### **METHODS**

### **Sample Collection and DNA extraction**

Samples were collected between the year 2015 and 2016 by Mr. Najih Lazar, the fisheries advisor on the USAID-Sustainable Fisheries Management Project (SFMP) in Ghana. A total of fifty (50) specimens of round sardinella were obtained from landing sites in selected countries (Table1). The dorsal fin from the base was cut and placed in a 1.5-2 ml polypropylene vials/tube. The fin tissue was then preserved in 95% ethanol and stored in -200C until it is ready to be shipped. Genomic DNA from fin clips were extracted using Qiagen DNeasy kit (Qiagen, Valencia, CA) following the manufacturer’s protocol. Genomic DNA quality and concentration were checked using Nanodrop 2000 and Qubit 2.0 fluorometer instruments.

**Table 1.**Details on sampling for RAD sequencing from seven(7) sampling sites. The numbers of samples (*n*)

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Region coordinates | Collection date | Number of samples |
| Mauritania | 3.11 | August 2016 | 50 |
| Senegal | 3.12 | December 2016 | 50 |
| Guinea | 3.3 | January 2017 | 50 |
| Ghana | 3.4 | August 2017 | 50 |
| Togo | 3.4 | November 2016 | 50 |
| Benin | 3.4 | - | 50 |
| Angola | 3.6 | September 2016 | 50 |

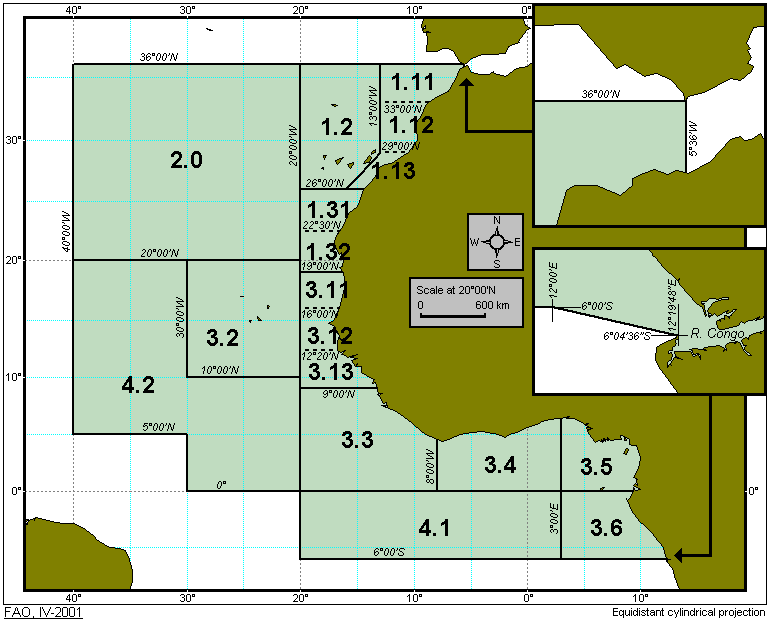


Figure 2: Map of areas sampled. Areas sampled are from subdivision point 1.32 to 3.6.

**ddRADSeq Library preparation and sequencing**

Libraries for double digest RAD‐seq(ddRAD-seq) were prepared using the enzymes Tas1 and MspI and sequencing was carried out using Illumina Hiseq4000 at Texas Corpus Christi sequencing core facility.

### **SNP discovery and genotyping**

The *dDocent* pipeline version 2.5.3 (Puritz, Hollenbeck, & Gold, 2014) was used for quality filtering of demultiplexed data, de novo assembly, read trimming, read mapping, and variant calling. For denovo assembly, assembly parameters of k1=2, k2=2, and clustering similarity equalled 90% were used after examining multiple iterations, for read mapping, the default parameters were used, with mismatch parameter equals 4 and the gap opening penalty of 6 . Variant calling was carried out using freebayers which was subjected to several filtering steps to reduce false positives.

### **SNP filtering**

Raw variant calls were filtered using the program VCFtools and several bash scripts (pop\_missing\_filter.sh, dDocent\_filters.sh,rad\_haplotyper.pl, remove.bad.hap.loci.sh) (Puritz, Hollenbeck, & Gold, 2014). Loci with minimum depth less than 5 were filtered out. SNP loci that were not called in 25% individuals in each of the population were removed using the script pop\_missing\_filter.sh. dDocent\_filters were used to filter loci based on FreeBayes INFO criteria and depth. Variant calls were collapsed into SNP and Indel calls, Indels were removed with VCFtools to produce a VCF file of only SNP calls. SNPs were then filtered to only include loci with 2 alleles using VCFtools and haplotyped using the script rad\_haplotyper.pl to search for loci that are paralogs. Loci that had more than 5 individuals marked as paralogous and more than 30 individuals marked for potential genotyping errors were removed using the script remove.bad.hap.loci.sh . loci were finally haplotyped and converted into a final GENEPOP file for analysis.