

SOFTWARE

Open Access

# EvoFreq: visualization of the Evolutionary Frequencies of sequence and model data



Chandler D. Gatenbee<sup>1†</sup>, Ryan O. Schenck<sup>1,2†</sup>, Rafael R. Bravo<sup>1</sup> and Alexander R. A. Anderson<sup>1\*</sup> 

## Abstract

**Background:** High throughput sequence data has provided in depth means of molecular characterization of populations. When recorded at numerous time steps, such data can reveal the evolutionary dynamics of the population under study by tracking the changes in genotype frequencies over time. This necessitates a simple and flexible means of visualizing an increasingly complex set of data.

**Results:** Here we offer EvoFreq as a comprehensive tool set to visualize the evolutionary and population frequency dynamics of clones at a single point in time or as population frequencies over time using a variety of informative methods. EvoFreq expands substantially on previous means of visualizing the clonal, temporal dynamics and offers users a range of options for displaying their sequence or model data.

**Conclusions:** EvoFreq, implemented in R with robust user options and few dependencies, offers a high-throughput means of quickly building, and interrogating the temporal dynamics of hereditary information across many systems. EvoFreq is freely available via <https://github.com/MathOnco/EvoFreq>.

**Keywords:** Evolution, Tumor, Visualization software

## Background

Changes in genotype frequencies are often visualized using Muller plots, wherein each polygon represents a genotype (clone), and the thickness of the polygon indicates either the number of individuals with that genotype, or the frequency of the genotype in the total population at each time point. Nesting of genotypes represents evolutionary relationships, i.e. one genotype emerging from within another genotype's polygon indicates that the former was created by an individual with the latter genotype. Muller plots thus provide an excellent way to visualize how the genetic composition of a population changes over time. As these changes are governed by mutation, selection, drift, and gene flow (evolutionary forces), the visualization of these dynamics facilitates an understanding of which forces are dominating the population.

One example where this form of data visualization has proven insightful is in understanding tumor evolution and subclonal compositions from both mechanistic models and bulk or multi-region sequencing. The hereditary nature of mutations in somatic tissue and cancers allows us to recapitulate, the temporal dynamics associated with mutational events through subclonal reconstruction. To date, two visualization packages exist in R to visualize this temporal data fishplot [1] and ggMuller [2]. We constructed an alternative library that includes the features of these two libraries while expanding functionalities (see Table 1). EvoFreq offers users ease of implementation and a way to generate many visualizations for assessing the temporal dynamics of clonal Evolutionary Frequencies. EvoFreq is available at <https://github.com/MathOnco/EvoFreq>.

## Implementation

EvoFreq is built on top of the leading visualization library in R, ggplot2 [3]. Input data can be either clone sizes or frequencies, and in long or wide formats. Given such data,

\*Correspondence: [Alexander.Anderson@moffitt.org](mailto:Alexander.Anderson@moffitt.org)

<sup>†</sup>Chandler D. Gatenbee and Ryan O. Schenck contributed equally to this work.

<sup>1</sup>Integrated Mathematical Oncology, Moffitt Cancer Center, 33612 Tampa, FL, USA

Full list of author information is available at the end of the article



**Table 1** Comparison of the features within the currently available packages for plotting evolutionary dynamics

	ggmuller	fishplot	EvoFreq
Frequency data		✓	✓
Size data	✓		✓
Long format	✓		✓
Wide format		✓	✓
Smoothing		✓	✓
Uses ggplot	✓		✓
Animations			✓
Create graphs			✓
Integration with subclonal reconstruction software		✓	✓

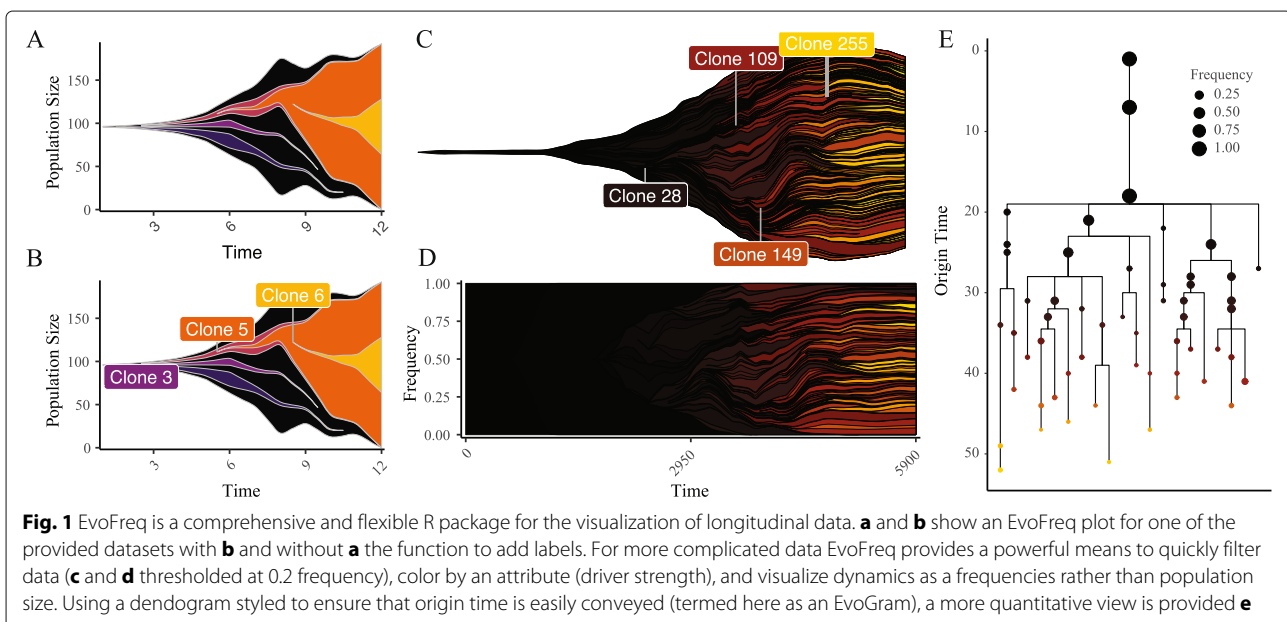
EvoFreq is capable of visualizing the evolutionary dynamics from ClonEvol, PhyloWGS, and CALDER

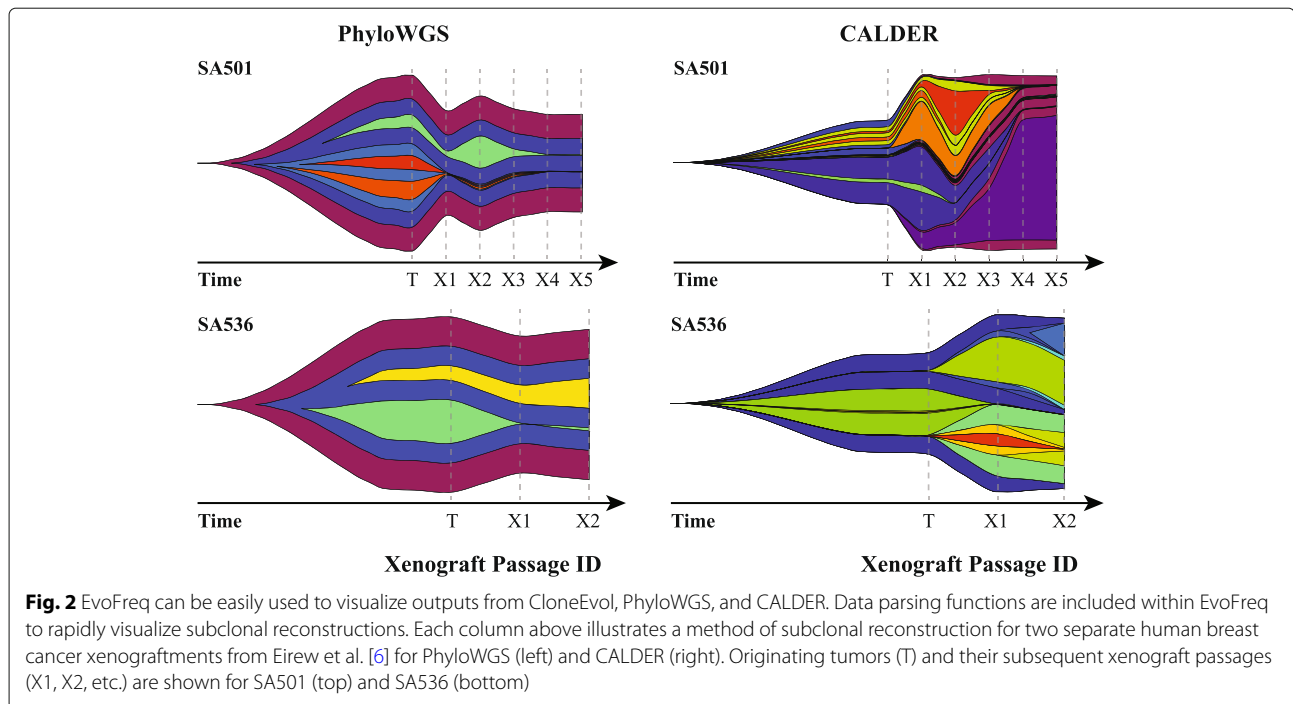
EvoFreq can create Muller plots and/or dendrograms, revealing the clonal dynamics over time. Additional customizations include the ability to color clones based on a user defined attribute (such as fitness), provide custom colors, and label polygons. Due to EvoFreq's utilization of ggplot2 as the primary, underlying library, further customization to EvoFreq's plots is possible. Using the optional dependency [4], users may also create animations of evolving Muller plots and "growing" dendrograms. EvoFreq is capable of visualizing frequency dynamics at a single point in time, as a phylogenetic tree representation, a graph representation, or as a frequency plot over time similar to FishPlot and ggMuller, but with extended capabilities.

## Results

At the core of EvoFreq is the ability to visualize relational data structures over time, whether this is from simulations or inferred from data. Here we illustrate the usefulness of generating these results using EvoFreq. Our first example utilizes data generated by West et al. which quantifies how spatial constraints alter the evolutionary trajectory of a single tumor using a passenger-driver model [5]. EvoFreq was used extensively within this publication and a subset of this data is shown in Fig. 1. This figure highlights how the user can manually or automatically add labels to provide informative details of clones, a particularly useful feature of EvoFreq.

Second, we analyze sequence data using three different clonal reconstruction tools from 15 initial engraftments and serial propagation of primary and metastatic breast cancers [6] and visualize these results using EvoFreq. We applied ClonEvol [7], PhyloWGS [8], and CALDER [9] to infer the clonal dynamics from each of the longitudinal xenografts (select inferences are illustrated in Fig. 2). Initial processing and reformatting of the somatic single nucleotide variant (SNV), copy-number alteration (CNA), and loss of heterozygosity (LOH) data from whole-genome shotgun sequences (WGSS) and Affymetrix SNP Array 6.0 was carried out to extract read data using custom python scripts and prepare inputs for PhyloWGS and CALDER. Each of these tools requires different pre-processing and infers subclonal reconstructions in different ways. We have incorporated functions within EvoFreq to parse outputs of each of these tools to visualize inferred clonal dynamics using EvoFreq. When the output provides numerous solutions for subclonal reconstructions,





the user has an option of returning one or all of these solutions. Examples of this process have also been provided within EvoFreq's documentation.

## Conclusions

We present EvoFreq as a versatile library capable of, generating publication and presentation ready images as well as video for, visualizing clonal frequencies over time. EvoFreq's design allows for broad access to all users through robust support for input data and input validation functions. EvoFreq can be used for all relational data structures providing many different visualizations from one user-facing library, making it applicable in a number of fields. EvoFreq has currently been adopted by a number of research groups, focused on cancer genomics and mechanistic modelling, and has been used in more than four studies already [10–13]. All source code, read me, and issue support is available at <https://github.com/MathOnco/EvoFreq>.

## Availability and requirements

**Project name:** EvoFreq

**Project home page:** <https://github.com/MathOnco/EvoFreq>

**Operating system:** Operating system independent.

**Programming languages:** R

**Other requirements:** ggplot2.

**License:** GNU GPLv3

**Any restrictions to use by non-academics:** None

## Abbreviations

SNV: Single nucleotide variant; CNA: Copy number alteration; LOH: Loss of heterozygosity

## Acknowledgements

Not applicable.

## Authors' contributions

CDG and ROS wrote the code base and packaged EvoFreq for distribution and prepared the manuscript. RRB assisted in the development of algorithms needed for graphical representation of frequencies. ARAA provided guidance on code development, oversaw all work efforts and provided funding. All authors read, edited, and approved the manuscript.

## Funding

The authors gratefully acknowledge funding from both the Cancer Systems Biology Consortium (CSBC) and the Physical Sciences Oncology Network (PSON) at the National Cancer Institute, through grants U01CA232382 (supporting ARAA) and U54CA193489 (supporting CDG, ARAA). ARAA would also like to acknowledge support from the Moffitt Center of Excellence for Evolutionary Therapy. ROS is supported by the Wellcome Trust (grant no. 108861/7/15/7) and the Wellcome Centre for Human Genetics (grant no. 203141/7/16/7). No funding body played a role in the design of the study, analysis and interpretation of data, or in writing the manuscript.

## Availability of data and materials

All data used in this manuscript is freely available at the European Genome-phenome Archive under accession number EGAS00001000952. The code repository can be found at <https://github.com/MathOnco/EvoFreq>.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Integrated Mathematical Oncology, Moffitt Cancer Center, 33612 Tampa, FL, USA. <sup>2</sup>Wellcome Centre for Human Genetics, University of Oxford, OX3 7BN Oxford, UK.

Received: 21 August 2019 Accepted: 22 October 2019

Published online: 16 December 2019

### References

1. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, Mardis ER, Wilson RK. Visualizing tumor evolution with the fishplot package for R. *BMC Genomics*. 2016;17(1):880. <https://doi.org/10.1186/s12864-016-3195-z>.
2. Noble R. Ggmuller: Create Muller Plots of Evolutionary Dynamics. 2018. R package version 0.5.1. <https://CRAN.R-project.org/package=ggmuller>.
3. Wickham H. Ggplot2: Elegant Graphics for Data Analysis: Springer; 2016. <http://ggplot2.org>.
4. Pedersen TL, Robinson D. Gganimate: A Grammar of Animated Graphics. 2019. R package version 1.0.3.9000. <http://github.com/thomasps5/gganimate>.
5. West J, Schenck R, Gatenbee C, Robertson-Tessi M, Anderson ARA. Tissue structure accelerates evolution: premalignant sweeps precede neutral expansion: Cold Spring Harbor Laboratory; 2019. <https://doi.org/10.1101/542019>.
6. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C, Wan A, Laks E, Biele J, Shumansky K, Rosner J, McPherson A, Nielsen C, Roth AJL, Lefebvre C, Bashashati A, de Souza C, Siu C, Aniba R, Brimhall J, Oloumi A, Osako T, Bruna A, Sandoval JL, Algara T, Greenwood W, Leung K, Cheng H, Xue H, Wang Y, Lin D, Mungall AJ, Moore R, Zhao Y, Lorette J, Nguyen L, Huntsman D, Eaves CJ, Hansen C, Marra MA, Caldas C, Shah SP, Aparicio S. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*. 2014;518:422.
7. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, Maher CA. Clonevol: clonal ordering and visualization in cancer sequencing. *Ann Oncol*. 2017;28(12):3076–82. <https://doi.org/10.1093/annonc/mdx517>.
8. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. Phylowgs: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol*. 2015;16(1):35. <https://doi.org/10.1186/s13059-015-0602-8>.
9. Myers MA, Satas G, Raphael BJ. Inferring tumor evolution from longitudinal samples: Cold Spring Harbor Laboratory; 2019. <https://doi.org/10.1101/526814>.
10. West J, You L, Zhang J, Gatenby RA, Brown J, Newton PK, Anderson ARA. Towards multi-drug adaptive therapy: Cold Spring Harbor Laboratory; 2019. <https://doi.org/10.1101/476507>.
11. West J, Schenck RO, Gatenbee C, Robertson-Tessi M, Anderson ARA. Tissue structure accelerates evolution: premalignant sweeps precede neutral expansion: Cold Spring Harbor Laboratory; 2019. <https://doi.org/10.1101/542019>.
12. Schenck RO, Kim E, Bravo RR, West J, Leedham S, Shibata D, Anderson ARA. How homeostasis limits keratinocyte evolution. 2019. <https://doi.org/10.1101/548131>.
13. Gatenbee CD, Baker A-M, Schenck RO, Neves MP, Hasan SY, Martinez P, Cross WC, Jansen M, Rodriguez-Justo M, Sottoriva A, Leedham S, Robertson-Tessi M, Graham TA, Anderson ARA. Niche engineering drives early passage through an immune bottleneck in progression to colorectal cancer. 2019. <https://doi.org/10.1101/623959>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

