**Overview**

The State Fisheries Genomics Lab at Oregon State University and GTseek LLC developed a Pacific Albacore Tuna GTseq Panel in 2021. Genomic features (SNPs) were selected from three sources: FST outliers between North and South Pacific samples identified by (Vaux *et al.* 2021), spatial redundancy analysis, and presumed neutral loci. After feature selection, primers were designed and two rounds of panel optimization were completed to produce the final panel. A previously developed PCR based sex assay was adapted to function in the GTseq panel and was evaluated using sex identified individuals. The final panel consisted of 288 SNP markers and the sex marker.

**Feature Selection**

*Vaux 2021 Dataset*

Vaux et al (2021) successfully genotyped 308 albacore at ~13k RADseq loci. The final dataset included 234 individuals from 10 locations throughout the North Pacific and 74 individuals from two locations in the South Pacific. This dataset and the underlying raw RADtag sequences were used for feature selection

*FST Markers*

Vaux et al (2021) identified 84 SNPs that were found in common across multiple FST outlier scans comparing North and South Pacific samples. All 84 SNPs were selected as genomic features for primer design.

The intended function of these SNPs in the panel is to discriminate between albacore with North and South Pacific ancestry.

*Spatial Redundancy Analysis*

Write methods results summary.

The intended purpose of these spatial markers in the panel is twofold. First these markers may provide additional power to discriminate between albacore with North and South Pacific ancestry. Second, these markers may reveal genetic structure at finer spatial scales, for example within the North Pacific.

*Neutral Markers*

Neutral SNPs are based on the Vaux et al 2021 sequence data but used a different filtering approach. The empirical estimated site frequency spectrum (SFS) indicated that the majority of the variants in among samples was substantially below the minor allele frequency cutoff used to generate the SNP dataset for the spatial and FST outlier analyses. To avoid biasing the SFS for the neutral SNPs, a new SNP dataset was called using less stringent filtering. The neutral SNPs were then drawn from this dataset.

The intended purpose of the neutral SNPs in the panel to provide a more suitable set of markers for estimating population genetic parameters that are intended to reflect genome-wide, neutral patterns.

**Primer Design**

A total of 495 genomic features were selected for primer design. RADtag sequences from Vaux et al (2021) containing the selected features were provided for GTseq primer design. Using these data as well as flanking sequence from the 2013 Pacific Bluefin assembly (RefSeq: GCA\_000418415.1). After several rounds of optimization, primer sets meeting the physical properties for GTseq and filtered for expected performance in multiplex PCR were designed (Nmarkers = 332). Primers for 15 neutral markers with lengths greater than 60bp were excluded to reduce primer costs (Nmarkers = 317).

**Validation 1**

An initial test library (validation library 1) was sequenced on an Illumina NextSeq500 instrument with 75bp single end chemistry. Validation library 1 used a primer pool of 317 primer pairs (but note that GTseek LLC results for this panel included the 15 neutral markers with long primers, so all results use Nmarkers = 332), and 148 unique albacore, 138 of which were previously genotyped in Vaux et al (2021).

Validation library 1 returned an average on-target rate of 31% and 28 primer pairs were identified that contributed to most of the off-target sequences.

**Validation 2**

The 28 primer pairs producing primer artifacts were omitted from a new primer pool and validation library 2 was prepared using the remaining 289 primer pairs. Validation library 2 was sequenced on a Miseq3 instrument using 75bp paired end sequencing. The overall on-target rate for the validation library was 61% among a subset of 100 individuals that performed well. Further analysis of the dataset identified a few other primers contributing to the off-target signal (it doesn’t look like these were ever removed???)

**Final Panel**

The final panel includes 289 markers (Table 5). Full panel information is available at the [SFGL GTseq Github Repository](https://github.com/State-Fisheries-Genomics-Lab/GT-seq).

|  |  |
| --- | --- |
| **Marker Type** | **n** |
| FST Outlier | 54 |
| Spatial Outlier | 117\* (81) |
| Neutral | 153 |
| Total | 289 |

**Table 5:** Marker Counts in the final panel (Panel\_289). \*Note that there are 117 spatial outliers, but 36 are also FST outliers

Vaux F, Bohn S, Hyde JR, O'Malley KG (2021) Adaptive markers distinguish north and south pacific albacore amid low population differentiation. *Evolutionary Applications* **14**, 1343-1364.