Provenance determination

Two-hundred and fifty eastern samples have been submitted to Diversity Array Technologies for DArT-Cap sequencing. The schedule of delivery of the data and analyses are provided in Table 2.

The existing DArT-Cap data for the western stock has been extracted and quality controlled and is ready for analysis as soon as the data for the eastern samples is available.

Sex determination

The PCR step of the sex determination workflow has been completed for 500 samples (250 eastern and 250 western). These samples have been submitted to the AGRF for sequencing and approximated 75% have been sequenced. The schedule for completion of the sequencing and availability of the data is provided in table 2.

Epigenetic age calibration and prediction

DNA extractions for the epigenetic age calibration is complete and the samples are currently undergoing the multiplex PCR stage prior to sequencing. Approximately 25% of the multiplex library preparation is complete.

The schedule for completion of the sequencing, bioinformatics and calibration of the epigenetic age model is provided in Table 2.

4. Discussion

This project aims to provide a demonstration of the feasibility of applying epigenetic ageing to BFT to facilitate the application of CKMR to this species. If successful, which we expect it will be given the high quality of the DNA extracted from the tissue samples and the level of replication for the main age range that would be covered in a potential CKMR study, then this method would not only facilitate the implementation of CKMR, but also its use for routine catch at age data for conventional stock assessment.

Unfortunately, substantial delays in the installation and programming of a replacement genetics robot in the CSIRO lab meant that the sequencing aspects of the genetics work-flows for each component (provenance, sex and epigenetic age) of the project have not been able to be completed within the timeframe of Phase 12 of the GBYP. Notwithstanding this delay, we are confident that the remaining activities will be completed in the coming month and in time for a paper on the results and conclusions to be submitted and presented to the BFT Species Group of SCRS. Given our experience with a range of other species, including the results for three species of tuna presented at the ICCAT CKMR workshop in March (Anon., 2023), we are confident that the results should bode well for the application of the method to BFT.

5. Acknowledgements

We thank Francisco Alemany and Igaratza Fraile for their cooperation in arranging transport of the tissue samples and associated data from the GBYP Tissue Bank and Matt Loretta, Alex Hanke and Walt Golet for the use of the western age data. Nicola Potter and Chloe Anderson provided excellent technical assistance with the genetics and Shane Bayliss assisted with the retrieval of the DArT-Cap genotype and meta-data for the western samples.
This work has been carried out under the ICCAT Atlantic-Wide Research Programme for BFT (GBYP), which is funded by the European Union, several ICCAT CPCs, the ICCAT Secretariat, and other entities (see https://www.iccat.int/gbyp/en/overview.asp). The content of this paper does not necessarily reflect ICCAT's point of view or that of any of the other sponsors, who carry no responsibility. In addition, it does not indicate the Commission's future policy in this area.

6. References

Anon., 2023. Report of the workshop on Bluefin Tuna close-kin mark recapture (CKMR), including biological sampling coordination. Madrid, Spain, hybrid, 1-6 March, 2023.

Davies C, Marsac F, Murua H, Fraile I, Fahmi Z,et al. 2020. Population Structure of IOTC species and sharks of interest in the Indian Ocean: Estimation with Next Generation Sequencing Technologies and Otolith Micro-chemistry. Final project Report to IOTC. 249p.

Engebretsen, S. and J. Bohlin, *Statistical predictions with glmnet.* Clinical Epigenetics, 2019. **11**(1): p. 123.

Grewe, P.M., Baylis, S., Bravington, M.V., Aulich, J., Devloo-Delva, F., Reglero, P., Alvarez-Berastegui, D., Martin, M., Gosselin, T., McDowell, J. Lauretta, M., Walter, J. and Davies, C.R. 2021. Genomic assessment of Atlantic bluefin tuna larvae collected from the Mediterranean Sea: Refinement of SNP markers for stock of origin assignment (Western, Gulf of Mexico versus Eastern, Mediterranean Sea) of adult fish used for genetic Close-Kin Mark-Recapture. Report to the Pew Charitable Trust. May 2021, 12p.

Li, H., et al., *Cost-reduction strategies in massive genomics experiments.* Marine Life Science & Technology, 2019. **1**(1): p. 15-21

Lu, A., et al., *Universal DNA methylation age across mammalian tissues.* bioRxiv, 2021: p. 2021.01.18.426733.

Mayne, B., et al., 2021. *Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod.* Molecular Ecology Resources, 2021. **21**(7): p. 2324-2332.

Mayne, B., et al., 2022. *Age prediction of green turtles with an epigenetic clock.* Molecular Ecology Resources, 2022. **00**: p. 1– 10

McDowell, J.R., Bravington, M., Grewe, P.M.. et al., 2022. Low levels of sibship encourage use of larvae in western Atlantic bluefin tuna abundance estimation by close-kin mark-recapture (2022). *Scientific Reports*, 12(1), 18606.DOI: 10.1038/s41598-022-20862-9.

Polanowski, A.M., et al., Epigenetic estimation of age in humpback whales. Molecular Ecology Resources, 2014. 14(5): p. 976-987.

TABLES

Table 1: Summary of source, sampling region and age of tissue samples/DNA available for eastern and western stocks of BFT for pilot study on the suitability of epigenetic ageing. GBYP = Grand Bluefin Year Program administered by ICCAT. MED = Mediterranean, NALT = North Atlantic, SALT = South Atlantic, WALT = Western Atlantic. NOAA/UM/DFO – collaborative sampling program between US-NOAA, University of Maine and Department of Fisheries and Oceans-Canada. CAN= Canada, MAI = Maine.

		Age class (years)																								
Source	Region	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**26**	**27**	**Total**
GBYP	MED	17	15	19	18	14	17	16	15	17	11	12	15	17	5	1	-	-	-	-	-	-	-	-	-	209
(Eastern)	NATL	0	0	0	0	1	0	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
	SATL	0	1	1	2	3	4	2	3	-	-	6	3	-	-	-	-	-	-	-	-	-	-	-	-	33
	WATL	-	-	-	-	-	-	1	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	4
	Total	17	16	20	20	18	21	20	19	19	20	18	18	18	5	1	0	0	0	0	0	0	0	0	0	**250**
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**26**	**27**	**Total**
NOAA/UM/DFO	CAN	-	-	1	-	-	8	24	24	26	15	31	33	21	14	8	9	2	2	1	2	2	-	-	1	231
(Western)	MAI	-	-	-	-	3	4	6	1	-	1	2	-	1	-	1	-	-	-	-	-	-	-	-	-	19
	Total	0	0	1	0	9	12	30	25	26	16	33	33	22	14	9	9	2	2	1	2	2	0	1	1	**250**

Table 2: ICCAT ABT epigenetic age pilot study: work plan and revised schedule as of 21 July 2023.

Project Component	Activity	Status	Revised completion date	Comment
Sample acquisition	Select and acquire GBYP samples	Complete	-	See: SelecSamples_Epig_Age_BFTE_20230120.xlsxList of samples selected from GBYP collection.
	DNA extraction and QC GBYP tissue samples	Complete	-	CSIRO Hobart genetics laboratory See: ABT_Concentrations_11July23.xlsx. This is a summary of the DNA QC following extraction used to assess their suitability for epigenetic age and other genetic analyses.
	Select western samples and locate archived DNA	Complete		See: Sum_ABT-Wprov_BYage_epigenetics_19July23. DNA will be sent to BM next week for epigenetic age. Sub-sample prepared for genetic sex determination.
Population assignment	DArT-Cap at DArT 250 eastern samples	In progress	16 August 2023	Invoiced for payment
	Retrieve western sequence data for analysis	Complete		
	Provenance analysis for eastern and western samples		23 August 2023	Or as soon as DArT data for eastern samples is received and QC'd
Genetic sex determination	PCR for sex determination	Complete		CSIRO Hobart genetics laboratory
	Sequencing at AGRF for sex determination	In progress – 50% complete	4 August 2023	Invoiced for payment
Epigenetic age	Multiplex PCR for library prep for sequencing	In progress	5 August 2023	CSIRO Perth genetics laboratory
	Sequencing at AGRF for methylation		30 August 2023	Invoiced for payment to secure sequencing runs.
	Bioinformatics for methylation scores		1 September 2023	This is a standard step with established HPC routines
	Collation & QC of final dataset for epigenetic age		1 September 2023	
	Calibration and prediction analysis		6 September 2023	This will be done with the refined statistical methods presented at the BFT CKMR workshop.
Reporting	Draft final report	Submitted	22 July 2023	Will be revised based on ICCAT review.
	Final report		8 September 2023	
	SCRS paper submitted		12 September 2023	
	SCRS Presentation to BFT Species Group		18-25 September 2023	

Appendix 1: Abridged Terms of Reference for ICCAT GBYP 02/2023 – Pilot study on epigenetic ageing technique for age estimation for Atlantic bluefin tuna Atlantic-wide research program for bluefin tuna (ICCAT GBYP – Phase 12).

The contractor will perform a pilot study on epigenetic ageing technique (DNA methylation-based analysis) for age estimation of Atlantic bluefin tuna, according with the following terms of reference:

1. The aim of the pilot study will be to evaluate the accuracy of the epigenetic clock in comparison with direct age readings. Precision vs costs must also be considered in this comparison.
2. At least 500 samples of Atlantic bluefin tuna specimens (250 from each management areas, West vs East Atlantic plus Mediterranean Sea) will be analyzed. In these 250 samples for each management area, the entire age range should be represented, with at least approximately 10 specimens by age class. Samples used for testing the epigenetic clock should include individuals of both sexes. Specimens captured in West or East Atlantic management areas should cover all possible geographical locations. Most of the samples caught in the East area should come from the Mediterranean.
3. The genetic origin of each sample must be accredited for those samples for which genetic origin has not been previously assigned.
4. Muscle samples from specimens will be provided to the analytical team from the GBYP and other laboratories tissue banks. Sample metadata to include sex, location and date of sampling, gear and sampling lab will be provided to the analytical team. Direct ageing by means of schlerochronological methods will also be provided based on standard and validated protocols.
5. In order to reduce costs and meet the deadlines for the pilot study, samples of specimens with genetic origin already assigned and/or whose otoliths have been previously read (with high reading confidence) should be favoured.

Clarification on "accuracy of epigenetic clock" in ToRs

The application of epigenetic ageing to a particular species/taxon involves a number of steps, including the identification of appropriate epigenetic sites (CpG sites) and optimisation of the multiplex PCR for sequencing; (see more detailed explanation in 2 Methods, below); followed by calibration of the analysis model with samples from "known age" animals (see more detailed explanation in 2 Methods, below). In the case of tuna, the "known age" samples are sourced from individuals with high confidence, otolith-based, age estimates. As a result, the accuracy of the calibrated epigenetic age estimates are influenced, to a degree, by the accuracy of the otolith-based age estimates used in the calibration, and by the age signal from the degree of methylation at multiple CpG sites. Given the lack of true "known age" fish (e.g., animals sampled from breeding programs) that could be used as an independent source of "true age" for BFT, comparing the accuracy of the epigenetic age independently of the otolith ages is not possible. We can, however, estimate the precision and the level of uncertainty associated with each source as part of the development of the calibration model. This is how we propose to assess the performance of the epigenetic age model.