

Name of the service: Service for three bark beetle genome sequencing

Required activities - their description, scope (e.g., the number of samples, forest area), time (requirements) etc.:

1. Pre-made Hi-C or Omni-C libraries sequencing Illumina or Novaseq PE150 Q30 \geq 85% (100x coverage)

Services: Library quality control (QC) (3 samples), Sequencing 40Gb/library (3 samples), data delivery via cloud or hard drive transfer.

2. Whole-genome Sequencing on Novaseq or the latest Illumina platform, PE150 Q30 \geq 85%. Illumina short insert DNA library 100x coverage.

Services: gDNA library prep and sequencing 40Gb/sample raw data in fast (3 samples), Data delivery via cloud or hard drive transfer.

3. Illumina short-read transcriptome sequencing (48 samples) for gene model Prediction. mRNA sequencing (Novaseq or Illumina PE150 Q30 \geq 85%)

Services: cDNA library preparation (poly-A selection) and sequencing 50M PE reads=15 Gb/sample, Data delivery via cloud or hard drive transfer.

4. Data analysis service (+ customized): Complete genome data analysis (genome-level assembly, annotation, comparative, and evolutionary analysis) – 3 genomes. The contractor will provide pacbio sequences of genomes for incorporating into the analysis.

Services: Genome level assembly service; Annotation service; Comparative and Evolutionary Analysis. Analysis should also include the following:

- a) Synteny analysis for 2 genomes, [*3x]
- b) Genome element Circos (for 3 genomes)
- c) Gain and loss analysis for 3-5 genomes – [*1]

Altogether in 4 shipments, (price for 4 shipments is included in the unit price stated in the Attachment No. 2 of the contract – Tender price , all other shipments will be charged)

The Client also requires the following to be included in the data analysis:

1. Survey Analysis Content

Data quality control: filtering reads containing adapter or with low quality

Genome size estimation

Heterozygous rate & Repeat content estimation

Contamination estimation

2. Pipeline for assembly (genome level) and annotation.

Assembly Analysis Content

Data quality control

Sequencing data error correction

Contig assembly

Scaffold assembly

Gap filling

Genome assembly results statistics

Assembly quality control

Completeness analysis of the genome.

3. Annotation Contents

Repetitive sequence annotation

Structural annotation

Gene function annotation

Non-coding RNA annotation

4. Comparative and Evolutionary Analysis

Gene family clustering

Phylogenetic tree construction (including estimation of divergence time)

Gene family expansion and contraction

Positive selection