



Construction and analysis of pangenomic graphs in soft wheat from 14 public genomes

Djémilatou Ouandaogo, Vincent Pailier, Hélène Rimbert, Pierre Sourdille, Pauline Lasserre-Zuber

► To cite this version:

Djémilatou Ouandaogo, Vincent Pailier, Hélène Rimbert, Pierre Sourdille, Pauline Lasserre-Zuber. Construction and analysis of pangenomic graphs in soft wheat from 14 public genomes. JOBIM 2025, Jul 2025, Bordeaux, France. , pp.1. hal-05194677

HAL Id: hal-05194677

<https://hal.inrae.fr/hal-05194677v1>

Submitted on 31 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Construction and analysis of pangenomic graphs in soft wheat from 14 public genomes

Djémilatou OUANDAOGO, Vincent PAILLER, Hélène RIMBERT, Pierre SOURDILLE and Pauline LASSERRE-ZUBER

Context

Bread wheat (*Triticum aestivum*) is a vital staple crop cultivated worldwide. However wheat has a large hexaploid (15 Gb, AABBDD, $2n = 6x = 42$) and complexe genome with 85% of transposable elements (TE) and numerous structural variations (SV) that hampers genetic gain. To accelerate wheat breeding & better understand its genetic diversity, we constructed a wheat pangenome graph from 14 genomes using Minigraph-Cactus [1].

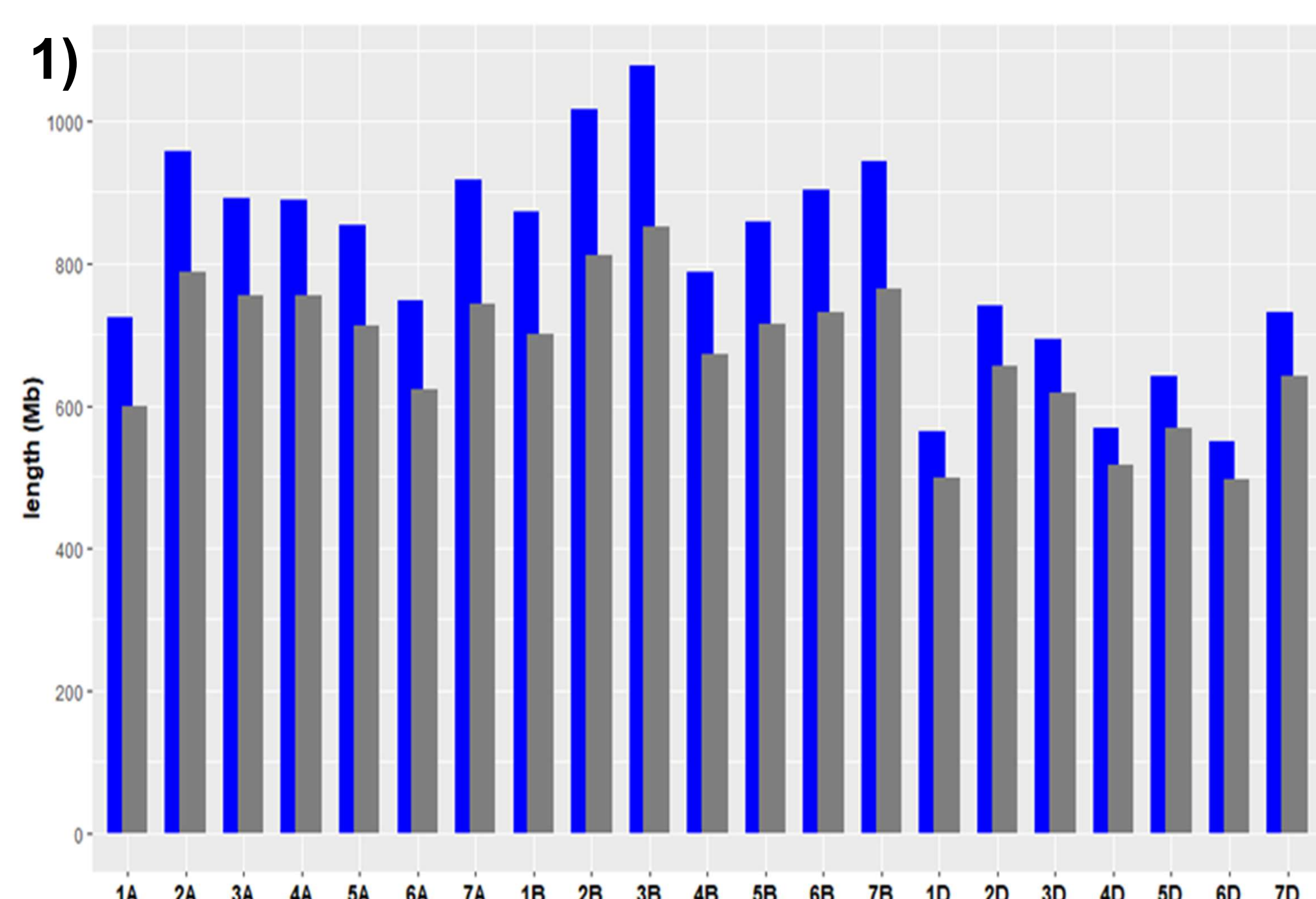
Material

- **13 public genomes:** Chinese Spring [2], Renan [3], Jagger, Julius, ArinaLrFor, LongReach_Lancer, Stanley, Landmark, Mace, Norin61, PI190962 (spelta), SY Mattis [4], zang1817 [5].
- **+ 1 genome** assembled and annotated by our team: Courtot
- **public gene annotations**

Method

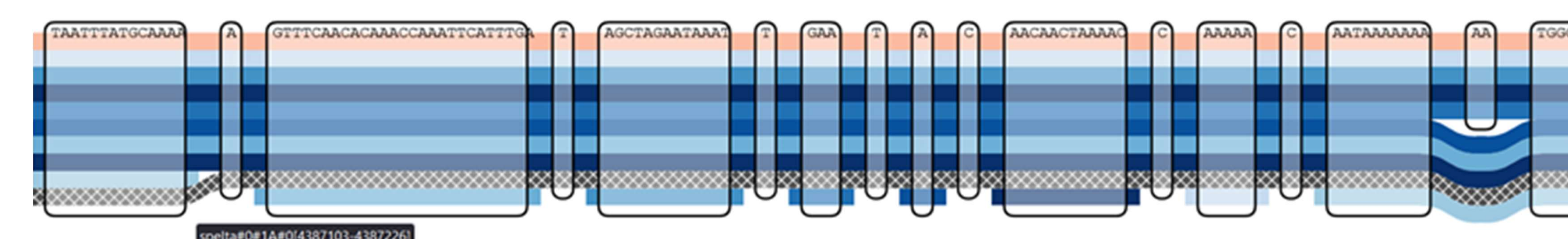
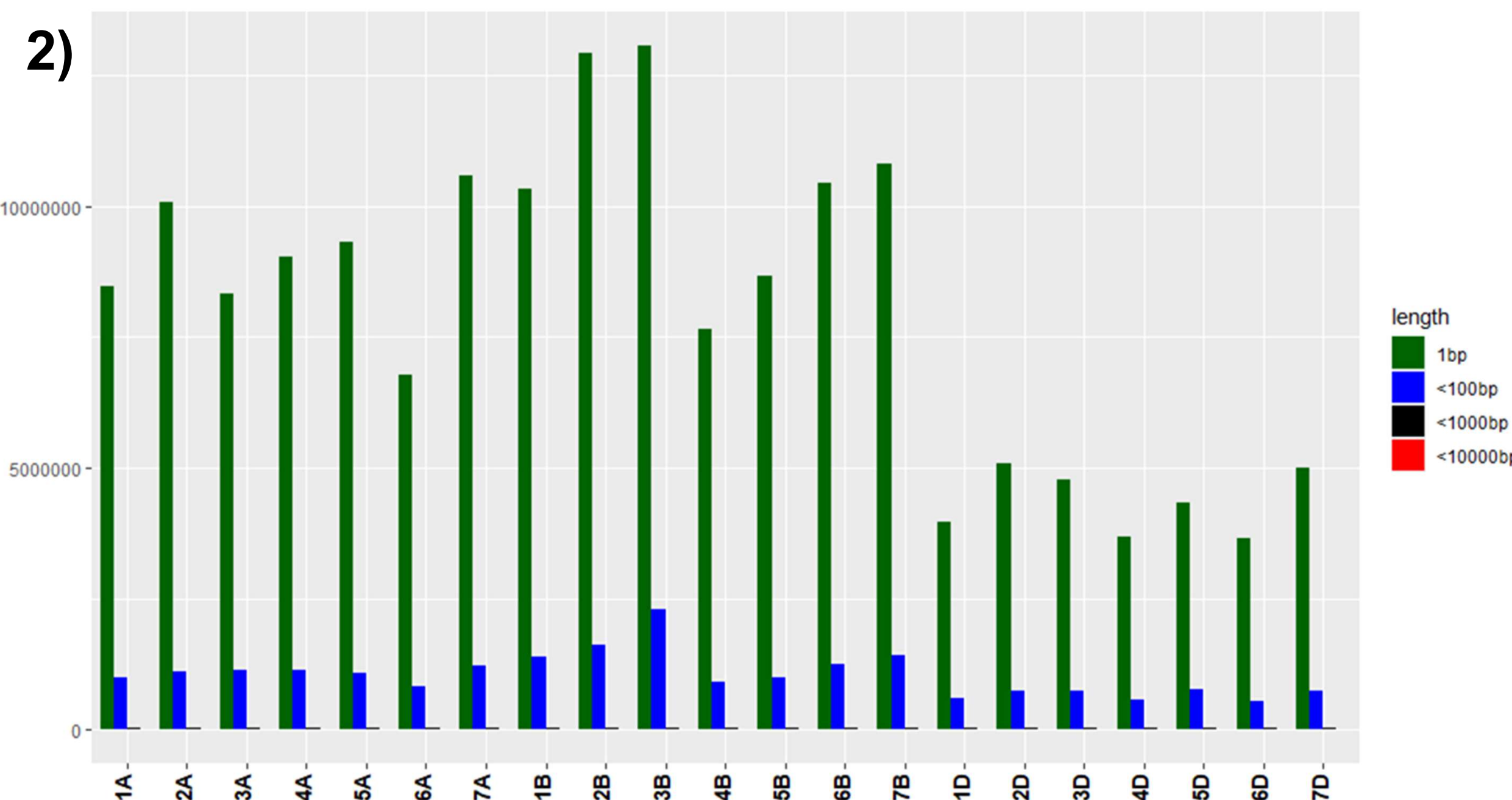
- **Minigraph-Cactus** [1]: pangenome graphs construction
- **vg**: VCF files and sub-graph extraction
- **GrAnnoT** [6]: gene annotation transfer
- **CLARITE_smk**: TE annotation [7] https://forge.inrae.fr/umr-gdec/clari-te_smk
- **tidyverse, ggplot2, UpSetR**: satatistics and visualization

Results

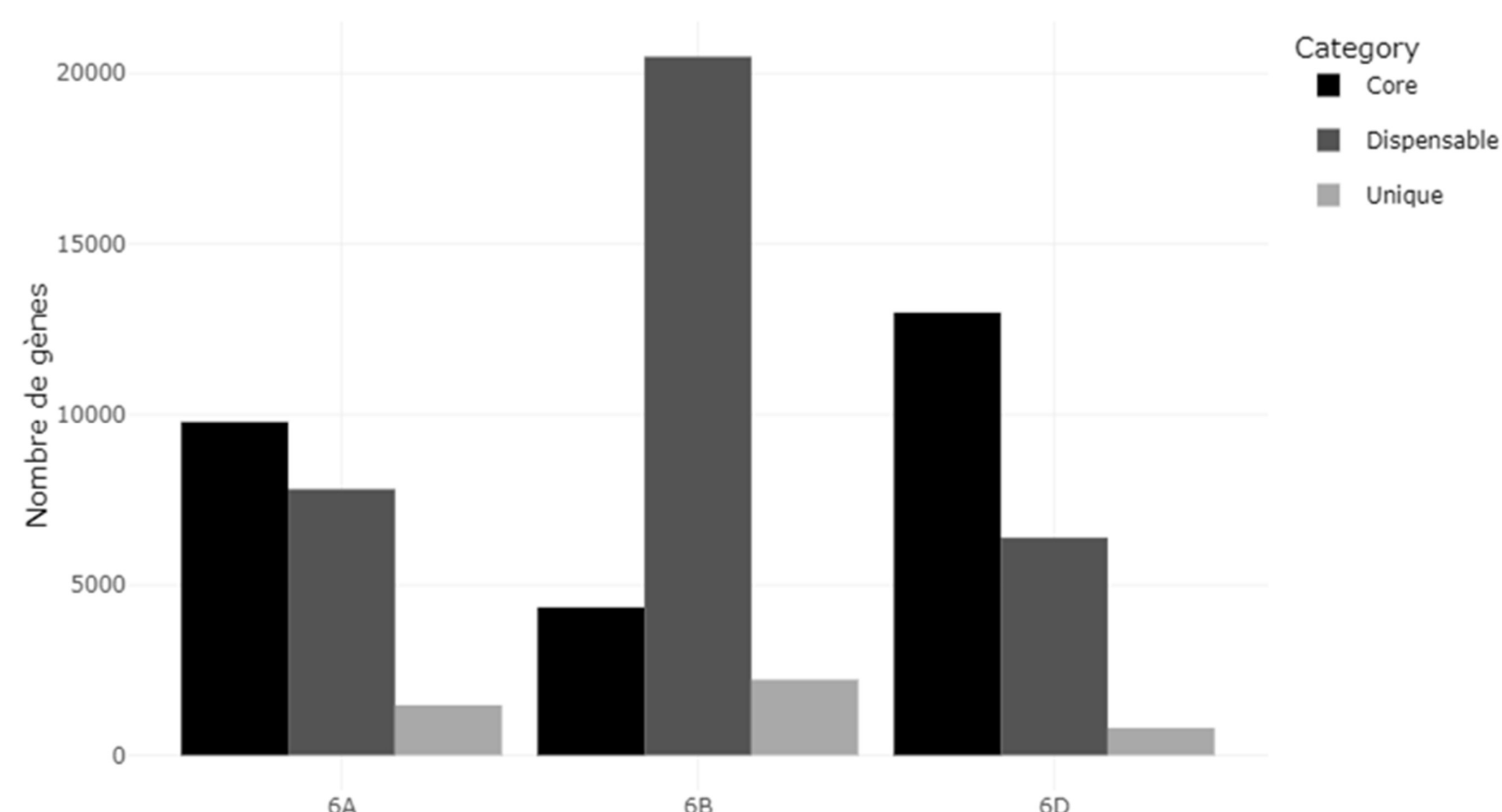


2) Structural Variations (SV) count per chromosome by length (bp):
SVs were mostly SNP. 2B, 3B, and 7B chromosomes showed higher amounts of SVs, explained by higher sizes and repeats content of sub-genome B.

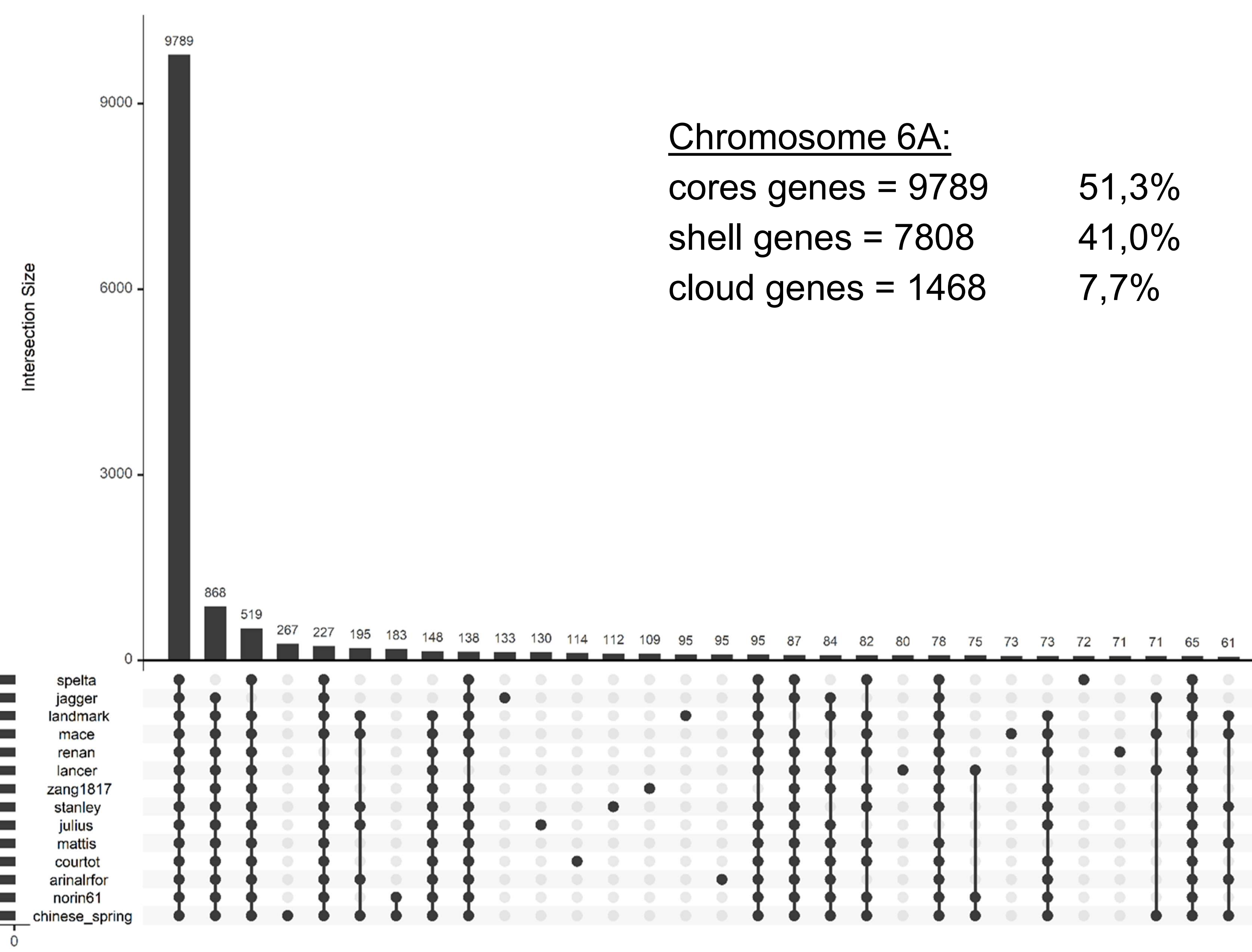
1) Chromosome sizes in linear genomes (REF) vs. Pangenome Graphs (PG):
Chromosomes in PG were consistently longer than those in REF (Chinese Spring), being more exhaustive by integrating variable sequences.



3) Gamma gliadin 1A locus sub-graph (SequenceTubeMap)



4a) Core, Dispensable & Shell genes in Chromosomes 6 ABD:
Chromosome 6B showed a high number of dispensable genes, indicating strong evolutionary dynamics, to investigate further.



4b) Genes sharing on Chromosome 6A: a majority of core genes were shared across genomes, alongside genome-specific & nearly specific sets, reflecting substantial inter-accession variability. These findings highlights the genetic richness and diversity revealed and comprised in the PG.

Perspectives and Conclusion

- Core, shell, and cloud genomes will be identified at the pangenome level and characterized for genes functions, notably shell and cloud genes that could be related to adaptive traits to climate change.
- Graph augmentation & sub-graph extraction will permit key agronomic loci analyses for genetic variability, such as for glutenin locus 1ABD (3)).
- TE will be analysed for their dynamic and detection of insertion variations will be applied using panREPET pipeline [8].
- Structural variants calling will be applied directly from the graph by aligning new sequencing data to supply association studies.
- Pangenome graphs offer a richer view of wheat genomic diversity than linear references, providing valuable insights for breeding.

INRAE Auvergne Rhône-Alpes

5, chemin de Beaulieu, 63000 Clermont-Ferrand
pauline.lasserre-zuber@inrae.fr

Acknowledgement to Mésocentre Clermont-Auvergne of the Université Clermont Auvergne for providing help, computing and storage resources.

References: [1] Hickey G, Monlong J, Ebler J, et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol* 2024. [2] IWGSC, Appels R, Eversole K, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 2018; 361: eaar7191. [3] Aury JM, Engelen S, Istace B, Monat C, Lasserre-Zuber P, et al. Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *GigaScience* 2022. [4] Walkowiak S, Gao L, Monat C, et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* 2020. [5] Guo W, Xin M, Wang Z, et al. Origin and adaptation to high altitude of Tibetan semi-wild wheat. *Nat Commun* 2020. [6] Marthe N, Zytynicki M, Sabot F. GrAnnoT, a tool for efficient and reliable annotation transfer through pangenome graph. *BioRxiv* 2025. [7] Daron J, Glover N, Pingault L, et al. Organization and evolution of transposable elements along the bread wheat chromosome 3B. *GenomeBiology* 2014. [8] Saidi S, Blaison M, Del Pilar Rodríguez-Ordóñez M, Johann Confais, Hadi Quesneville. The *Brachypodium distachyon* pangenome highlights Transposable Element dynamics in the species. *BioRxiv* 2024.