Statement of Work

Chromosome Conformation and Whole Genome Sequencing

Background

The USDA ARS NCWMAC Shellfish Genetics program is conducting research on eastern oysters in order to assess genetic diversity and develop improved germplasm.

Objectives

The objective of this work is to use chromosome conformation capture techniques to identify DNA sequences that are located on the same chromosome over long distances and use whole genome sequencing to jointly identify novel genetic variants in populations of eastern oysters.

Scope

The project requires DNA extraction, library preparation, and paired-end sequencing of two samples for chromosome conformation capture and 21 samples for whole genome sequencing (WGS). All libraries will be sequenced with 2 x 150 bp reads and target 100 million paired reads per sample (200 million reads total) for a coverage of 50x of a 600 Mbp genome. Sequence data will be demultiplexed.

Requirements

Chromosome conformation library preparation and sequencing

USDA-ARS will submit two flash frozen tissue samples to the contractor. The contractor will extract DNA, construct chromosome conformation capture libraries using a proximity ligation protocol, and sequence these separately to generate a minimum of 100M 150bp paired-end reads (200M reads total, 50x coverage of a 600 Mbp genome) per sample. The technique must use a non-specific endonuclease or equivalent method to give reads that are not meaningfully biased toward a specific sequence (such as a restriction enzyme cutsite), for example the OmniC kit or equivalent. The technique must have been demonstrated in the scientific literature to yield data of sufficient quality for investigating structural variation and assembling chromosome-scale genomes. Over 50% of the cis reads for each sample must be long-range cis pairs as assessed by mapping to the current NCBI RefSeq eastern oyster (*Crassostrea virginica*) genome assembly. Reads for each sample will have a minimum of 70% of bases with quality score above Q30. Sequence data will be collected using a technique that produces minimal insertion/deletion errors, such as the Illumina platform or equivalent. The contractor will deliver the demultiplexed sequence data as FASTQ files to USDA ARS.

Whole genomes library preparation and sequencing

USDA-ARS will submit 21 tissue samples preserved in ethanol or similar preservatives to the contractor. These samples will be delivered in up to two shipments and will be delivered separately from the chromosome conformation samples described above. The contractor will extract DNA then construct and QC whole genome sequencing libraries (one per individual) and sequence these to generate a

minimum of 100M 150bp paired-end reads (200M reads total, 50x coverage of a 600 Mbp genome) per sample. The sequence data will then be demultiplexed to represent individual samples. Reads for each sample will have a minimum of 70% of bases with quality score above Q30. Sequence data will be collected using a technique that produces minimal insertion/deletion errors, such as the Illumina platform or equivalent. The contractor will deliver the demultiplexed sequence data as FASTQ files to USDA ARS.

Deliverables

The contract will be considered complete upon receipt of demultiplexed sequence data meeting the technical requirements for each sample as well as a brief description of the library preparation protocol (e.g., names of kits and adaptors used) and sequencing instrument (e.g., model name and version) used within 60 days of sample submission.

Reporting Requirements

The sequencing data should be sent to Thomas A. Delomas, USDA ARS NCWMAC Shellfish Genetics, 483 CBLS, 120 Flagg Road, Kingston, RI 02881, or posted on a private server for download.

Other Considerations

Vendor must demonstrate previous experience successfully applying the technique proposed for chromosome conformation library preparation and sequencing (see requirements above) to one or more marine invertebrate species.