Bioinformatics, XXXX, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Applications Note

Subject Section

VRPG: an interactive web viewer for reference pangenome graph

Zepu Miao[1], Jia-Xing Yue[1,*]

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, China.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: With the increasing availability of high-quality genome assemblies, pangenome graph is emerging as a new paradigm in genomics research for identifying, encoding, and interpreting the genetic variation landscape at the population and species levels. To facilitate a better examination of pangenome graph towards novel biological insights, here we present VRPG, a web-based interactive viewer of reference pangenome graph. Compared with its counterparts, VRPG is the only one to provide native and efficient support for rGFA-formatted reference pangenome graph, capable of visualizing large-scale reference pangenome graphs built upon hundreds of genome assemblies. Moreover, VRPG shines with its unique features in highlighting the path of any constitutive assembly along the graph as well as in presenting its corresponding presence/absence and copy number variation, which is highly valuable in both large-scale genome comparison and fine-scale structural variant analysis. To further demonstrate its features and scalability, we applied VRPG to the cutting-edge yeast and human reference pangenome graphs derived from hundreds of input assemblies via a dedicated website.

Availability and implementation: VRPG is written in Python and HTML. It is free for use under the MIT license, available at https://github.com/codeatcg/VRPG .
Demonstration: https://www.evomicslab.org/app/vrpg/

Contact: yuejiaxing@gmail.com
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Long-read-based DNA sequencing has become the go-to-choice for most genome sequencing projects nowadays, empowering the production of chromosome-level telomere-to-telomere (T2T) genome assemblies for diverse organisms (including human) (Yue *et al.*, 2017; Jiao and Schneeberger, 2020; Nurk *et al.*, 2022). With such T2T reference assembly panel continuously expanding at both population and species levels, researchers began to use pangenome graphs to better represent the population- and species-wide genomic variation landscapes in a sequence-resolved manner (Eizenga *et al.*, 2020). Compared with the conventional single linear reference genome, pangenome graph offers more sensitive and accurate read mapping and variant discovery, especially in the presence of sequence polymorphisms and structural variants (Paten *et al.*, 2017). Therefore, a representative pangenome graph is expected to shed novel insights into the interpretation of the genotype-to-phenotype association and the discovery of missing heritability.

While a number of tools have been developed to build pangenome graph based on genome alignments, Minigraph (Li *et al.*, 2020) and seqwish (Garrison and Guarracino, 2022) are among the most popular ones. With Minigraph, a reference pangenome graph can be derived using the reference Graphical Fragment Assembly (rGFA) format. Compared with the standard GFA format, rGFA also records the origin of each graph constitutive segment from the input linear genomes and therefore allows for traversing the pangenome graph along a stable coor-

dinate system. This unique feature makes the reference pangenome graph a natural and intuitive extension to the conventional linear reference genomes.

An intuitive visualization of pangenome graph can greatly assist researchers to explore and understand the global and local genomic variation in their graph representation. To date, several tools have been developed for visualizing pangenome graph, among which Bandage (Wick *et al.*, 2015) and GFAviz (Gonnella *et al.*, 2019) focus more on large-scale topology, while SeauenceTubeMap (Beyer *et al.*, 2019) and MoMI-G (Yokoyama *et al.*, 2019) are more suitable for visualizing fine-scale sequence level details. However, none of these existing tools can be used to visualize the full-scaled rGFA-formatted reference pangenome graphs, due to limitations in either combability or scalability. Moreover, none of these tools can take advantage of the stable coordinate system built-in the reference pangenome graph, and therefore incapable of conveying important biological information such as assembly-specific presence/absence and copy number variation via a direct and intuitive graph representation.

Here we present VRPG, a web-based interactive viewer for reference pangenome graph with full scalability. VRPG takes full advantage of the stable coordinate information encoded in the rGFA format and renders visualization in a fashion that is highly informative in both large-scale genome comparison and fine-scale structural variant analysis.

2 Implementation and feature highlights

VRPG was implemented in Python and HTML. It is easy to deploy and highly efficient in visualizing complex reference genome graphs derived from hundreds of genome assemblies on the fly. In the graph rendered by VRPG, genomic segments from the pre-defined reference assembly are shown along the center line in purple, while segments from non-reference assemblies are shown around the center line in other colors (Figure 1). In VRPG, users can easily obtain the source assembly, genomic coordinates, and mapping depth of each segment by a simple click on the corresponding segment in the graph. Also, VRPG provides a "Highlight" feature for highlighting the path of any constitutive genome assembly in the pangenome graph. These two features in combination are highly valuable for both large-scale genome comparison and fine-scale structural variant analysis.

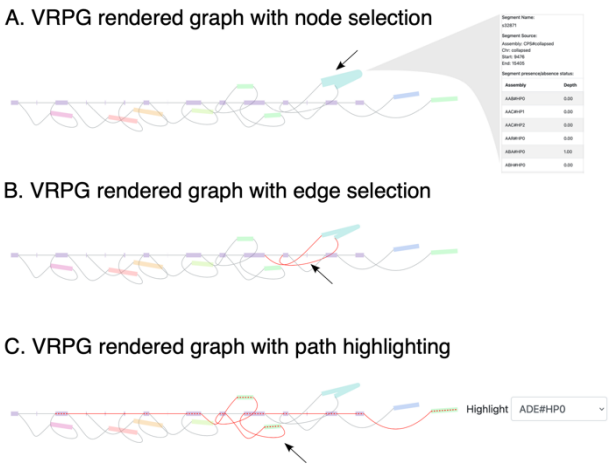

Fig. 1. The reference pangenome graph visualization with VRPG. A) VRPG visualization with the node s32871 selected. B) VRPG visualization with the two edges surround the node s32871 selected. C) VRPG visualization with the assembly-to-graph mapping path of the assembly ADE#HP0 highlighted.

3 Application demonstration

To demonstrate the application of VRPG in real world examples, we set up a demonstration website (https://www.evomicslab.org/app/vrpg/) to visualize reference pangenome graphs derived from 143 yeast and 90 human genome assemblies respectively. The human reference pangenome graph is a public dataset, whereas the yeast reference pangenome graph is built based on the *Saccharomyces cerevisiae* reference assembly panel (ScRAP) that we recently constructed (O'Donnell *et al.*, 2022) (Supplementary Note).

4 Conclusions

We developed VRPG, the first native viewer for reference pangenome graphs with full scalability. We expect VRPG to become a highly useful visualization tool to help researchers to explore the full power of reference pangenome graph.

Acknowledgements

We are grateful to Dr. Jing Li for valuable feedback on the manuscript. We thank Dr. Heng Li for making the human reference pangenome graph available.

Funding

This work is supported by National Natural Science Foundation of China (32070592 to J.-X. Y.), Natural Science Foundation of Guangdong Province (2022A1515010717 to J.-X. Y.), Guangdong Basic and Applied Basic Research Foundation (2019A1515110762 to J.-X. Y.), and Guangdong Pearl River Talents Program (2019QN01Y183 to J.-X. Y.).

Conflict of Interest: none declared.

References

Beyer,W. *et al.* (2019) Sequence tube maps: making graph genomes intuitive to commuters. *Bioinformatics*, **35**, 5318–5320.

Eizenga,J.M. *et al.* (2020) Pangenome Graphs. *Annual Review of Genomics and Human Genetics*, **21**, 139–162.

Garrison,E. and Guarracino,A. (2022) Unbiased pangenome graphs. *Bioinformatics*, btac743.

Gonnella,G. *et al.* (2019) GfaViz: flexible and interactive visualization of GFA sequence graphs. *Bioinformatics*, **35**, 2853–2855.

Jiao,W.-B. and Schneeberger,K. (2020) Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat Commun*, **11**, 989.

Li,H. *et al.* (2020) The design and construction of reference pangenome graphs with minigraph. *Genome Biology*, **21**, 265.

Nurk,S. *et al.* (2022) The complete sequence of a human genome. *Science*, **376**, 44–53.

O'Donnell,S. *et al.* (2022) 142 telomere-to-telomere assemblies reveal the genome structural landscape in Saccharomyces cerevisiae. 2022.10.04.510633.

Paten,B. *et al.* (2017) Genome graphs and the evolution of genome inference. *Genome Res.*, **27**, 665–676.

Wick,R.R. *et al.* (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, **31**, 3350–3352.

Yokoyama,T.T. *et al.* (2019) MoMI-G: modular multi-scale integrated genome graph browser. *BMC Bioinformatics*, **20**, 548.

Yue,J.-X. *et al.* (2017) Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, **49**, 913–924.