

TE Density: A Software Tool to Investigate the Biology of Transposable Elements

Scott J. Teresi, Michael B. Teresi, Patrick P. Edger

MICHIGAN STATE
UNIVERSITY

Teresi, S.J., Teresi, M.B. & Edger, P.P. TE Density: a tool to investigate the biology of transposable elements. *Mobile DNA* 13, 11 (2022).
<https://doi.org/10.1186/s13100-022-00264-4>

BACKGROUND:

- A TE's potential to impact phenotype is partially a function of its location in the genome.
- Previous research has shown TEs' ability to impact the expression of neighboring genes, however our understanding of this trend is hampered by the exceptional amount of diversity in the TE world, and a lack of publicly available computational methods that quantify the presence of TEs relative to genes.
- Here we have developed a tool to more easily quantify TE presence relative to genes through the use of only a gene and TE annotation, yielding a new metric we call TE Density.

RESULTS & DESIGN:

- TE Density is defined as the proportion of TE-occupied base-pairs relative to a gene and a given a measurement window-size.
- This new pipeline reports TE density for each gene in the genome, for each type descriptor of TE (order and superfamily), and for multiple positions and distances relative to the gene start/stop position (upstream, intragenic, and downstream) over sliding, user-defined windows.
- In this way, we overcome previous limitations to the study of TE-gene relationships by focusing on all TE types present in the genome, utilizing flexible genomic distances for measurement, and reporting a TE presence metric for every gene in the genome.

WHAT QUESTIONS CAN YOU ASK WITH THIS DATA?

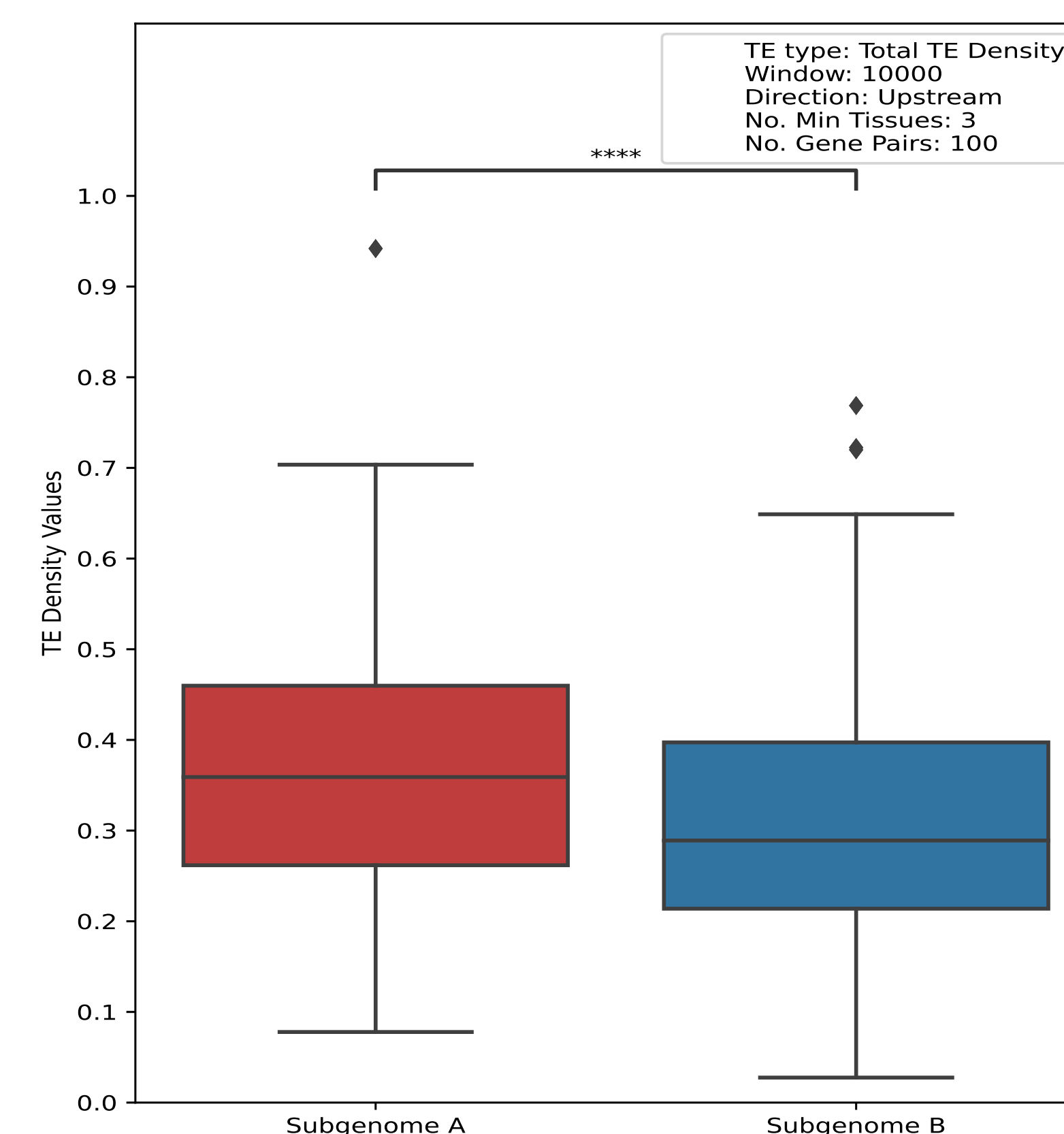
- How does TE presence relate to gene expression?
- Do orthologs differ in TE presence? Do polyploid subgenomes differ?
- What are some general genome-wide trends of TE presence relative to genes?
- Which genes are over- or under-enriched for TEs? If so, what patterns emerge for different TEs and different genes?

CODE AVAILABILITY & REQUIREMENTS:

- **Input data: gene annotation & TE annotation, that's all!**
- TE Density is open source and freely available at https://github.com/sjteresi/TE_Density. The code utilizes Python 3.8.0 and most genomes will require usage of a computing cluster.

POTENTIAL APPLICATIONS AND EXAMPLES:

Subgenome Differences:



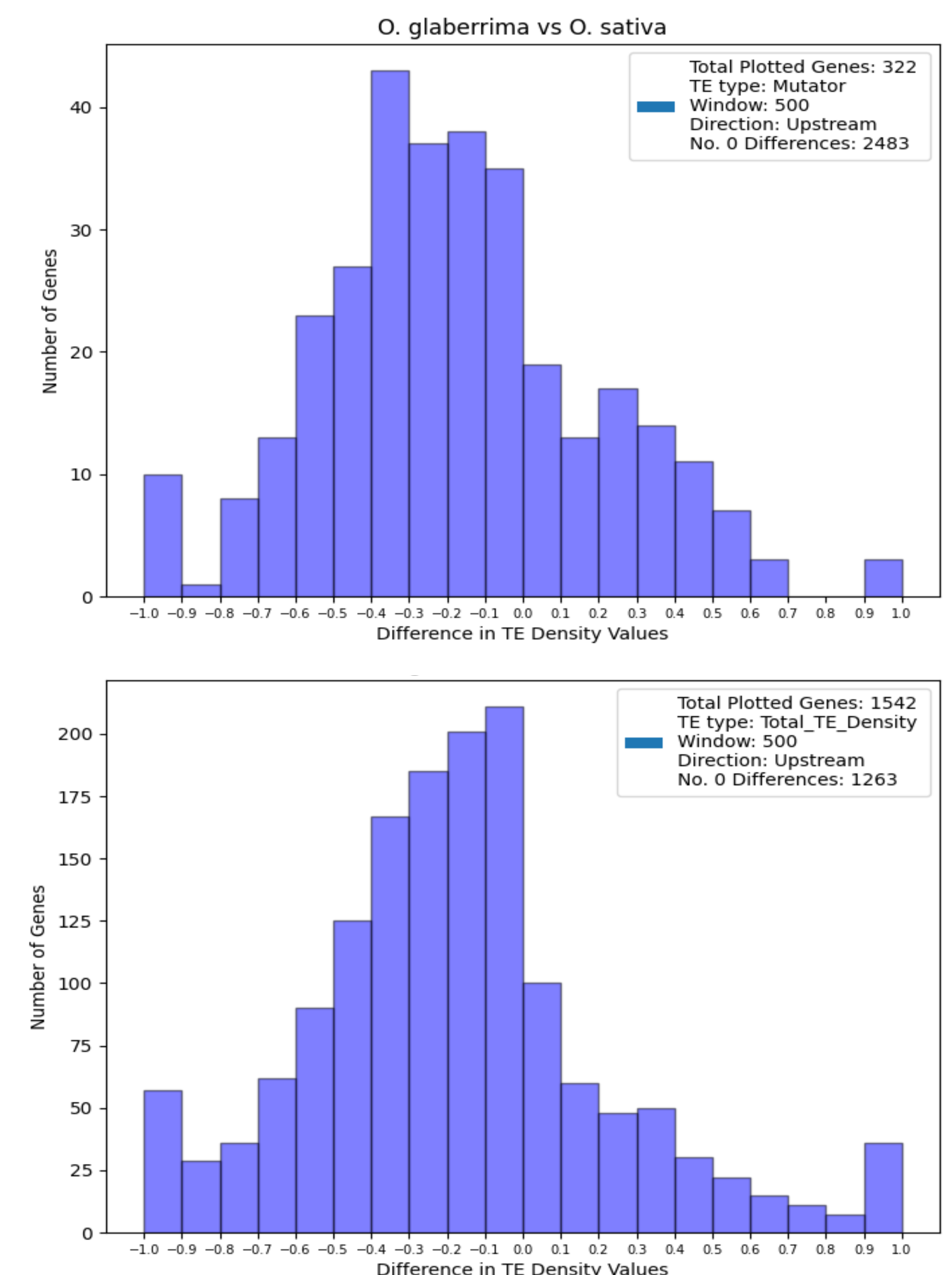
GO Enrichment:

PANTHER GO-Slim Biological Process	Fold Enrichment Value	FDR
Unclassified	1.11	0.00552
Biological process (GO:0008150)	0.49	0.00276
Cellular process (GO:0009987)	0.48	0.00687
Organic substance metabolic process (GO:0071704)	0.46	0.0279
Metabolic process (GO:0008152)	0.46	0.0175
Primary metabolic process (GO:0044238)	0.44	0.0283
Nitrogen compound metabolic process (GO:0006807)	0.42	0.0282
Cellular metabolic process (GO:0044237)	0.42	0.0157

Genomes Utilized Above:

- *Vaccinium corymbosum* (blueberry)
- *Oryza sativa* & *Oryza glaberrima* (Asian rice & African rice)
- *Luciobarbus capito* (carp)

Syntelog Differences:



Expression Profiles and Gene Counts:

