



HAL
open science

Pangenome graph manipulation for local visualisation with pancat

Siegfried Dubois, Matthias Zytnicki, Thomas Faraut, Claire Lemaitre

► **To cite this version:**

Siegfried Dubois, Matthias Zytnicki, Thomas Faraut, Claire Lemaitre. Pangenome graph manipulation for local visualisation with pancat. MIGGS1 2024 - 1st symposium on Methods for Interfacing with Graphs of Genomic Sequences, Sep 2024, Lille, France. pp.1-1. hal-04724519

HAL Id: hal-04724519

<https://hal.science/hal-04724519v1>

Submitted on 7 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



Abstract

Pangenome graph manipulation for local visualisation with pancat

Siegfried Dubois^{1,2} ✉, Matthias Zytnicki³, Thomas Faraut² and Claire Lemaître¹

¹Univ Rennes, Inria, CNRS, IRISA, Rennes, F-35000, France

²GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan, France

³MIAT, Université de Toulouse, INRAE, 31320, Castanet Tolosan, France

✉ **Corresponding author:** siegfried.dubois@inria.fr

A pangenome graph is a data structure designed to represent variations among a collection of genomes [1], often from the same species, encapsulating the information of an associated multiple sequence alignment for those genomes. This data structure features multiple levels of information. The *topology* describes how the nodes of the graph are linked to each other, while the *content* pertains to the sequences embedded in the graph, labeling the nodes and inscribing specific paths in the graph.

Pangenome graphs are enormous, with potentially billions of nodes and intricate connectivity patterns [2]. Visualizing these data structures is a significant challenge, and needs to be addressed to facilitate pangenome graph analysis. Multiple visualization tools are currently available, each adopting different approaches: for example, some take *paths* into account [3], others focus on global topology [4], and some are based upon the representation of the embedded sequences [5]. However, in many applications displaying the whole graph structure is not required and can be confusing.

In this work, we propose several approaches to visualize local regions of a graph, along with a visualization tool. We introduce algorithms to aid in the local representation of pangenomes, compatible with *state-of-the-art* GFA visualizers. Our methods include subgraph extraction from positions in a genome, compression of small polymorphisms without loss inside a graph, and abstraction of parts of a graph to reduce the number of elements to render. Additionally, we propose a method allowing the simultaneous visualization of two pangenome graphs built from the same genomes. We tested these methods on real data, highlighting hotspots of differences between graphs made with different tools.

Availability: implementations are part of a Python tool, pancat, available as a open-source project. <https://github.com/Tharos-ux/pancat>

References

- [1] J. Baaijens et al. Computational graph pangenomics: a tutorial on data structures and their applications. *Natural Computing*, March 2022.
- [2] F. Andreade et al. Comparing methods for constructing and representing human pangenome graphs. *Genome Biology*, November 2023.
- [3] S. Heumos et al. Pangenome graph layout by path-guided stochastic gradient descent. *Bioinformatics*, 2024.
- [4] R. Wick et al. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, October 2015.
- [5] W. Beyer et al. Sequence tube maps: making graph genomes intuitive to commuters. *Bioinformatics*, December 2019.