

## RESEARCH ARTICLE

## AncesTree: An interactive immunoglobulin lineage tree visualizer

Mathilde Foglierini<sup>1,2\*</sup>, Leontios Pappas<sup>1</sup>, Antonio Lanzavecchia<sup>1</sup>, Davide Corti<sup>3</sup>, Laurent Perez<sup>1</sup>

**1** Università della Svizzera italiana, Faculty of Biomedical Sciences, Institute for Research in Biomedicine, Bellinzona, Switzerland, **2** Swiss Institute of Bioinformatics, Lausanne, Switzerland, **3** Humabs Biomed SA, Vir Biotechnology, Bellinzona, Switzerland

\* [mathilde.perez@irb.usi.ch](mailto:mathilde.perez@irb.usi.ch)

## Abstract

High-throughput sequencing of human immunoglobulin genes allows analysis of antibody repertoires and the reconstruction of clonal lineage evolution. The study of antibodies (Abs) affinity maturation is of specific interest to understand the generation of Abs with high affinity or broadly neutralizing activities. Moreover, phylogenetic analysis enables the identification of the key somatic mutations required to achieve optimal antigen binding. The Immcantation framework provides a start-to-finish set of analytical methods for high-throughput adaptive immune receptor repertoire sequencing (AIRR-Seq; Rep-Seq) data. Furthermore, Immcantation's Change-O package has developed IgPhyML, an algorithm designed to build specifically immunoglobulin (Ig) phylogenetic trees. Meanwhile Phylip, an algorithm that has been originally developed for applications in ecology and macroevolution, can also be used for the phylogenetic reconstruction of antibodies maturation pathway. To complement Ig lineages made by IgPhyML or Dnaml (Phylip), we developed AncesTree, a graphic user interface (GUI) that aims to give researchers the opportunity to interactively explore antibodies clonal evolution. AncesTree displays interactive immunoglobulins phylogenetic tree, Ig related mutations and sequence alignments using additional information coming from specialized antibody tools. The GUI is a Java standalone application allowing interaction with Ig tree that can run under Windows, Linux and Mac OS.

## OPEN ACCESS

**Citation:** 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>

**Editor:** Mihaela Pertea, Johns Hopkins University, UNITED STATES

**Received:** February 13, 2020

**Accepted:** June 14, 2020

**Published:** July 10, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1007731>

**Copyright:** © 2020 Foglierini et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Software, documentation, source code and examples are available at <https://github.com/MathildeFogPerez/ancesree>

This is a *PLOS Computational Biology* Software paper.

## Introduction

Development of Next Generation Sequencing (NGS) methodology and its use for high-throughput sequencing of the Adaptive Immune Receptor Repertoire (AIRR-seq) has provided unprecedented molecular insight into the complexity of the humoral adaptive immune response by generating Ig data sets of 100 million to billions of reads. Different computational methods have been developed to exploit and analyze these data [1]. Retracing the antigen-driven evolution of Ig repertoires by inferring antibody evolution lineages is a powerful

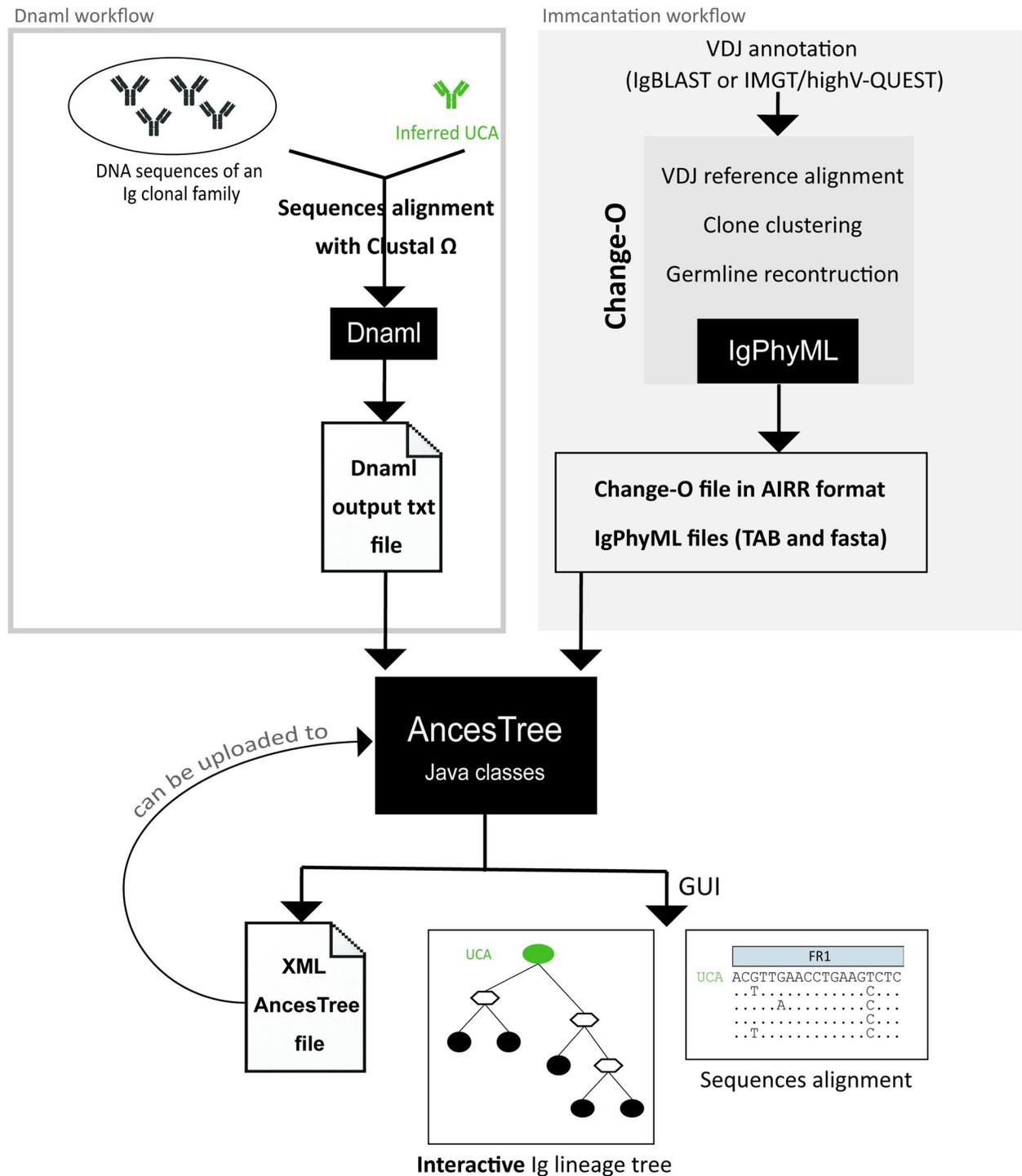
**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

method to understand how vaccines or pathogens shape the humoral immune response [2–5]. Indeed, Abs maturation is the result of clonal selection during B cell expansion and a clonal lineage is defined as immunoglobulin sequences originating from the same recombination event occurring between the V, D and J segments [6]. B cell receptor (BCR) engagement by a given antigen will trigger somatic hypermutations (SHMs) events generating a large BCR diversity. This process leads to antibodies with mutated Ig variable regions, thus forming a specific B-cell lineage that extends from the naive unmutated B-cells, to somatically hypermutated and class switched memory B or plasma-cells [7]. Lineage tree building requires a common preprocessing step, the clonal lineage assignment [8]. A common starting approach is to initially cluster sequences by their V and J genes and by their CDR3 length. Commonly used tools capable of aligning Ig sequences are MiXCR, IMGT, IgBlast, SONAR, IGoR and iHM-Munealign [9–13]. One of the major drawbacks of the previously mentioned tools is the reliance of the initial alignment with the germline and the exclusion of insertion/deletions (indels) events in the lineage. To circumvent those problems other methods were developed: i) Partis and SONAR [14, 15] can perform both unseeded and seeded lineage assignment ii) Clonify, using a hierarchical clustering approach, performs unseeded lineage assignment [16]. Nonetheless, there is no consensus as to which phylogenetic method is optimal to infer the ancestral evolutionary relationships among Ig sequences [17, 18]. Literally, several methods have been used, such as Levenshtein distance (LD), neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BEAST) [19–22]. The DNA Maximum Likelihood program (Dnaml) of the PHYLIP package [23], is a ML method that has been originally developed for applications in ecology. It is also commonly used to infer B cell clonal lineages [24–29]. Visualization of the phylogeny is performed using Dendroscope [30, 31]. Meanwhile, a framework was developed to provide a start-to-finish toolbox to process high-throughput AIRR-seq datasets. The Immcantation framework (<https://immcantation.readthedocs.io/en/stable/>) is currently the gold standard for antibody repertoire analysis. The Change-O tool [32], which is part of Immcantation, was developed to make i) a V(D)J reference alignment standardization after sequences annotation by IMGT/High-VQUEST [33] or IgBlast ii) clonal clustering iii) germline reconstruction iv) conversion and annotation. The IgPhyML algorithm, which is part of Change-O, allows the reconstruction of phylogenetic tree by implementing substitution models that correct for the context-sensitive nature of SHM, and combines information from multiple lineages to give more precisely estimated repertoire-wide model parameter estimates. Currently, there is no efficient bioinformatics tool allowing an interactive display of phylogenetic tree inferred from Ig sequences. Here we developed AncesTree, an Ig lineage tree visualizer that also integrates information coming from most used antibody bioinformatics tools: IgBlast, IMGT, Change-O, Kabat numbering [34] and BASELINE [35]. AncesTree enables users to interact with a tree containing up to thousands Ig sequences, which were generated by Dnaml or IgPhyML, via the GUI. It is a standalone application that is platform independent and only need JAVA JRE 12 or higher as prerequisite software installed.

## Design and implementation

The AncesTree workflow is presented in **Fig 1**, it consists of three different main steps: Input, Processing and Outputs. Importantly, phylogenetic tree analyses coming from two different tools can be used by AncesTree (Dnaml or Immcantation). If the Dnaml workflow is used, AncesTree will parse the Dnaml output text file. If the Immcantation workflow is used (i.e. RepSeq data), the Change-O tab file in AIRR format, the IgPhyML tab file and its related fasta file (with the reconstructed intermediate sequences) are used as input. Once AncesTree



**Fig 1. AncesTree workflow.** Dnaml workflow is used (left part of the figure), DNA sequences are aligned with Clustal  $\Omega$  and processed by Dnaml, a phylogenetic tree is generated. If the Immccantation workflow is used (right part of the figure), reads are processed through Change-O pipeline and IgPhyML will generate the trees. AncesTree processes the different inputs and reconstructs the phylogenetic tree with all information related to Ig. The tree is displayed in a GUI and an Extensible Markup Language (XML) file is produced (that could be used as direct input into AncesTree).

<https://doi.org/10.1371/journal.pcbi.1007731.g001>

processes the input file(s), it will generate a tree in a graphic interface to allow direct interactivity. Features specific for Ig analysis are included in the GUI.

## Input

AncesTree has two possibilities to display an Ig lineage tree: Dnaml workflow or Immcantation workflow (Fig 1). The first one is the Dnaml workflow. The required input for AncesTree usage is the output text file generated by Dnaml. Optionally, a fasta file with data obtained from IMGT can also be used to have full AncesTree features. A clonal family is composed of heavy (or light) V(D)J sequences and their related unmutated common ancestor (UCA). The UCA can be inferred with Antigen Receptor Probabilistic Parser (ARPP) UA Inference software [36] or Cloanalyst [37]. Then, sequences are aligned with Clustal  $\Omega$  [38] and the generated file in PHYLIP format can be provided for Dnaml. Next, Dnaml is launched with the following settings: 'O' for the outgroup root with the number corresponding to the UCA position provided in the PHYLIP input text file and '5' to reconstruct hypothetical sequences. The generated 'outfile' text file can be used as input for AncesTree. To visualize the different frameworks (FW) and complementary-determining (CDR) regions that composed the Ig variable region, a fasta file can be uploaded. The user provides a fasta file containing the following information: the UCA V(D)J sequence in IMGT format including gaps, and the end positions of each region included in the fasta identifier (separated by a space). This information is easily retrieved using IMGT/V-QUEST [33] with the UCA nucleotide sequence as input. The second possibility for AncesTree to display an Ig lineage tree is through the Immcantation workflow. In most of the cases, this workflow is applied to RepSeq data coming from NGS experiments, but it can also be applied to a small set of sequences. Change-O is processed to run the following steps: i) align to V(D)J reference (sequences are annotated by IMGT/High-VQUEST or IgBlast prior to this step), ii) filter in the productive sequences, cluster sequences by clone, reconstruct germline by clone and iii) convert the change-O file into AIRR format [39]. To infer phylogenetic trees with the clones of interest, IgPhyML is run with the '-asr' option, allowing to reconstruct the intermediate sequences of the tree. Of note, this option is only available with the docker image develop version of Immcantation. AncesTree takes as input the Change-O file in AIRR format, the related IgPhyML tab file and the fasta IgPhyML file produced by the '-asr' option. Finally, the user has to specify with a drop down menu which clone id (from the IgPhyML tab file) he wants to be displayed into AncesTree GUI.

## Processing

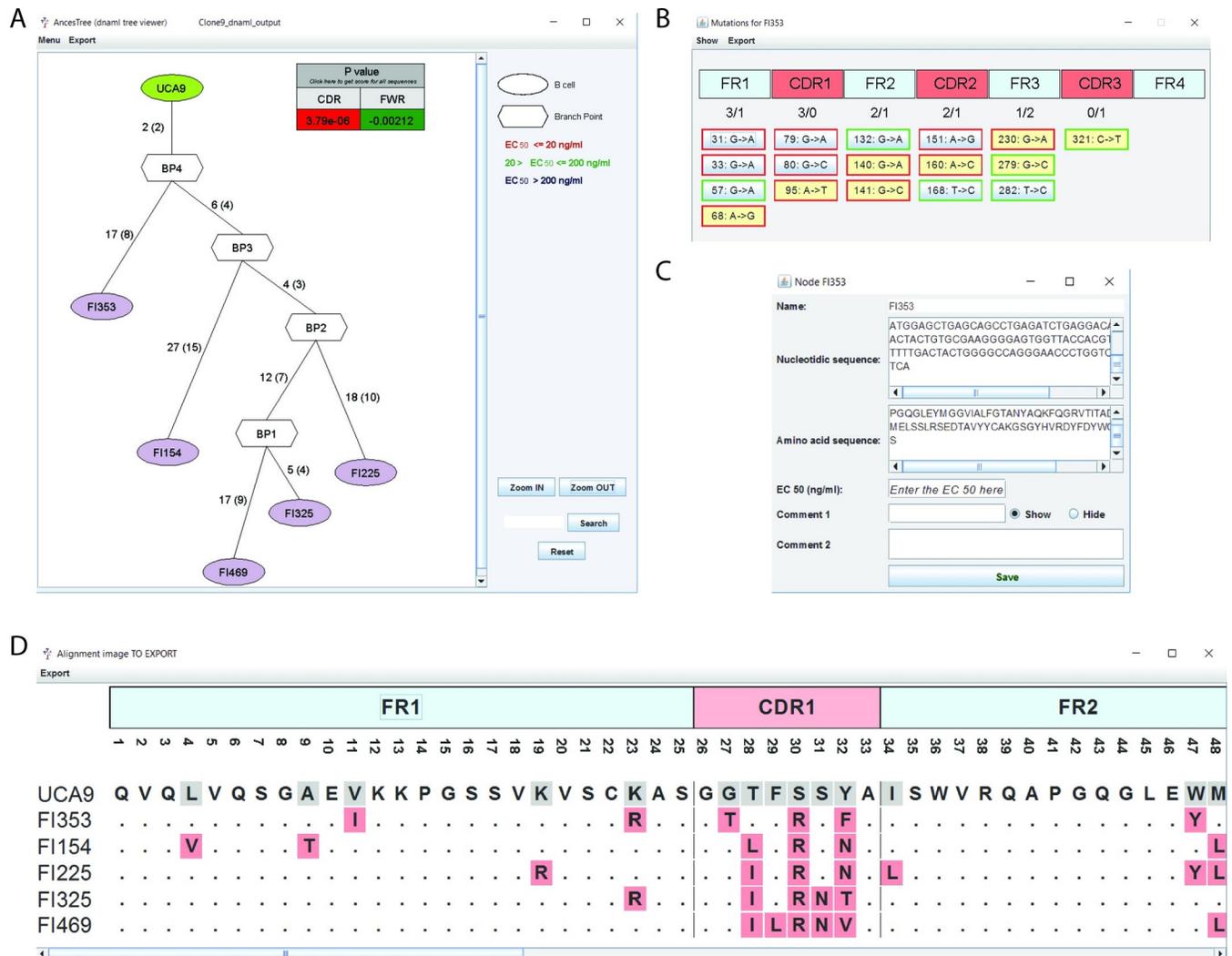
AncesTree parses the Dnaml output file or the Change-O and IgPhyML files. The Dnaml file is a text file and does not require a tree in Newick format. Naturally, the relationship between the different nodes of the tree is already stored, in addition to the sequence of each node in the Dnaml output text file. Conversely, in the case of IgPhyML input, the tree is reconstructed from a newick format. The theoretical intermediate reconstructed sequences are renamed branch points (BPs). In cases of ambiguous nucleotide notation (IUPAC nomenclature) with the Dnaml input file, AncesTree selects the nucleotide with the highest probability based on the Ig sequences retrieved after this BP. Of note, IgPhyML already makes this correction with the '-asr' option set to 0.1. AncesTree has the ability to collapse a node if the sequences are identical, for example in the case of a theoretical BP corresponding to an existing Ig. Moreover, AncesTree will also draw different nodes clustered together in the case of identical Ig sequences, thus providing a clear topology view of the tree.

## Outputs

After running AncesTree, a sub-folder is automatically created in the 'output' folder using the name of the Dnaml output file or the name of the IgPhyML file with the selected clone id. The

resulting folder will contain all produced files such as a XML file that can be used for direct loading into the GUI.

AncesTree displays the processed tree in the main panel of the GUI (Fig 2A). The number of nucleotide and amino acid mutations are written on the edge between each node/sequence (with amino acid mutations shown in parenthesis) and it is clickable, enabling the opening of a new window frame that displays the detailed location of each mutation (Fig 2B). Of note, the color of the box around each mutated codon indicates whether the mutation is replacement (R) in red or silent (S) in green. This information is also available as R/S numbers under each region. The user can view the amino acid (a.a.) mutations, and have access by default to the Kabat numbering of the related a.a. position (without internet access, AncesTree will use the a. a. absolute position). To obtain the nucleotide or protein sequence of a node, the user can click on it (Fig 2C). The user also has the possibility to enter the EC<sub>50</sub> for the specified Ig. The sequence alignments (DNA or protein) are accessible in a new frame via the ‘Menu’ button on



**Fig 2. Snapshot of AncesTree GUI.** (A) The tree generated by Dnaml is displayed in the main panel. The BASELINE analysis for the clonal family is displayed in the right upper corner. (B) The mutations between two nodes can be displayed in a separate window and they are positioned using IMGT sequence annotation. (C) The user can have access to each specific node to obtain the related sequences (DNA or protein) and add comments. (D) An alignment is generated with the UCA appearing in the first lane, and a ruler indicates the different regions that compose an Ig sequence.

<https://doi.org/10.1371/journal.pcbi.1007731.g002>

the top (Fig 2D). Finally, the alignment view is customizable: the sequences can be selected or deselected, as well as different positions or regions. Different color modes can be chosen.

If the user is interested in a BASELINE analysis of a clonal family of interest, and if the optional input fasta file (with the UCA VDJ sequence including gaps) was provided with the Dnaml input, AncesTree will automatically generate the fasta input file needed for this software (<http://selection.med.yale.edu/baseline/>). Once BASELINE has finished to process, the output file can be loaded into AncesTree to have a nice graphic view of antigen-driven selection occurring for this particular clonal family. All graph generated can be exported in PNG or EPS format and the alignment can be exported as Tab-separated Values (TSV) file.

## Results

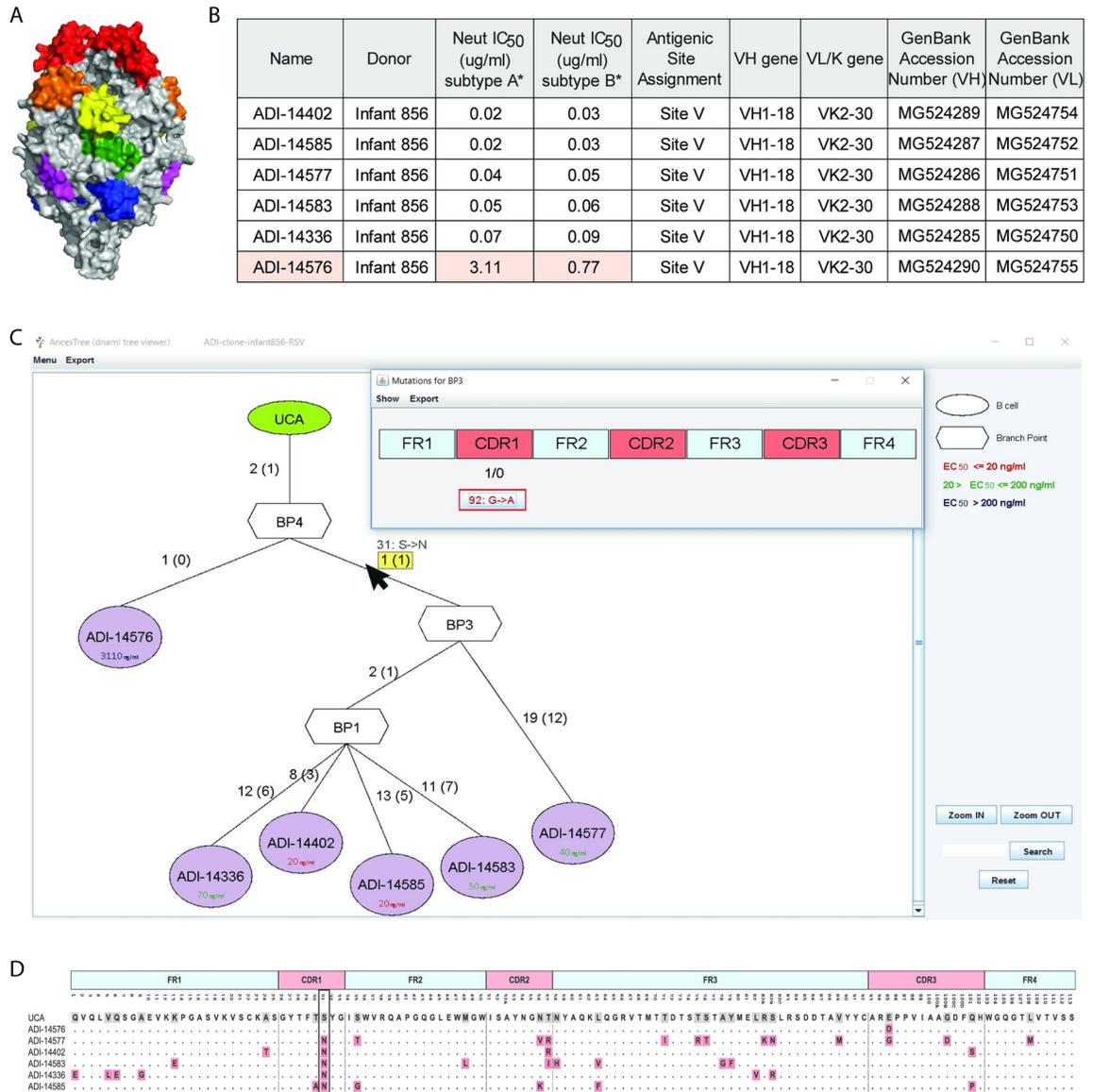
To demonstrate the utility of AncesTree we analyzed a case study by performing the analysis of an Ig lineage tree targeting the fusion protein (F) of the Respiratory Syncytial Virus (RSV). RSV is an enveloped RNA virus belonging to the recently defined *Pneumoviridae* family [40]. Infection of healthy adults by RSV typically results in mild respiratory symptoms. However, viral infection of infants and older adults, accounts for a substantial hospitalization burden in both age groups [41]. Indeed, RSV infection is the second cause of infant mortality worldwide after malaria [42]. Understanding the immunological basis for the development of potent neutralizing antibodies is a key step for the development of an effective vaccine for RSV.

### Case study: Exploration of Ig lineage targeting the Fusion protein of the Respiratory Syncytial Virus (F-RSV)

To demonstrate the practical use of AncesTree, we re-analyzed an Ig dataset generated post infection by Respiratory Syncytial Viral infection (HRSV). The dataset was collected by isolating antibodies direct against the F-RSV protein, a class I fusion protein mediating viral entry into host cells [43]. The Ig sequences were clustered by grouping antibodies sharing the same VH and VL gene usage, HCDR3 length and identity (at least 85% for HCDR3). Among the clusters generated, we chose Igs targeting the antigenic site V of F-RSV located near amino acid 447 between the  $\alpha 3$  helix and  $\beta 3/\beta 4$  hairpin of F-RSV in prefusion (Fig 3A). About 70% of the mAbs targeting this site use the same VH and VL germline pair (VH1–18 and VK2–30) [43–45]. We identified an Ig family of interest containing potent neutralizers targeting site V with one outlier, the mAb ADI-14576, being less potent and with a 10-fold decrease in binding affinity (Fig 3B). We used Dnaml to generate a VH sequences phylogenetic tree and launched AncesTree to analyze and interact with the produced phylogenetic tree (Fig 3C). The  $EC_{50}$  (ng/ml) related to the neutralization assay against RSV subtype A is reported in each node (of note,  $EC_{50}$  against subtype B is in the same range for each Ig). Surprisingly, a common mutation 92: G->A (kabat position 31: S->N) is shared between all the Igs, except for ADI-14576 that does not share this mutation. The alignment of the Ig protein sequences highlights clearly this shared mutation (Fig 3D). A result suggesting that ADI-14576 underwent less affinity maturation and therefore diverges from all the other family members. Interestingly, the 31: S->N mutation is located in the HCDR1 and asparagine residues are often involved in protein binding sites. It is tempting to speculate that the Serine to Asparagine substitution is in part responsible for the higher potency and binding titer of the antibodies.

### Concluding remarks

To summarize, we developed an intuitive, easy and interactive GUI allowing the visualization and exploration of antibody clonal evolution. Our application is open access and platform independent. AncesTree only needs the file(s) that can be produced either by Dnaml or by



**Fig 3. Clonal family against F-RSV protein antigenic site V.** (A) Shown is the prefusion conformation of F-RSV trimer (PDB ID: 4MMU) [46]. The antigenic sites are colored, site Ø (red), I (blue), II (yellow), III (green), IV (purple) and V (orange). (B) Table showing the different characteristic of a mAbs clonal family isolated from an infant (≥ 6 months) after RSV infection. The Igs neutralization titers are shown as well as their related Germline annotations. ADI-14576 is highlighted because of its lower neutralization value in comparison to the other mAbs of the same clonal family. Phylogenetic analysis of the VH chain of a clonal family F-RSV specific. (C) Phylogenetic tree displayed in Ancestree where the user clicked on the mutation shared by all Igs below BP3 node (31: S->N). (D) Protein alignment of the different Ig sequences, the mutation 31: S->N is boxed.

<https://doi.org/10.1371/journal.pcbi.1007731.g003>

IgPhyML, which is one of the most used tool for antibody repertoire analysis. In the latter, Ancestree processed the Change-O file in AIRR format that allows to share RepSeq data in a standardized format. The possibility to visualize tree independently of the used pipeline allows a broad Ancestree usage. Ancestree was successfully tested with phylogenetic trees up to 500 unique sequences processed by IgPhyML and more than a thousand sequences processed by Dnaml. Our interface will provide the users a practical tool containing several useful features that will be of high utility for the immunologists' community and especially those with little or no computational skills.

## Acknowledgments

The authors acknowledge present and past members of the Lanzavecchia's group for comments and feedback on the software.

## Author Contributions

**Conceptualization:** Mathilde Foglierini.

**Data curation:** Mathilde Foglierini.

**Formal analysis:** Mathilde Foglierini, Leontios Pappas.

**Investigation:** Mathilde Foglierini, Laurent Perez.

**Methodology:** Mathilde Foglierini.

**Project administration:** Antonio Lanzavecchia.

**Resources:** Antonio Lanzavecchia.

**Software:** Mathilde Foglierini.

**Supervision:** Antonio Lanzavecchia, Davide Corti, Laurent Perez.

**Validation:** Leontios Pappas.

**Visualization:** Mathilde Foglierini.

**Writing – original draft:** Mathilde Foglierini, Laurent Perez.

**Writing – review & editing:** Mathilde Foglierini, Davide Corti, Laurent Perez.

## References

1. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. *Frontiers in Immunology*. 2018; 9(224).
2. Wang C, Liu Y, Cavanagh MM, Le Saux S, Qi Q, Roskin KM, et al. B-cell repertoire responses to varicella-zoster vaccination in human identical twins. *Proc Natl Acad Sci U S A*. 2015; 112(2):500–5. <https://doi.org/10.1073/pnas.1415875112> PMID: 25535378
3. Hoehn KB, Fowler A, Lunter G, Pybus OG. The Diversity and Molecular Evolution of B-Cell Receptors during Infection. *Mol Biol Evol*. 2016; 33(5):1147–57. <https://doi.org/10.1093/molbev/msw015> PMID: 26802217
4. Zhu J, Ofek G, Yang Y, Zhang B, Louder MK, Lu G, et al. Mining the antibodyome for HIV-1–neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proceedings of the National Academy of Sciences*. 2013; 110(16):6470–5.
5. Jackson KJ, Liu Y, Roskin KM, Glanville J, Hoh RA, Seo K, et al. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell host & microbe*. 2014; 16(1):105–14.
6. Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983; 302(5909):575–81. <https://doi.org/10.1038/302575a0> PMID: 6300689
7. Adler R. Janeway's immunobiology. *Choice: Current Reviews for Academic Libraries*. 2008; 45(10):1793–4.
8. Yermanos AD, Dounas AK, Stadler T, Oxenius A, Reddy ST. Tracing Antibody Repertoire Evolution by Systems Phylogeny. *Frontiers in Immunology*. 2018; 9(2149).
9. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nature methods*. 2015; 12(5):380–1. <https://doi.org/10.1038/nmeth.3364> PMID: 25924071
10. Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res*. 2015; 43(Database issue):D413–22. <https://doi.org/10.1093/nar/gku1056> PMID: 25378316

11. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* 2013; 41(Web Server issue):W34–40. <https://doi.org/10.1093/nar/gkt382> PMID: 23671333
12. Marcou Q, Mora T, Walczak AM. High-throughput immune repertoire analysis with IGoR. *Nat Commun.* 2018; 9(1):561. <https://doi.org/10.1038/s41467-018-02832-w> PMID: 29422654
13. Gaeta BA, Malming HR, Jackson KJ, Bain ME, Wilson P, Collins AM. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics.* 2007; 23(13):1580–7. <https://doi.org/10.1093/bioinformatics/btm147> PMID: 17463026
14. Ralph DK, Matsen FA. Likelihood-Based Inference of B Cell Clonal Families. *PLoS Comput Biol.* 2016; 12(10):e1005086. <https://doi.org/10.1371/journal.pcbi.1005086> PMID: 27749910
15. Schramm CA, Sheng Z, Zhang Z, Mascola JR, Kwong PD, Shapiro L. SONAR: A High-Throughput Pipeline for Inferring Antibody Ontogenies from Longitudinal Sequencing of B Cell Transcripts. *Front Immunol.* 2016; 7:372. <https://doi.org/10.3389/fimmu.2016.00372> PMID: 27708645
16. Briney B, Le K, Zhu J, Burton DR. Clonify: unseeded antibody lineage assignment from next-generation sequencing data. *Scientific Reports.* 2016; 6(1):23901.
17. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends in immunology.* 2015; 36(11):738–49. <https://doi.org/10.1016/j.it.2015.09.006> PMID: 26508293
18. Yaari G, Vander Heiden JA, Uduman M, Gadala-Maria D, Gupta N, Stern JN, et al. Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. *Front Immunol.* 2013; 4:358. <https://doi.org/10.3389/fimmu.2013.00358> PMID: 24298272
19. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med.* 2014; 6(248):248ra107.
20. Barak M, Zuckerman NS, Edelman H, Unger R, Mehr R. IgTree: creating Immunoglobulin variable region gene lineage trees. *J Immunol Methods.* 2008; 338(1–2):67–74. <https://doi.org/10.1016/j.jim.2008.06.006> PMID: 18706908
21. Andrews SF, Kaur K, Pauli NT, Huang M, Huang Y, Wilson PC. High preexisting serological antibody levels correlate with diversification of the influenza vaccine response. *J Virol.* 2015; 89(6):3308–17. <https://doi.org/10.1128/JVI.02871-14> PMID: 25589639
22. Wu X, Zhang Z, Schramm CA, Joyce MG, Kwon YD, Zhou T, et al. Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection. *Cell.* 2015; 161(3):470–85. <https://doi.org/10.1016/j.cell.2015.03.004> PMID: 25865483
23. Felsenstein J. PHYLIP—phylogeny inference package (version 3.2). *Cladistics.* 1989; 5:164–6.
24. Liao HX, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature.* 2013; 496(7446):469–76. <https://doi.org/10.1038/nature12053> PMID: 23552890
25. Pappas L, Foglierini M, Piccoli L, Kallewaard NL, Turrini F, Silacci C, et al. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature.* 2014; 516(7531):418–22. <https://doi.org/10.1038/nature13764> PMID: 25296253
26. Kallewaard NL, Corti D, Collins PJ, Neu U, McAuliffe JM, Benjamin E, et al. Structure and Function Analysis of an Antibody Recognizing All Influenza A Subtypes. *Cell.* 2016; 166(3):596–608. <https://doi.org/10.1016/j.cell.2016.05.073> PMID: 27453466
27. Tan J, Pieper K, Piccoli L, Abdi A, Perez MF, Geiger R, et al. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature.* 2016; 529(7584):105–9. <https://doi.org/10.1038/nature16450> PMID: 26700814
28. Di Niro R, Lee SJ, Vander Heiden JA, Elsner RA, Trivedi N, Bannock JM, et al. Salmonella Infection Drives Promiscuous B Cell Activation Followed by Extrafollicular Affinity Maturation. *Immunity.* 2015; 43(1):120–31. <https://doi.org/10.1016/j.immuni.2015.06.013> PMID: 26187411
29. Revell LJ, Chamberlain SA. Rphylip: an R interface for PHYLIP. *Methods in Ecology and Evolution.* 2014; 5(9):976–81.
30. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics.* 2007; 8:460. <https://doi.org/10.1186/1471-2105-8-460> PMID: 18034891
31. Wu X, Zhang Z, Schramm CA, Joyce MG, Kwon YD, Zhou T, et al. Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection. *Cell.* 2015; 161(3):470–85. <https://doi.org/10.1016/j.cell.2015.03.004> PMID: 25865483

32. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*. 2015; 31(20):3356–8. <https://doi.org/10.1093/bioinformatics/btv359> PMID: 26069265
33. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*. 2008; 36(Web Server issue):W503–8. <https://doi.org/10.1093/nar/gkn316> PMID: 18503082
34. Abhinandan KR, Martin AC. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. *Mol Immunol*. 2008; 45(14):3832–9. <https://doi.org/10.1016/j.molimm.2008.05.022> PMID: 18614234
35. Yaari G, Uduman M, Kleinstein SH. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. *Nucleic Acids Res*. 2012; 40(17):e134. <https://doi.org/10.1093/nar/gks457> PMID: 22641856
36. Kepler TB. Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000Res*. 2013; 2:103. <https://doi.org/10.12688/f1000research.2-103.v1> PMID: 24555054
37. Kepler TB, Munshaw S, Wiehe K, Zhang R, Yu JS, Woods CW, et al. Reconstructing a B-Cell Clonal Lineage. II. Mutation, Selection, and Affinity Maturation. *Front Immunol*. 2014; 5:170. <https://doi.org/10.3389/fimmu.2014.00170> PMID: 24795717
38. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics*. 2014; 48:3.13.1–6.
39. Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, et al. Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data. *Frontiers in Immunology*. 2017; 8(1418).
40. Collins PL, Fearn R, Graham BS. Respiratory syncytial virus: virology, reverse genetics, and pathogenesis of disease. *Curr Top Microbiol Immunol*. 2013; 372:3–38. [https://doi.org/10.1007/978-3-642-38919-1\\_1](https://doi.org/10.1007/978-3-642-38919-1_1) PMID: 24362682
41. Widmer K, Griffin MR, Zhu Y, Williams JV, Talbot HK. Respiratory syncytial virus- and human metapneumovirus-associated emergency department and hospital burden in adults. *Influenza Other Respir Viruses*. 2014; 8(3):347–52. <https://doi.org/10.1111/irv.12234> PMID: 24512531
42. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859):2095–128. [https://doi.org/10.1016/S0140-6736\(12\)61728-0](https://doi.org/10.1016/S0140-6736(12)61728-0) PMID: 23245604
43. Goodwin E, Gilman MSA, Wrapp D, Chen M, Ngwuta JO, Moin SM, et al. Infants Infected with Respiratory Syncytial Virus Generate Potent Neutralizing Antibodies that Lack Somatic Hypermutation. *Immunity*. 2018; 48(2):339–49. e5.
44. Gilman MS, Castellanos CA, Chen M, Ngwuta JO, Goodwin E, Moin SM, et al. Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors. *Science immunology*. 2016; 1(6).
45. Mousa JJ, Kose N, Matta P, Gilchuk P, Crowe JE Jr. A novel pre-fusion conformation-specific neutralizing epitope on the respiratory syncytial virus fusion protein. *Nat Microbiol*. 2017; 2:16271. <https://doi.org/10.1038/nmicrobiol.2016.271> PMID: 28134924
46. McLellan JS, Chen M, Joyce MG, Sastry M, Stewart-Jones GB, Yang Y, et al. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science*. 2013; 342(6158):592–8. <https://doi.org/10.1126/science.1243283> PMID: 24179220