6.13

Chromosome-scale graph pangenomics of *Ricinus communis*: decoding intraspecific diversity to accelerate sustainable crop improvement

Sofia Sisti[1], Giulio Bianca[1], Mirko Coggi[1], Guglielmo Bruno[1], Beatrice Branchini[1], Simone Scalabrin[2], Sara Castelletti[2], Davide Scaglione[2], Guido Walter Di Donato[1]

[1]GenoGra s.r.l., Milan, Italy, [2]IGA Technology Services, Udine

BACKGROUND

Ricinus communis is among the most promising non-food biofuel feedstocks: its seeds reach 45–55% oil, the highest of any non-food crop, feeding a global market worth +$12B.

- **Intraspecific diversity remains underexplored:** a single linear reference misses key variation slowing breeding for oil yield and stress resilience to 10–15-year cycles[1].
- The **bottleneck has shifted to secondary analysis**, where reference bias hides the complex, rare and structural variation that matters most[2].

AIM OF THE STUDY

This study aims to build a complete, **chromosome-resolved graph pangenome** of *R. communis* that captures species-wide variation enriched **with functional and agronomic-trait annotations**, and turn this raw diversity into ranked, queryable breeding candidates prioritised by predicted functional impact.

SUMMARY

An open, chromosome-resolved pangenome atlas integrating structural and functional variation across 103 *Ricinus communis* accessions, with haplotype-level resolution

- By integrating **VEP and trait annotations**, a single automated query links each variant to its gene, agronomic trait and predicted functional impact.
- This enables millions of variants to be narrowed down to breeding-ready candidates, including a **moderate-impact missense variant on chromosome 2** associated with single-seed weight.

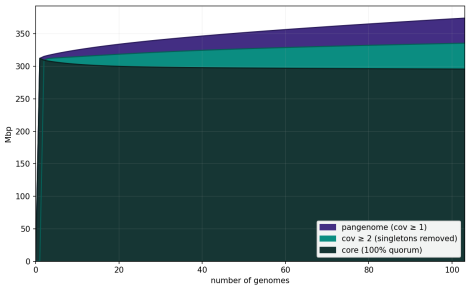
PROSPECTS

- Integrate **additional phenotyped accessions** to enable robust GWAS to prioritise trait-associated variants and validate markers
- Use the **pangenome** to interpret variants of unknown significance
- Field validation to support **marker-assisted selection**, helping accelerate breeding programmes

RESULTS

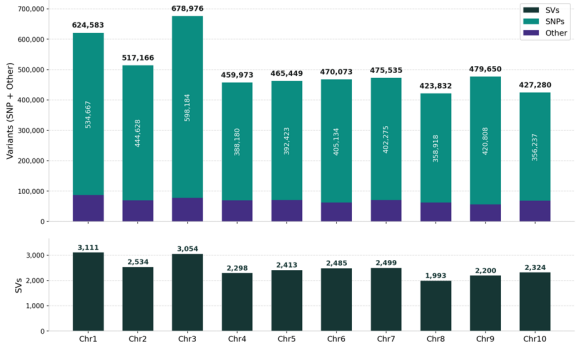
1 - An open, expanding pangenome

Across 103 accessions, the pangenome expands to **~375 Mbp**, while the **core genome** stabilises at **~300 Mbp**. The growing gap represents accessory and singleton sequences **missing from any linear reference**, highlighting the **intraspecific diversity** captured by the **genome graph**.



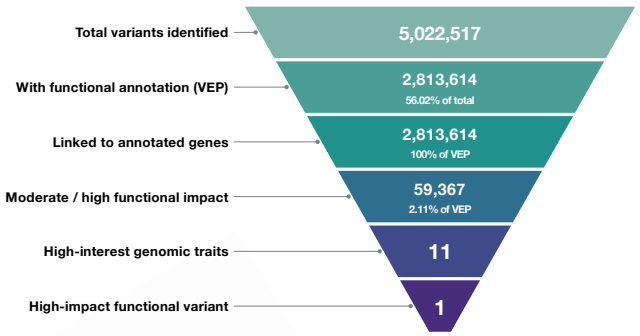
2 - A dense, multi-class variant landscape

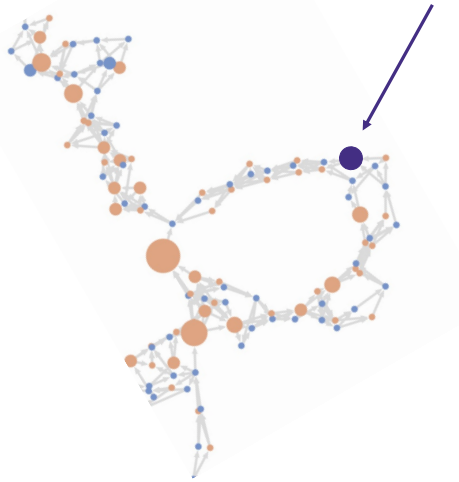
The pangenome graph encodes **5,022,517 variants** across the ten chromosomes, capturing a diversity that no single linear reference can represent. **SNPs represent ~86%**, while indels and structural variants account for the remaining **~14%**, represented as alternative graph paths.



3 - From millions of variants to a few breeding candidates

A single automated query on the **GenoGra platform** filters millions of variants by **VEP annotation, genic location,** and **colocalisation with trait loci**, returning a concise set of candidates with their associated gene, trait, and impact class.





As a proof of concept, a single query identified a **moderate-impact missense variant** on **chromosome 2**, associated with **single-seed weight.** On the **GenoGra** platform, the corresponding graph node is highlighted in **violet**, turning a complex genomic signal into an **immediately interpretable breeding target.**

[1] Xu, Wei, et al. "Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean." Genome biology 22.1 (2021): 113.
[2] Foster, Jeffrey T., et al. "Single nucleotide polymorphisms for assessing genetic diversity in castor bean (Ricinus communis)." BMC Plant Biology 10.1 (2010): 13.