The Journal of Open Source Software

Olmsted: Rich Interactive Visualization of B Cell Lineages

David H. Rich [1], **Elias Harkins** [1], **Christopher T. Small**[1], **Kevin Sung** [1], **Laura Doepker** [2], **Duncan Ralph** [1], **Amrit Dhar**[1], **Mackenzie S. Kopp** [2], **Julie Overbaugh** [2], **and Frederick A. Matsen IV** [1¶]

1 Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA **2** Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA ¶ Corresponding author

DOI: N/A

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: 01 January 1970

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Olmsted is an open-source, browser-based application for visually exploring B cell repertoires and clonal family tree data. In the human immune system, affinity maturation of B cell receptor sequences coding for immunoglobulins (antibodies) begins with a diverse pool of randomly generated naive sequences and leads to a collection of evolutionary histories. High-throughput DNA sequencing of the B cell repertoire, combined with computational reconstruction of these evolutionary histories, produces large collections of clonal families and their associated phylogenetic trees. Olmsted enables researchers to scan across collections of clonal families using summary statistics, then interactively explore individual families to visualize phylogenies and amino acid mutations that occurred during affinity maturation.

Statement of Need

Researchers studying B cell responses increasingly rely on high-throughput sequencing to characterize antibody repertoires. Computational tools can reconstruct the evolutionary histories of these sequences, resulting in large collections of clonal families—groups of sequences descended from a common naive ancestor—and phylogenetic trees describing their diversification. However, researchers often lack tools to explore these reconstructions in the detail necessary to choose sequences for functional, structural, or biochemical studies.

Existing visualization tools address different analytical goals. AncesTree (Foglierini et al., 2020) provides detailed single-tree exploration with amino acid mutation display, but requires Java installation, processes one lineage at a time without a repertoire overview, and handles heavy or light chains separately rather than as pairs. ViCloD (Jeusset et al., 2023) focuses on large-scale intraclonal diversity analysis across hundreds of thousands of sequences, primarily for characterizing B cell tumors; it analyzes heavy chains only. AIRRscape (Waltari et al., 2022) enables comparison across multiple repertoires to identify convergent antibody responses at the population level; it also visualizes heavy chains only. ImmuneDB (Rosenfeld et al., 2016) presents collections in paginated list form for database-style querying.

Olmsted combines repertoire-level overview with detailed lineage exploration: researchers can scan across all clonal families in a scatterplot, then drill down to examine phylogenetic trees with aligned amino acid sequences showing mutations from the naive ancestor. This multi-scale navigation, delivered through a zero-installation web application that keeps data client-side, supports practical decisions about which sequences to prioritize for downstream experimental studies. Specifically, Olmsted allows users to:

- View all clonal families simultaneously in a configurable scatterplot

- Filter and select families based on biological criteria (mutation frequency, sequence count, V/J gene usage)
- Examine phylogenetic trees with aligned sequences showing mutations from the naive ancestor
- Trace the mutational history of individual sequences back to their germline origin
- Visualize paired heavy and light chain sequences together, which is essential for selecting antibodies for expression

Olmsted requires no installation—users simply visit olmstedviz.org and upload their data directly in the browser. Data processing is handled by a companion command-line tool, olmsted-cli, which converts common immunoinformatics formats into Olmsted's input format.

Features

Client-Side Data Management

Olmsted requires no installation, account creation, or data upload to external servers. Users simply visit olmstedviz.org and load data via drag-and-drop or file browser. All processing occurs client-side using browser-based storage (IndexedDB), ensuring that sensitive patient data never leaves the researcher's machine. Datasets persist across browser sessions, combining the convenience of a web application with the privacy of local software.

Data Preparation with olmsted-cli

The olmsted-cli command-line tool converts data into Olmsted's JSON format. It supports two input formats:

- **AIRR format**: The JSON-based standard developed by the Adaptive Immune Receptor Repertoire Community (AIRR Community, 2024; Rubelt et al., 2017; Vander Heiden et al., 2018)
- **PCP (Parent-Child Pair) format**: A CSV-based format with explicit parent-child relationships

Example usage:

```
olmsted process -i data.json -o output.json -n "My Dataset"
```

Once this data preparation is done once it creates a file that can be shared with collaborators who can use the web interface only.

Interactive Visualization

The visualization interface consists of four linked sections:

1. **Clonal Families Scatterplot**: Each clonal family appears as a point, with configurable axes, colors, and faceting. Users can brush-select regions or click individual points.
2. **Selected Clonal Families Table**: Displays metadata for selected families, including V/J gene assignments and a visual representation of the recombination event.
3. **Clonal Family Tree**: Shows the phylogenetic tree alongside a sequence alignment. Colors indicate amino acid mutations relative to the naive sequence. The tree supports zooming and panning.
4. **Ancestral Sequences**: For a selected leaf, displays the complete mutational path from the naive sequence.

Implementation

Olmsted is built with React and Redux, with visualizations implemented in Vega and Vega-Lite. The codebase originated as a fork of Nextstrain's Auspice (Hadfield et al., 2018). The application can be deployed as a static single-page application or run locally via Docker.

Acknowledgements

We thank Trevor Bedford, James Hadfield, and other authors of Nextstrain, on which Olmsted's source code is based. We are grateful to the AIRR data representation working group for their cooperative development of the AIRR lineage schema. This work was supported by National Institutes of Health grants R01 AI146028, R01 GM113246, R01 AI120961, R01 AI138709, U19 AI117891, and U19 AI128914, and by the Howard Hughes Medical Institute.

References

AIRR Community. (2024). *AIRR clones and trees schema*. https://docs.airr-community.org/.

Foglierini, M., Pappas, L., Lanzavecchia, A., Corti, D., & Perez, L. (2020). AncesTree: An interactive immunoglobulin lineage tree visualizer. *PLoS Comput. Biol.*, *16*(7), e1007731. https://doi.org/10.1371/journal.pcbi.1007731

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*, *34*(23), 4121–4123. https://doi.org/10.1093/bioinformatics/bty407

Jeusset, L., Abdollahi, N., Verny, T., Armand, M., De Septenville, A. L., Davi, F., & Bernardes, J. S. (2023). ViCloD, an interactive web tool for visualizing B cell repertoires and analyzing intraclonal diversities: Application to human B-cell tumors. *NAR Genom. Bioinform.*, *5*(2), lqad064. https://doi.org/10.1093/nargab/lqad064

Rosenfeld, A. M., Meng, W., Luning Prak, E. T., & Hershberg, U. (2016). ImmuneDB: A system for the analysis and exploration of high-throughput adaptive immune receptor sequencing data. *Bioinformatics*, *33*(2), 292–293. https://doi.org/10.1093/bioinformatics/btw593

Rubelt, F., Busse, C. E., Bukhari, S. A. C., Bürckert, J.-P., Mariotti-Ferrandiz, E., Cowell, L. G., Watson, C. T., Marthandan, N., Faison, W. J., Hershberg, U., Laserson, U., Corrie, B. D., Davis, M. M., Peters, B., Lefranc, M.-P., Scott, J. K., Breden, F., AIRR Community, Luning Prak, E. T., & Kleinstein, S. H. (2017). Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. *Nature Immunology*, *18*(12), 1274–1278. https://doi.org/10.1038/ni.3873

Vander Heiden, J. A., Marquez, S., Marthandan, N., Bukhari, S. A. C., Busse, C. E., Corrie, B., Hershberg, U., Kleinstein, S. H., Matsen IV, F. A., Ralph, D. K., Rosenfeld, A. M., Schramm, C. A., AIRR Community, Christley, S., & Laserson, U. (2018). AIRR community standardized representations for annotated immune repertoires. *Frontiers in Immunology*, *9*, 2206. https://doi.org/10.3389/fimmu.2018.02206

Waltari, E., Nafees, S., McCutcheon, K. M., Wong, J., & Pak, J. E. (2022). AIRRscape: An interactive tool for exploring B-cell receptor repertoires and antibody responses. *PLoS Comput. Biol.*, *18*(9), e1010052. https://doi.org/10.1371/journal.pcbi.1010052