

Project Log – Transposable Element (TE) Analysis

Project: GENE DUPLICATION AND TRANSPOSABLE ELEMENT DYNAMICS OF *Oryza sativa* AND *Oryza brachyantha* (GROUP 3)

Period: November 14 – January 7

Contributor: Priyasha PAUL (20245619)

November 14

- Defined the scope of the transposable element (TE) analysis within the broader comparative genomics project.
- Reviewed project instructions to identify required analyses related to TE content, composition, and spatial distribution.
- Identified *Oryza sativa* and *Oryza brachyantha* as focal species for comparative TE analysis.
- Downloaded TE annotation files (GFF3 format) for both species from APTEdb.
- Downloaded corresponding genome annotation files (GFF3) from Ensembl Plants.
- Performed an initial inspection of annotation formats and feature types to ensure compatibility.

November 16

- Imported TE and gene annotation files into R using the rtracklayer package.
- Verified feature types present in gene annotations (gene, mRNA, CDS, UTRs).
- Verified TE annotation categories and classification levels (Class, Order, Family).
- Filtered gene annotations to retain protein-coding gene features.
- Cleaned TE annotations and harmonized chromosome naming conventions.
- Converted annotations into GRanges objects for downstream genomic analyses.
- Computed total TE copy numbers for each species.
- Calculated total TE genomic contribution (base pairs).
- Performed initial comparison of TE abundance between species.

November 19

- Classified TEs by major orders (LTR, LINE, SINE, TIR, MITE, Helitron, etc.).
- Calculated TE copy number and genomic contribution per TE order.
- Generated order-level summary tables for both species.

November 20

- Created comparative TE order tables between *O. sativa* and *O. brachyantha*.
- Generated bar plots comparing TE order composition between species.
- Identified LTR retrotransposons as the dominant contributors in *O. sativa*.
- Performed exploratory LTR family-level summaries (copy number and total bp).
- Evaluated LTR family differences between species.

- Decided to retain family-level analyses for Supplementary material only.
- Designed a window-based approach for chromosomal distribution analysis.
- Partitioned chromosomes into non-overlapping 100 kb windows.
- Computed TE density (bp/kb) per window for both species.

November 23

- Computed gene density (gene count per window) using the same 100 kb windows.
- Merged TE density and gene density into unified datasets for correlation analysis.
- Performed Spearman rank correlation tests between TE density and gene density.
- Observed a strong negative correlation in *O. sativa* and a positive correlation in *O. brachyantha*.
- Recorded correlation coefficients and p-values.

November 25

- Generated scatter plots of TE density versus gene density.
- Interpreted differences in genome organization between the two species.

November 26

- Generated chromosomal TE density profiles using 100 kb windows.
- Visualized TE density along individual chromosomes for both species.
- Identified strong TE clustering in *O. sativa* and smoother profiles in *O. brachyantha*.
- Selected representative chromosomes for main Results figures.
- Prepared additional chromosome plots for Supplementary figures.

November 28

- Defined gene-associated regions as ± 2000 bp flanking sequences around genes.
- Generated gene flanking regions using GenomicRanges.
- Sampled random genomic regions of identical size as background controls.
- Calculated TE density in gene-flanking regions and random regions.
- Performed Wilcoxon rank-sum tests to compare TE densities.
- Observed strong TE depletion near genes in *O. sativa* and weaker depletion in *O. brachyantha*.

November 30

- Compiled summary statistics (mean, median, standard deviation) for TE densities.
- Finalized statistical outputs for inclusion in the Results section.

December 1 – December 23

No project work conducted (due to exams).

December 24

- Resumed project work and reviewed all TE analyses for completeness.
- Organized TE results into genome-wide, chromosomal-scale, and gene-scale analyses.
- Identified tables and figures for main text versus Supplementary material.
- Drafted the Materials and Methods section for the TE analysis.
- Clearly described data sources, window-based analyses, and statistical tests.

December 28

- Drafted the Results section for the TE analysis.
- Integrated numerical results, statistical tests, and figure references.
- Drafted the Discussion section interpreting TE-related findings.
- Linked results to genome size variation and gene–TE spatial relationships.

January 3

- Added and verified literature references supporting TE classification and distribution.
- Checked consistency between Methods, Results, and Discussion sections.
- Ensured all required analyses and statistics were included.

January 4

- Cleaned scripts by removing unused exploratory code.
- Finalized Supplementary tables and figures.

January 7

- Completed the transposable element analysis component.
- Performed final review and prepared materials for submission.