

Input data

- Reference genome (.fa)
- Annotation file (.gff3)
- RestrictionDigest position_FB_BF output
- User-defined size selection ranges

1 Generate gene-associated regions

Script: 1_generate_ensembl_gene_flanking_regions.py

Output:

GFF3 file with non-overlapping genes, and their associated 1500 bp upstream and 1500 bp downstream regions.

2 Convert RestrictionDigest output to BED

Script: 2_restrictiondigest_position_to_bed.py

Output:

BED file of AB/ BA restriction fragments

3 Summarize digestion fragments and feature intersections

Script: 3_summarize_fragment_gff_intersections.py

Output:

Data frame containing fragments count by sample size, estimated sampled genome percentage, genomic feature recovery

4 Select the best global epi-ddRADseq desing

Script: 4_select_best_epi_ddrad_desing.py

Output:

Two data frame: one containing best enzyme pair + size selection range and a second with the top-5 ranked candidate designs

5 Select the best epi-ddRADseq desing per genomic feature

Script: 5_select_best_epi_ddrad_desing_per_element.py

Output:

Dataframe with best design per genomic feature sampling (genes, upstream/ downstream regions, NLR-loci, TE, ecc)

Final output

- Species-specific epi-ddRADseq design
- target-aware enzyme-pair and size-selection recommendations.