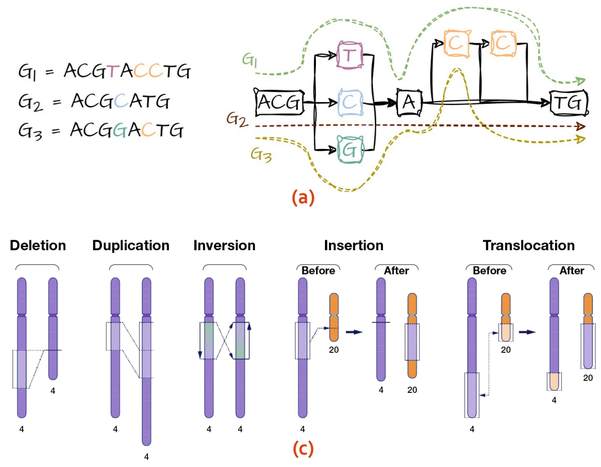
Pangenome-based characterization of novel genetic variants

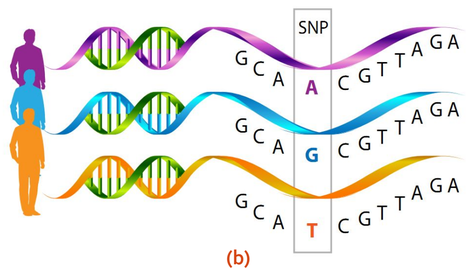
Background

(a) **Pangenome:** collection of genomes from a population
- ↪ *helps in reducing the "reference-bias"*

(b) **Short Variants:** SNPs/indels
- ↪ *health and disease (e.g., diabetes)*
- ↪ *drug response*
- ↪ *other phenotypic traits*

(c) **Structural Variants:** big (≥50bp) rearrangements
- ↪ *genetic disorders (e.g., autism, schizophrenia…)*
- ↪ *cancer (e.g., melanoma, breast, prostate…)*



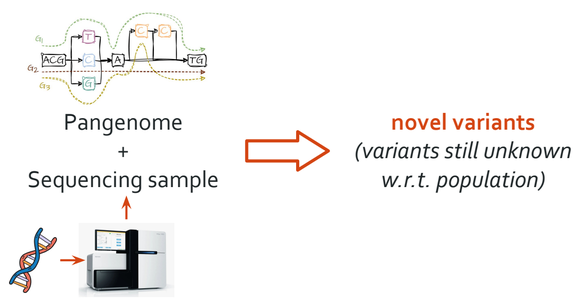


Images from genome.gov and thesciencenotes.com

Goal

"Joint variant calling and pangenome augmentation"

pangenome-based, assembly-free, and mapping-free

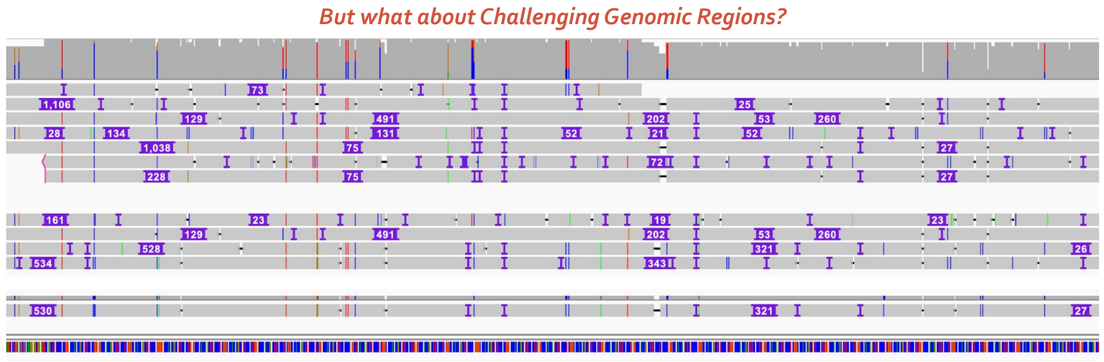


State-of-the-art

A lot of ***reference-based*** approaches
- **short reads alignment (✗ inaccurate)**
  - …
- **long reads alignment (✗ "inaccurate")**
  - *SVIM [Bioinformatics, 2019]*
  - *cuteSV [Genome Biology, 2020]*
  - *SVDSS [Nature Methods, 2023]*
  - *sniffles2 [Nature Biotechnology, 2024]*
  - …
- **genome assembly (✗ expensive)**
  - *dipcall [Nature Methods, 2018]*
  - *PAV [Science, 2021]*
  - *VolcanoSV [Nature Communication, 2024]*
  - …

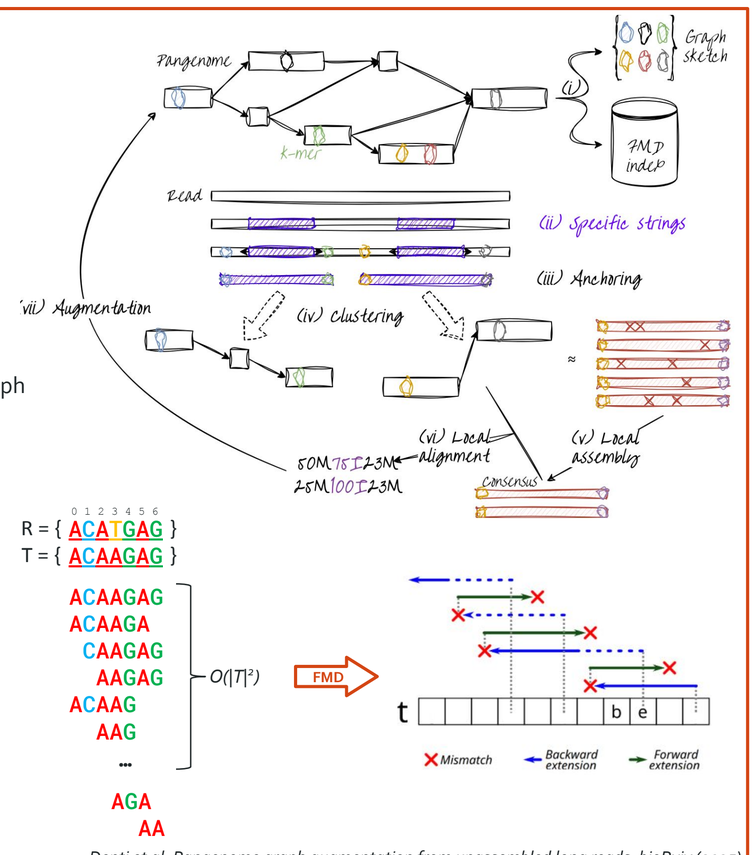
Few **pangenome-based** approaches **(✗ expensive)** *[biorXiv, 2024]*
- based on genome assembly and read alignments to graphs

But what about Challenging Genomic Regions?



Approach

(i) index pangenome
(ii) compute ***specific strings***
(iii) anchor specific strings to graph
(iv) cluster specific strings
(v) "summarize" clusters
(vi) realign consensus back to local graph
(vii) augment graph and get variations



Denti et al. Pangenome graph augmentation from unassembled long reads. bioRxiv (2025)

Results

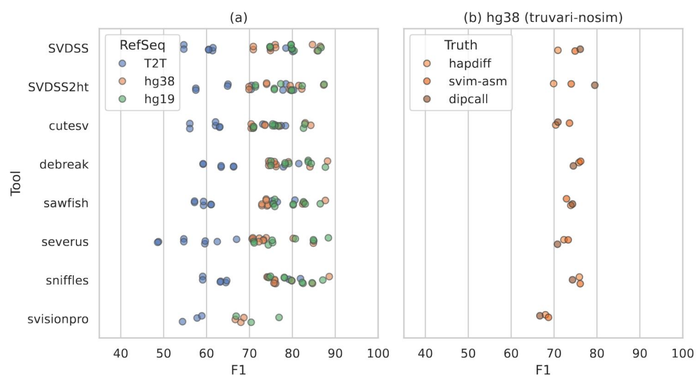
Currently testing C++ implementation on HPRC pangenomes and simulated PacBio HiFi data

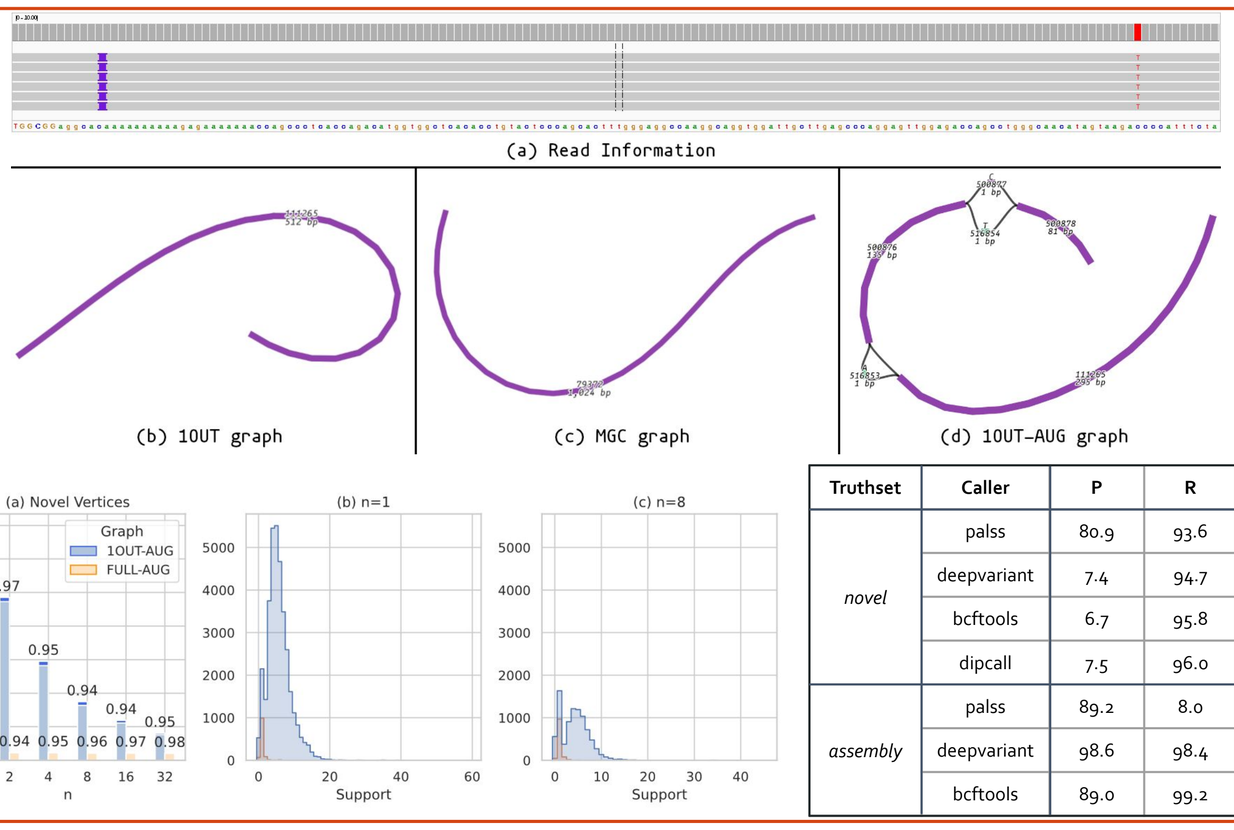
⇒ augmentation is effective *(+15 challenging regions with 65 specific variants w.r.t. assembly-based)*

⇒ good precision (although there is still room for improvements)

⇒ promising results on novel SNP calling

⇒ SVs evaluation under investigation *(although particularly complex*)*





Truthset	Caller	P	R	F1
novel	palss	80.9	93.6	86.8
	deepvariant	7.4	94.7	13.7
	bcftools	6.7	95.8	12.6
	dipcall	7.5	96.0	13.9
assembly	palss	89.2	8.0	14.7
	deepvariant	98.6	98.4	98.5
	bcftools	89.0	99.2	93.8

**Denti et al. Anyone can be the best: Impact of diverse methodologies on the evaluation of structural variant callers. bioRxiv (2025)*

This work is supported by the European Union's Horizon programme under grant agreement No. 101180581 (ASVA-CGR)

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
Univerzita Komenského v Bratislave

MATFYZ CONNECTIONS

Luca Denti, Tomáš Vinař, Broňa Brejová
Comenius University Bratislava, Slovakia

Paola Bonizzoni, Rayan Chikhi, Thomas Krannich
University of Milano-Bicocca, Italy / Institut Pasteur, France / Robert Koch Institute, Germany

Fereydoun Hormozdiari
UC Davis, California (US)