1. Bioinformatics Analysis

|  |
| --- |
| **Data Quality Control** |
| Distribution of Sequencing Error Rate |
| Distribution of A/T/G/C Base Content |
| Data Filtering: Removal of reads with adaptor contamination, uncertain nucleotides(N)with > 10% content and low quality nucleotides |
| Statistic Summary of Data Quality: Amount of data output, Error rate, Q20%, Q30% and GC content, etc |

|  |  |
| --- | --- |
| **Standard Analysis (Prokaryotic Species with Reference)** | |
| Data Quality Control: Filtering reads containing adapter or uncertain nucleotides or of low quality, and statistic summary of data quality | |
| Mapping Clean Reads to Reference Genome | |
| Novel Gene Prediction | |
| Gene Expression Quantification | |
| Correlation Analysis Between Groups (**Only for Samples with biological replicates**) | |
| Differential Expression Analysis (**Only for Grouping with control samples**) | |
| Functional Analysis of **Differentially Expressed Genes** (DEGs) (**Only for Grouping with control samples**) | Gene Ontology (**GO**) Enrichment Analysis |
| Kyoto Encyclopedia of Genes and Genomes (**KEGG**) Pathway Enrichment Analysis |
| Protein Protein Interaction Analysis (**Only for species**  **with well-known PPI network database**) |
| Gene Structural Analysis | Operon Prediction |
| Transcription Start Site (**TSS**) and Transcription Termination Site (**TSS**) Prediction |
| Promoter prediction |
| UTR Analysis | UTR Prediction and Length Distribution |
| 5’ SD Sequence Prediction |
| 3’ Rho-independent Terminator Prediction |
| Antisense Transcripts Prediction | |
| Small RNA Analysis | sRNA Prediction and Length Distribution |
| sRNA Secondary Structure Prediction |
| sRNA Targeted Gene Prediction |
| sRNA Expression Quantification |