



HAL
open science

Understanding the limits of pangenome graphs for the analysis of large inversions in a complex of butterfly species

Sandra Romain, Fabrice Legeai, Claire Lemaitre

► To cite this version:

Sandra Romain, Fabrice Legeai, Claire Lemaitre. Understanding the limits of pangenome graphs for the analysis of large inversions in a complex of butterfly species. 2024 - International Environmental and Agronomical Genomics symposium, Feb 2024, Toulouse, France. pp.1-1. hal-04721489

HAL Id: hal-04721489

<https://hal.science/hal-04721489v1>

Submitted on 4 Oct 2024

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 4.0 International License

Understanding the limits of pangenome graphs for the analysis of large inversions in a complex of butterfly species

Sandra Romain¹, Fabrice Legeai², Claire Lemaitre¹

¹ Univ Rennes, Inria, CNRS, IRISA, 35042, Rennes, France;

² IGEPP, INRAE, Institut Agro, Univ Rennes, 35653, Le Rheu, France

Among structural variants, inversions are of particular interest in ecology studies as they can introduce phenotypic diversity and have a lasting impact on population genetics by locally impairing recombination^[1]. Pangenome graphs (PG) are a way to represent all scales of genomic diversity in a species. Several tools have been developed to construct PG from genome alignments and to detect variants from the graph topology. PG were shown to be particularly efficient for identifying and genotyping deletions and insertions in model organisms^[2]. However they have not yet been thoroughly assessed on inversion polymorphism.

We propose here to evaluate the ability of PG tools to represent and detect inversions in the case of a complex of butterfly species (*Coenonympha* genus, 9% nt divergence) for which we already detected a dozen of large (> 100 kb) inversions. To compare the tools, we selected a chromosome with 2 large, 6 medium sized (> 1 kb), and 9 small (< 1 kb) inversions, and built PG with 4 state of the art tools. Minigraph^[3] and minigraph-Cactus^[4] failed to build an accurate PG for such a degree of sequence divergence, while PGGB^[5] and Cactus^[6] found at most two inversions. In order to understand how PG handle inversions, we simulated the 17 known inversions in several synthetic chromosomes with increasing levels of single nucleotide divergence. We found that in such simplified graphs, most large simulated inversions are well represented with most tools. However, for smaller inversions and when the sequence divergence is higher, there is a significant variability in how the inversions are represented between PG tools. We analyzed the various obtained topological motifs, leading to methodological avenues for improving the detection of inversions in PG.

Keywords : Pangenome graphs ; structural variants ; inversions ; comparative genomics ; method benchmark

[1] Kirkpatrick M. (2010). How and Why Chromosome Inversions Evolve. *PLOS Biology*, 8(9): e1000501.

[2] Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., ... & Paten, B. (2023). A draft human pangenome reference. *Nature*, 617(7960), 312-324.

[3] Li, H., Feng, X., & Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. *Genome biology*, 21, 1-19.

[4] Hickey, G., Monlong, J., Ebler, J., Novak, A. M., Eizenga, J. M., Gao, Y., ... & Paten, B. (2023). Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology*, 1-11.

[5] Garrison, E., Guarracino, A., Heumos, S., Villani, F., Bao, Z., Tattini, L., ... & Prins, P. (2023). Building pangenome graphs. *bioRxiv*, 2023-04.

[6] Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., ... & Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 587(7833), 246-251.