1.11

Exploring the genomic diversity of *Eragrostis tef* through pangenome graph analysis

Sofia Sisti[1], Giulio Bianca[1], Mirko Coggi[1], Guglielmo Bruno[1], Beatrice Branchini[1], Ettore Riccucci[2], Leonardo Caproni[2] Simone Scalabrin[3], Sara Castelletti[3], Matteo Dell'Acqua[2], Davide Scaglione[3], Guido Walter Di Donato[1]

[1]GenoGra s.r.l., Milan, Italy, [2]Institute of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, [3]IGA Technology Services, Udine

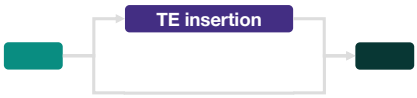
BACKGROUND

Eragrostis tef is an allotetraploid gluten free cereal (2n = 4x = 40) of the Horn of Africa, with **duplicated gene** copies on the A and B subgenomes.

- **Seed colour sets its value**, white grain sells higher and brown is richer in protein, and the flavonoid and anthocyanin pathway behind it is often switched off by transposable element insertions or by lost gene copies.
- **Structural variants** behind a high value trait: invisible on a single reference, explicit alleles on a **pangenome graph**, as below.

SINGLE LINEAR REFERENCE

PANGENOME GRAPH



AIM OF THE STUDY

Rebuild the pangenome. One graph per chromosome from the 26 teff assemblies, separating the A and B subgenomes.

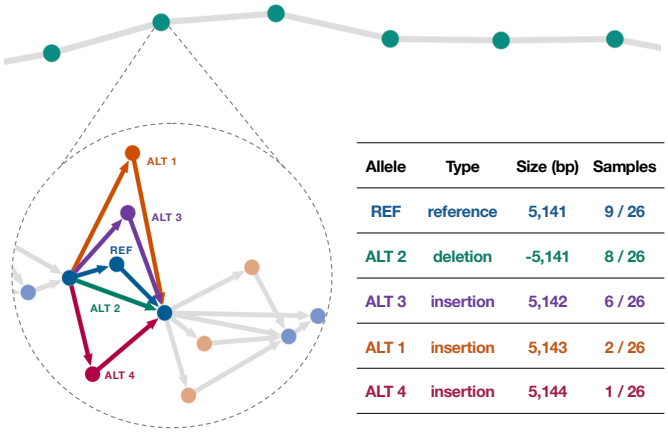
Read the variation. Compare the structural alleles at the seed colour locus on chr 4A and 4B.

Link paths to biology. Load annotations and seed colour metadata to analyze accessions carrying the 5,142 bp insertion.

RESULTS

1 - Structural variation in the GenoGra platform

From the chromosome scale assemblies of the Sant'Anna collection the *teff* pangenome was built: **one graph per chromosome** across all 26 assemblies. The graph is browsed first as a **high-resolution view** of Chr4B, then as a **zoom on the seed colour locus** where the TE insertion and the deletion appear.

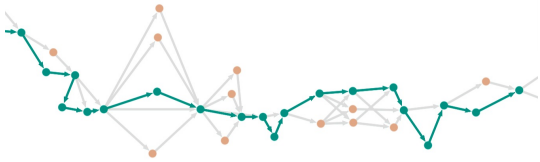


Allele	Type	Size (bp)	Samples
REF	reference	5,141	9 / 26
ALT 2	deletion	-5,141	8 / 26
ALT 3	insertion	5,142	6 / 26
ALT 1	insertion	5,143	2 / 26
ALT 4	insertion	5,144	1 / 26

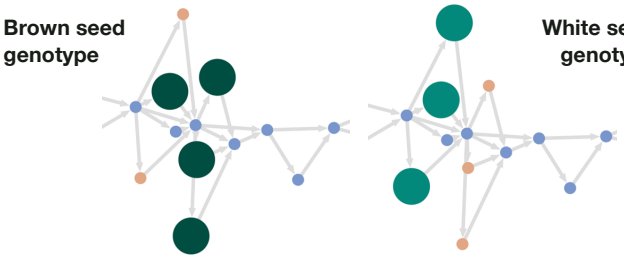
Bubbles are resolved into insertions, deletions and presence absence variants, and each sample path shows **which accessions carry them**, brown or white seeded.

2 - Annotation on the graph

An annotation file can be loaded for **each sample**, so gene models are read on its own path and not borrowed from one genome. Shown here is the annotation of the **TE gene** on the reference used, at the seed colour locus.



3 - Paths filtered by seed colour



6 of 11 brown phenotype samples carry the insertion and **5** carry the deletion. **Every white** phenotype sample carries the insertion, suggesting **its association with seed colour.**

SUMMARY

- **Reads land on the right copy:** homoeologous regions stay distinct, so chr4A and chr4B are not confused.
- **From variant to marker:** the insertion itself can be genotyped as a marker, instead of nearby SNPs that only stand in for it.
- **Trait read in one view:** sample paths and seed colour metadata are read on the same graph, in one browser.

PROSPECTS

Move to **genotyping on the graph**, where reads are mapped on all paths and structural alleles are scored directly, giving more accurate calls than a single reference and ready to use seed colour markers.

Jones, M.R.W., Kebede, W., Teshome, A. *et al.* Population genomics uncovers loci for trait improvement in the indigenous African cereal tef (*Eragrostis tef*). *Commun Biol* **8**, 807 (2025). https://doi.org/10.1038/s42003-025-08206-5