Bioinformatics, 2025, **41(6)**, btaf332
https://doi.org/10.1093/bioinformatics/btaf332
Advance Access Publication Date: 6 June 2025
Applications Note

OXFORD

Genome analysis

SVbyEye: a visual tool to characterize structural variation among whole-genome assemblies

David Porubsky[1,2,*], Xavi Guitart[1], DongAhn Yoo[1], Philip C. Dishuck[1], William T. Harvey[1], Evan E. Eichler[1,3]

[1]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, United States
[2]Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, 69117, Germany
[3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, United States

*Corresponding author. University of Washington School of Medicine, 3720 15th Ave NE, S413A, Seattle, WA 98195-5065, United States. E-mail: porubsky@uw.edu.

Associate Editor: Janet Kelso

Abstract

Motivation: We are now in the era of being able to routinely generate highly contiguous (near telomere-to-telomere) genome assemblies of human and nonhuman species. Complex structural variation and regions of rapid evolutionary turnover are being discovered for the first time. Thus, efficient and informative visualization tools are needed to evaluate and directly observe structural differences between two or more genomes.

Results: We developed SVbyEye, an open-source R package to visualize and annotate sequence-to-sequence alignments along with various functionalities to process these alignments. The tool facilitates the characterization of complex structural variants in the context of sequence homology helping resolve the mechanisms underlying their formation.

Availability and implementation: SVbyEye is available on GitHub (https://github.com/daewoooo/SVbyEye) and via Zenodo (https://doi.org/10.5281/zenodo.15303553).

1 Introduction

Informative and efficient visualization of genomic structural variation is an important step to evaluate the validity of the most complex regions of the genome, helping us to develop new hypotheses and draw biological conclusions. With advances in long-read sequencing technologies, such as PacBio HiFi (high-fidelity) (Wenger *et al.* 2019), and ONT (Oxford Nanopore Technologies) (Deamer and Branton 2002), we are now able to fully assemble even the most complex regions of the genome, such as segmental duplications (Vollger *et al.* 2022), acrocentric regions (Guarracino *et al.* 2023), and centromeres (Logsdon *et al.* 2024) into continuous, highly accurate linear assemblies—also known as telomere-to-telomere or T2T assemblies (Jarvis *et al.* 2022, Nurk *et al.* 2022). A large part of our understanding of the evolution of complex biological systems comes from comparative analyses, including direct visual observations (Paparella *et al.* 2023, Yoo *et al.* 2025).

The challenge with these analyses is that many large-scale structural changes between genomes are mediated by large, highly identical repeat sequences that are not readily annotated by existing software. This necessitates the development of visualization tools to complement T2T comparative studies. We developed SVbyEye for three purposes: (i) to directly characterize structurally complex regions, including insertions, duplications, deletions and inversions, by comparison to a linear genome reference; (ii) to place these changes in the context of sequence homology by characterizing associated sequence identity; and (iii) to define the breakpoints, including the length and orientation of homologous sequence mediating the rearrangement. SVbyEye is inspired by the previously developed tool called Miropeats (Parsons 1995) and brings its visuals to the popular scripting language R and visualization paradigm using ggplot2 (Wickham 2009).

2 Methods

SVbyEye uses as input DNA sequence alignments in Pairwise mApping Format (PAF), a TAB-delimited text format that can be easily generated with minimap2 (Li 2018). In principle, any sequence-to-sequence aligner that can export alignments to a standard PAF should be sufficient. We note, however, that we tested our tool only using minimap2 alignments. Such alignments can be read using the "readPaf" function. Subsequently, imported alignments can be filtered and flipped into the desired orientation using "filterPaf" and "flipPaf" functions, respectively.

2.1 Visualization modes

There are four visualization modes offered by SVbyEye: visualization of pairwise alignments, alignments of more than two sequences, alignments within a single sequence, and whole-genome alignments. SVbyEye exports all visualizations as ggplot2 objects and thus other ggplot2 plotting functions

Received: 1 October 2024; **Revised:** 2 May 2025; **Editorial Decision:** 28 May 2025; **Accepted:** 4 June 2025

("geoms") and themes can be applied on reported plot objects (Supplementary Notes, available as supplementary data at *Bioinformatics* online).

The main function of this package, called "plotMiro," serves to visualize pairwise sequence alignments in a horizontal layout with the target (reference) sequence at the top and the query at the bottom (Fig. 1A). The user has control over a number of visual and alignment processing features. For instance, users can color sequence alignments by their orientation or percentage of matched bases

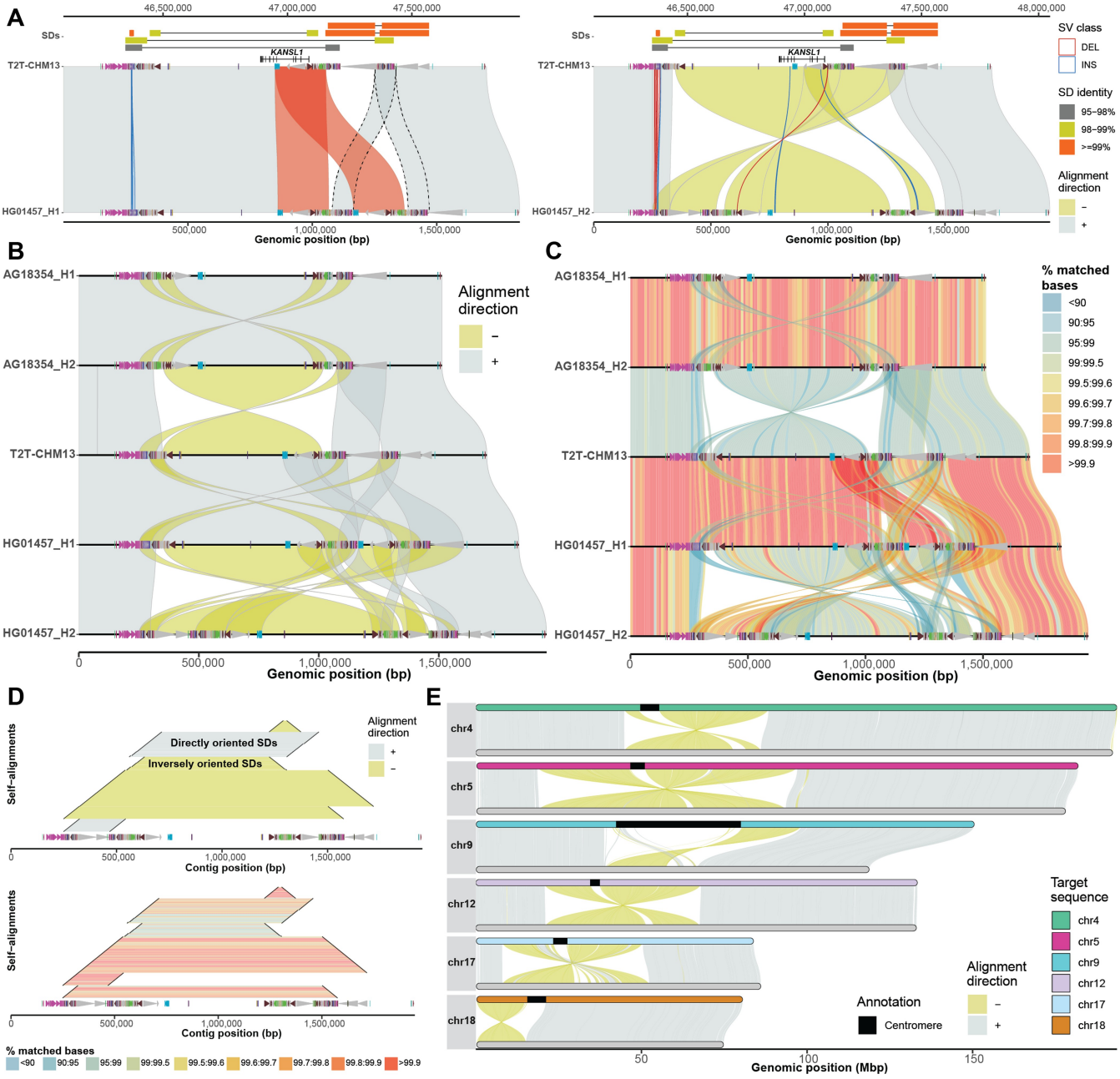

Figure 1. Example of SVbyEye visualization modes. (A) The plot depicts a minimap alignment of a 1.7 Mbp region from chromosome 17q21.31 of two human sequences: HG01457 haplotypes (query) versus T2T-CHM13 reference (target). Segmental duplication (SD) pairwise alignments are shown (top) (connected by horizontal line) colored by their sequence identity with gene annotation (*KANSL1* exons) depicted below as annotated in the UCSC Genome Browser. Minimap2 alignments are shown as polygons between query (bottom) and target (top) sequences colored by the alignment direction (forward "+" direction and reverse "−" direction). Duplicon annotations as defined by DupMasker (Jiang *et al.* 2008) are indicated for both query and target sequences by colored arrowheads pointing forward or backward based on their orientation. A structural variation embedded within the SDs between query and target sequences (≥1 kbp) is highlighted as blue (INS - insertion) and red (DEL - deletion) outlines facilitating breakpoint definition. (B) A "stacked" SVbyEye plot depicting the 17q21.31 region for two chimpanzee haplotypes (AG18354 H1 and H2) followed by three human haplotypes from T2T-CHM13 and HG01457. Each sequence is compared to the sequence immediately above and clearly defines a 750 kbp inversion between chimpanzee and human flanked by inverted repeats. A larger 900 kbp inversion polymorphism is also identified in human mediated by inverted SDs. (C) The plot shows the same alignments as in B but with a "% identity grid" colored by the percentage of matched bases per 10-kbp-long sequence bin. Human inversion shows significant divergence indicating a deeper coalescence of the 17q21.31 region (Zody *et al.* 2008). (D) A "horizontal dotplot" visualization shows self-alignments of HG01457 (haplotype 2) indicating the size (black lines), alignments direction (forward "+" direction and reverse "−" direction; top panel), and pairwise identity (colored grid; bottom panel). The largest and most identical segments are preferred sites for non-allelic homologous recombination (NAHR) breakpoints. (E) A T2T view of six chimpanzee (AG18354) chromosomes (query, bottom) aligned to human syntenic chromosomes (T2T-CHM13, target, top). This view readily defines the extent of paracentric and pericentric inversions.

(Supplementary Notes, available as supplementary data at *Bioinformatics* online).

SVbyEye also allows visualization of alignments of more than two sequences. This can be done by aligning multiple sequences to each other using so-called all-versus-all (AVA) or stacked alignments and submitting them to the "plotAVA" function. In this way, alignments are visualized in subsequent order where the alignment of the first sequence is shown with respect to the second and then second sequence to the third and so on (Fig. 1B). By default, the sequence order is defined by an increasing number of mismatches or can be defined by the user. Many of the same parameters from "plotMiro" also apply to "plotAVA" as well. We illustrate a use of binning PAF alignments into defined bins (by setting a parameter "binsize") and coloring them by the percentage of matched bases—a useful feature to reflect regional or pairwise differences in sequence identity (Fig. 1C).

To accommodate visualization of regions that are homologous to each other within a single sequence, we implemented the "plotSelf" function. This function takes PAF alignments of a sequence to itself and visualizes them in a so-called horizontal dotplot (Fig. 1D). Such visualization can tell us a relative orientation, identity, and size of intrachromosomally aligned regions, an important feature of segmental duplications that predispose intervening sequence to recurrent rearrangements (Itsara *et al.* 2012, Coe *et al.* 2014, Porubsky *et al.* 2022). We note that self-alignments can also be visualized as arcs or arrowed rectangles connecting aligned regions (Supplementary Notes, available as supplementary data at *Bioinformatics* online).

To allow for a full overview of whole-genome assembly with respect to a reference, SVbyEye offers a "plotGenome" function. With this function whole-genome alignments can be visualized to observe large structural rearrangements, such as large para- and pericentromeric inversions between the chimpanzee and human genomes (Fig. 1E).

2.2 Alignment processing and annotation functionalities

SVbyEye has the ability to break PAF alignments at the positions of insertions and deletions and thereby delineate their breakpoints. This is done by parsing alignment CIGAR strings if reported in the PAF file. Thus, by setting the minimum size of insertions and deletions to be reported, one can visualize structural variations as red (deletions) and blue (insertions) outlines (Fig. 1A). For further interrogation, users can also opt to report insertion and/or deletion boundaries in a data table format using the "breakPaf" function.

An important feature of SVbyEye is its capability to annotate query and target sequences with genomic ranges such as gene position, position of segmental duplications, or other DNA functional elements. This is done by adding extra annotation layers on top of the target and/or query alignments using the "addAnnotation" function (Supplementary Notes, available as supplementary data at *Bioinformatics* online). Annotation ranges are visualized as either a rectangle or an arrowhead. Arrowheads are especially useful for conveying an orientation of a genomic range. Similar to PAF alignments, annotation ranges can also be colored by a user-defined color scheme (Fig. 1A). If there is a need to highlight specific PAF alignments between a query and a target, one can do so with the "addAlignments" function that adds selected alignment (s) over the original plot highlighted by a unique outline and/or color (Fig. 1A).

There are several other useful functionalities that come with SVbyEye, for instance, lifting coordinates from target to query and vice versa provided by the "liftRangesToAlignment" function. Users can also subset alignments from a desired region on a target sequence using the "subsetPaf" function. Lastly, there is a possibility to disjoin PAF alignments at regions where two and more alignments overlap each other with the "disjoinPafAlignments" function to provide exact boundaries of duplicated regions (Fig. 1A).

3 Conclusions

We developed SVbyEye, a data visualization R package, to facilitate direct observation of structural differences between two or more sequences. SVbyEye provides several visualization modes depending on the desired application. It offers ample ways to annotate both query and target sequences along with many functionalities to process alignments in PAF.

Acknowledgements

This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a non-exclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions

David Porubsky (Conceptualization [lead], Software [lead], Writing—original draft [lead], Writing—review & editing [equal]), Xavi Guitart (Validation [lead], Writing—review & editing [supporting]), DongAhn Yoo (Validation [supporting], Writing—review & editing [supporting]), Philip C. Dishuck (Validation [supporting], Writing—review & editing [supporting]), William T. Harvey (Validation [supporting], Writing—review & editing [supporting]), and Evan E. Eichler (Conceptualization [supporting], Funding acquisition [lead], Resources [supporting], Supervision [lead], Writing—original draft [supporting], Writing—review & editing [supporting])

Supplementary data

Supplementary data is available at *Bioinformatics* online.

Conflict of interest: E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc.

Funding

This research was supported, in part, by funding from the National Institutes of Health (NIH) [R01 HG002385 and R01 HG010169 to E.E.E.]. E.E.E. is an investigator of the Howard Hughes Medical Institute.

Data availability

The phased genome assembly for human sample HG01457 used in this study is available via NCBI at https://www.ncbi.nlm.nih.gov and can be accessed with the accession number PRJEB76276 (Logsdon et al., 2024).

The phased genome assembly for chimpanzee sample AG18354 used in this study is available via GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/genbank/ and can be accessed with the accession number GCA_028858775.2 (Yoo *et al.* 2025).

References

Coe BP, Witherspoon K, Rosenfeld JA *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 2014;**46**:1063–71.

Deamer DW, Branton D. Characterization of nucleic acids by nanopore analysis. *Acc Chem Res* 2002;**35**:817–25.

Guarracino A, Buonaiuto S, de Lima LG *et al.*; Human Pangenome Reference Consortium. Recombination between heterologous human acrocentric chromosomes. *Nature* 2023;**617**:335–43.

Itsara A, Vissers LELM, Steinberg KM *et al.* Resolving the breakpoints of the 17q21.31 microdeletion syndrome with next-generation sequencing. *Am J Hum Genet* 2012;**90**:599–613.

Jarvis ED, Formenti G, Rhie A *et al.*; Human Pangenome Reference Consortium. Semi-automated assembly of high-quality diploid human reference genomes. *Nature* 2022;**611**:519–31.

Jiang Z, Hubley R, Smit A *et al.* DupMasker: a tool for annotating primate segmental duplications. *Genome Res* 2008;**18**:1362–8.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.

Logsdon GA, Ebert P, Audano PA *et al.* Complex genetic variation in nearly complete human genomes. *bioRxiv* 2024. https://doi.org/10.1101/2024.09.24.614721

Logsdon GA, Rozanski AN, Ryabov F *et al.* The variation and evolution of complete human centromeres. *Nature* 2024;**629**:136–45.

Nurk S, Koren S, Rhie A *et al.* The complete sequence of a human genome. *Science* 2022;**376**:44–53.

Paparella A, L'Abbate A, Palmisano D *et al.* Structural variation evolution at the 15q11-q13 disease-associated locus. *Int J Mol Sci* 2023;**24**:15818.

Parsons JD. Miropeats: graphical DNA sequence comparisons. *Comput Appl Biosci* 1995;**11**:615–9.

Porubsky D, Höps W, Ashraf H *et al.*; Human Genome Structural Variation Consortium (HGSVC). Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* 2022;**185**:1986–2005.e26.

Vollger MR, Guitart X, Dishuck PC *et al.* Segmental duplications and their variation in a complete human genome. *Science* 2022;**376**:eabj6965.

Wenger AM, Peluso P, Rowell WJ *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 2019;**37**:1155–62.

Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer, 2009.

Yoo D, Rhie A, Hebbar P *et al.* Complete sequencing of ape genomes. *Nature* 2025;**641**:401–18.

Zody MC, Jiang Z, Fung H-C *et al.* Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat Genet* 2008;**40**:1076–83.

Bioinformatics, 2025, 41, 1–4
https://doi.org/10.1093/bioinformatics/btaf332
Applications Note