ViCloD, an interactive web tool for visualizing B cell repertoires and analyzing intraclonal diversities: application to human B-cell tumors

Lucile Jeusset[1,2], Nika Abdollahi[1,3], Thibaud Verny[1,4], Marine Armand[2], Anne Langlois De Septenville[2], Frédéric Davi[2,*] and Juliana Silva Bernardes [1,*]

[1]Sorbonne Université, CNRS, UMR 7238, Laboratoire de Biologie Computationnelle et Quantitative, Paris, France, [2]Sorbonne Université, AP-HP, Hôpital Pitié-Salpêtrière, Department of Biological Hematology, Paris, France, [3]IMGT®, the international ImMunoGeneTics Information System, CNRS, Institute of Human Genetics, Montpellier University, France and [4]Ecole des Mines ParisTech, Paris, France

ABSTRACT

High throughput sequencing of adaptive immune receptor repertoire (AIRR-seq) has provided numerous human immunoglobulin (IG) sequences allowing specific B cell receptor (BCR) studies such as the antigen-driven evolution of antibodies (soluble forms of the membrane-bound IG part of the BCR). AIRR-seq data allows researchers to examine intraclonal differences caused primarily by somatic hypermutations in IG genes and affinity maturation. Exploring this essential adaptive immunity process could help elucidate the generation of antibodies with high affinity or broadly neutralizing activities. Retracing their evolutionary history could also clarify how vaccines or pathogen exposition drive the humoral immune response, and unravel the clonal architecture of B cell tumors. Computational methods are necessary for large-scale analysis of AIRR-seq properties. However, there is no efficient and interactive tool for analyzing intraclonal diversity, permitting users to explore adaptive immune receptor repertoires in biological and clinical applications. Here we present ViCloD, a web server for large-scale visual analysis of repertoire clonality and intraclonal diversity. ViCloD uses preprocessed data in the format defined by the Adaptive Immune Receptor Repertoire (AIRR) Community. Then, it performs clonal grouping and evolutionary analyses, producing a collection of useful plots for clonal lineage inspection. The web server presents diverse functionalities, including repertoire navigation, clonal abundance analysis, and intraclonal evolutionary tree reconstruction. Users can download the analyzed data in different table formats and save the generated plots as images. ViCloD is a simple, versatile, and user-friendly tool that can help researchers and clinicians to analyze B cell intraclonal diversity. Moreover, its pipeline is optimized to process hundreds of thousands of sequences within a few minutes, allowing an efficient investigation of large and complex repertoires.

INTRODUCTION

B and T lymphocytes are the main cellular components of the adaptive immune system. They detect and react to substances (antigens) from infectious pathogens or non-infectious origin leading to their elimination. The ability to recognize a huge array of antigen relies on specific cell surface receptors, respectively the B-cell receptor (BCR) and the T-cell receptor (TCR), which display a corresponding tremendous variability. The BCR is made up of a recognition unit formed by an immunoglobulin (IG) protein and an associated signaling unit, the Igα and Igβ molecules. IG consists of four polypeptide chains, two heavy chains and two light chains, joined by disulfide bonds. Besides the membrane-bound form, IG can also be secreted in the blood plasma as circulating antibodies.

The BCR diversity lies in the variable region of the IG gene and results from complex genetic mechanisms that occur during B cell ontogeny in the bone marrow. This process, called VDJ recombination [1], joins different types of genes: V (variable), D (diversity) in heavy chain, and J (joining). This combinatorial diversity is further enhanced by a junctional diversity due to the imprecise joining of these genes with deletion and insertion of random nucleotides at the recombination sites, thereby creating an extremely diverse zone within the variable region, the complementary

*To whom correspondence should be addressed. Tel: +33 1 44 27 73 38; Email: juliana.silva_bernardes@sorbonne-universite.fr
Correspondence may also be addressed to Frédéric Davi. Email: frederic.davi@aphp.fr

determining region 3 (CDR3) (2). Further diversity occurs after the antigen encounter by naive B cells, with a process called affinity maturation, which occurs in specialized structures of secondary lymphoid organs, the germinal centers (3). There, point mutations are introduced in the variable region of IG genes at very high rates, about 10^6 times higher than observed for other genes, by a mechanism termed somatic hypermutation (SHM) (4). This is followed by a Darwinian selection process whereby B cells expressing BCR with higher affinity resulting from SHM survive and proliferate, whereas B cells with lower affinity are eliminated (5). The selected B lymphocytes become either memory B cells or can further differentiate into antibody-secreting plasma cells (6). The resulting overall diversity of the B-cell repertoire is considerable, with in theory at least 10^{12} possible BCRs in humans (7,8).

However, perturbations and restriction in repertoire diversity, accompanied by emergence of expanded clones in some instances, may occur in various pathological conditions such as infections (9,10), auto-immune diseases (11–13), allergies (14–16), after vaccination (17,18) or upon ageing (17,19). The most extreme situation is represented by B cell tumors where a single clone dominates the repertoire (20). Importantly, physiological processes such as SHM are maintained in the neoplastic cells (21) and intraclonal diversification resulting from SHM may occur in B cell tumors originating from germinal centers such as follicular lymphoma (22). As antigen stimulation is now recognized to play an important role in the development of many types of B cell malignancies (23,24), the question whether SHM and improvement of antigen BCR fitness also contributes to tumor clonal evolution remains to be elucidated.

Rapid advances in high-throughput sequencing have opened up new possibilities for studying B cell clonal expansion and intraclonal compositions. With decreasing costs of DNA sequencing technology, numerous IG sequences with inherent complexity, variability, and mutational load are accessible, including in clinical contexts. Thus, the availability of adaptive immune receptor repertoire sequencing (AIRR-seq) data has motivated researchers with different backgrounds (biological, computational, and statistical) to investigate and examine the adaptive immune complexity.

Several methods and computational tools have been developed to treat different AIRR-seq analysis steps, producing multiple integrated context-specific software (25–28). Most existing tools were designed for V(D)J assignments (29), or the identification of groups of clonally-related sequences (30–32). Some solutions exist to tackle and determine the heterogeneity of BCR repertoires (28,33–36). Such tools quantify the diversity of a BCR repertoire by estimating clonal group sizes, characterizing CDR3 regions, identifying the most used V, D, and J genes (alleles), and the most frequent VJ combinations. Few tools allow the analysis of intraclonal diversity/evolution of a B-cell repertoire (37–41). AncesTree (37) is a stand-alone graphic interface to visualize clonal lineage trees interactively. It does not perform any data processing; instead, it parses the outputs of Dnaml (42) and IgPhyML (43). AncesTree displays pre-computed trees, the number of nucleotide and amino acid mutations, and multiple sequence alignments. Although trees are interactive, displaying the location of each mutation, users

must pre-process several tasks to explore intraclonal diversity, such as clonal grouping, clone selection and lineage tree reconstruction.

GLaMST is another stand-alone tool (38); it uses the minimum spanning tree algorithm to infer clonal lineage trees from high-throughput sequencing data. Like Ancestree, GlaMST does not process the raw data, and the intraclonal diversity analysis requires external tools. Intraclonal diversity can also be investigated by running tools available at the Immcantation framework. Unlike GLaMST and Ancestree, Immcantation provides start-to-finish tools to explore intraclonal diversities in a clonal lineage. First, users must identify a clonal lineage of interest, for instance, the most abundant one, and then use Alakazam to compute clonal lineage trees. Mutational profiling can also be obtained by running shazam (44). However, installing and manipulating data require some computational expertise when using the Immcantation framework. In addition, trees and mutational profiles are not interactive, and navigating among the most relevant clonal lineages to investigate their evolutionary events can be difficult.

IgIDivA (40) is the first tool dedicated to intraclonal diversity analysis; it first identifies all nucleotide variants inside a group of sequences annotated with the same IGHV, IGHJ genes, and identical CDR3 amino acid segments. Then, it constructs a network/graph centered on the most abundant variant rather than a tree grown from the hypothetical ancestral unmutated sequence, as done by Gctree (45), GlaMST (38) and ClonalTree (46). IgIDivA requires inputs produced by TRIP (47), another tool from the same team that needs to be previously installed by users. TRIP adopts a different definition of clonally related sequences that do not allow amino acid variations within the CDR3 region.

Most existing intraclonal diversity analysis tools are stand-alone programs, requiring computational skills for installation and pre-processing tasks such as clonal grouping and lineage tree inferences. Furthermore, such tools do not provide a user-friendly web interface with anywhere access that could assist users without prior programming knowledge. Moreover, most existing tools still need interactive plots to help users get insights from their analyses. Thus, we are currently lacking a large-scale, time-efficient and versatile AIRR-seq tool for deep analysis of intraclonal diversity and evolution.

Here, we propose ViCloD, a web-based interactive tool for visualizing B cell receptor repertoire clonality and analyzing intraclonal diversity. ViCloD enables repertoire analysis and proposes a dynamic repertoire topology panorama. Users can navigate among B-cell clonal lineages and explore their intraclonal diversity and evolutionary events. For this purpose, we have developed a set of algorithms to analyze millions of IGH sequences within a few minutes (30,46). Such methods are time efficient and compatible with both research and clinical settings. Users can download all processed data, save the generated plots, perform their analyses locally, and visualize them in the ViCloD web server. Our web server is a user-friendly data visualization tool to help the research/medical community select helpful information from BCR repertoire analyses. It allows non-computational experts to analyze intraclonal diversities, boosting their

research within the immune repertoire investigation. ViCloD provides unique functionalities absent in other web servers dedicated to AIRR-seq data processing, analysis, and visualization. It is a versatile AIRR-seq tool that can potentially help experimental/clinical investigations of B-cell immune repertoires. To demonstrate its functionalities and potential, we applied it to repertoires of human B-cell tumors. As these tumors constitute massive clonal populations, they offer the possibility to evaluate bioinformatic tools on large datasets of sequences. However, ViCloD can be used to investigate B-cell clonal expansions and intraclonal composition of any type of repertoire.

MATERIALS AND METHODS

ViCloD workflow

The ViCloD workflow is presented in Figure 1; it consists of five main steps: upload input, data processing (pipeline), repertoire composition, clonal composition and lineage architecture, described below.

Uploading input. As IGH captures most of the intraclonal diversification, ViCloD has been evaluated so far only on IGH repertoires. IGH sequences should be first filtered upon quality control and annotated to determine their IGHV, IGHD, IGHJ genes and localize their CDR3 region. ViCloD accepts data from IMGT/HighV-QUEST (48), the international standard web software for annotating IG sequences, see Figure 1-A. In principle, the output of any V(D)J annotation tool could be used as long as it provides the minimum information needed for the analysis, specified in http://www.lcqb.upmc.fr/viclod/Format. ViCloD requires AIRR formatted input files (49) that facilitate using other V(D)J assignment tools rather than IMGT/HighV-QUEST. Users can upload the AIRR file on the main page of the ViCloD web server, enter their email address to receive a link to their outputs, and set a threshold t for eliminating infrequent reads; t can have an absolute or percentage value. The analysis is relatively quick; for instance, analyzing a monoclonal repertoire containing around 300K sequences can take around two minutes. However, the computational time can vary depending on the number of simultaneous users and the repertoire structure (the size/amount of clonal lineages and sequence mutations); see some examples in Supplementary Table S1. Users can also submit files previously processed by ViCloD: downloaded from the web server or generated by the stand-alone version available at https://zenodo.org/record/7994894. Thus, users can store and re-analyze their data without any time constraints.

Data processing. Once IGH sequences are uploaded, the first step is to group them into clonally-related sequences. We use MobiLLe (30), a multi-objective-based clustering for inferring B-cell clonal lineages from high-throughput AIRR-seq data. Within each B-cell clonal lineage detected by MobiLLe, we group sequences with the same IGHV, IGHJ genes, and identical CDR3 amino acid sequences, forming *subclones*. Note that we adopted the IMGT definition (50) to define a subclone, whereas we consider a clonotype a group of identical nucleotide sequences, see Supplementary Figure S1. For each clonal lineage, we try to recon-

struct its evolutionary tree by using the N most abundant subclones. We use ClonalTree (46), a time-efficient and accurate approach to reconstruct B-cell lineage trees that incorporates genotype abundances into minimum spanning tree algorithms. ClonalTree requires a file in FASTA format containing sequences and their abundances as input. Each sequence in the input file represents a subclone. Since sequences in a given subclone can differ due to SHM, we select the most abundant sequence to represent it. The abundance corresponds to the total number of sequences within a subclone. ClonalTree also requires the hypothetical ancestral unmutated sequence (HAUS) representing the tree root. To infer the HAUS, we consider the IGHV and IGHJ germline sequences provided by IMGT/HighV-QUEST. For the junction part, we select the sequence with the lowest number of mutations on the IGHV gene compared to the germline, then we extract its junction region. Finally, we concatenate the three parts (IGHV, IGHJ, and junction region) to obtain the HAUS. The Data processing pipeline is illustrated in Figure 1-B.

Repertoire composition. We used nested circles to represent B-cell clonal lineages within a repertoire (with a limit of 1000 circle displayed), Figure 1C. The outer circle represents the entire BCR repertoire, while internal circles correspond to B-cell clonal lineages. The size of each circle correlates with the clonal lineages' abundance. A bar plot on the same web page shows the abundance of the ten most predominant clonal lineages. Users can choose between a standard or logarithmic scale; they can also select a threshold of clonal lineage abundance when analyzing the clonality of a repertoire, for instance, $>5\%$. When such a threshold is set, only B-cell clonal lineages with an abundance higher than this value are retained in the BCR repertoire visualization; all others are hidden. A table at the bottom of the page summarizes some essential properties of each inferred B-cell clonal lineage, such as its identifier, its abundance, the IGHV/IGHJ gene annotations, the CDR3 amino acid sequence and the 'functionality' status of the most abundant clonal lineage sequence, where 'yes' indicates productive sequences, and 'no' unproductive ones. The latter is due to stop codons or out-of-frame IGHV-IGHD-IGHJ junctions.

Clonal composition. ViCloD allows analyzing clonal composition for B cell lineages with abundance higher than 1%. For that, users can click on one of the labeled circles representing the most abundant B-cell clonal lineages. The circle expands, and other circles, representing subclones, appear (see dotted arrow in Figure 1C, D). The circle area correlates with the subclone abundance within the B-cell clonal lineage; see panel 'Clonal lineage population' in Figure 1D. A circular B-cell lineage tree generated by ClonalTree is displayed, representing the evolutionary relationships among such subclones and the HAUS. Several options for tree representation are available. For instance, nodes can be colored by subclones, a different color for each one, or by functionality, where green and red nodes represent productive and unproductive rearrangements. It is also possible to display the length of tree branches representing distances between sequences. The HAUS is represented by a triangle, and the
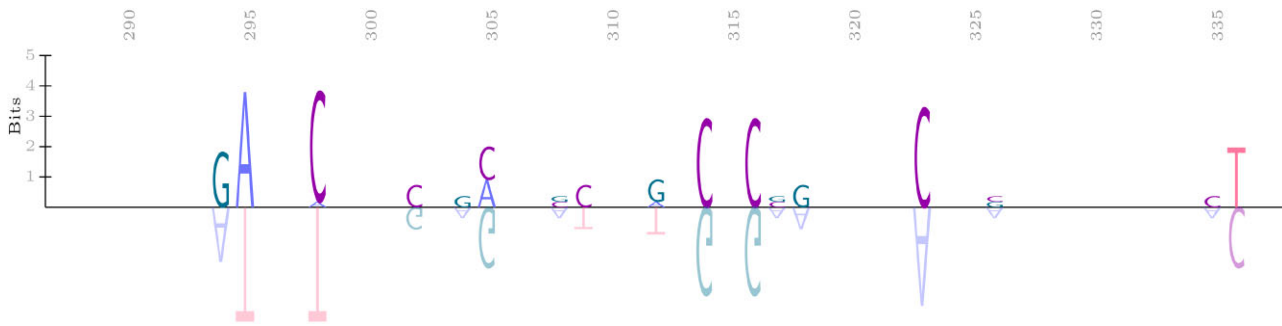
Figure 1. ViCloD flowchart. (**A**) Input. IGH sequences of a given repertoire are previously annotated and uploaded within the ViCloD web server as an AIRR file. (**B**) Data processing. MobiLLe processes the AIRR file and infers B-cell clonal lineages, which are then analyzed by clonalTree to reconstruct their evolutionary tree. (**C**) Repertoire composition. Clonal lineages are represented as nested circles, a bar plot shows the abundance of the most predominant, and a table summarizes some of their essential properties. (**D**) Clonal composition. Users can explore clonal composition by selecting a B-cell clonal lineage. Circles represent subclones, and a circular B-cell lineage tree represents the evolutionary relationships. A table shows additional properties of subclones. (**E**) Lineage architecture. Users can explore a B-cell lineage inferred tree by an interactive tree plot, a circular bar chart (mutational profile), a CDR3 logo, and multiple sequence alignments.

most abundant subclone by a square to facilitate tree interpretation. At the bottom of the page, additional information about subclones, such as their identifiers, abundances, and CDR3 amino acid sequences, is displayed.

The lineage tree inferred by the ClonalTree algorithm can contain (at most) 200 of the most abundant subclones. Interpreting the complete lineage tree can be challenging for large clonal lineages with many subclones. Thus, we propose two simplified tree topologies while conserving nodes with high abundance and their path until the root. We developed two strategies for pruning the tree (51) for a more comprehensible visualization, see Supplementary Figure S2. These simplified versions may be more useful than an entire tree with many details. The first pruning strategy transforms the tree shown into Supplementary Figure S2-A in the tree of Supplementary Figure S2B; it eliminates nodes with no descendant or an abundance lower than a predefined threshold (by default 0.01%). This approach is more helpful for trees containing few nodes. Trees with many subclones can require the second pruning strategy that eliminates nodes with low abundance if they present high similarity with more abundant subclones (Supplementary Figure S2B, C).

For both strategies, we eliminate just leaf nodes and keep the N most abundant subclones in the simplified tree (by default, $N = 10$). We try to preserve the tree's evolutionary paths by eliminating only leaves' nodes, which could be destroyed when an internal node is removed. We keep the N largest subclones since they contain relevant information to analyze the lineage tree. We apply the first strategy once, and on the resulting tree, we repeat the second one several times until achieving at most 30 nodes. Users can navigate between the complete lineage tree and pruned trees on the website.

Lineage architecture. The last ViCloD module provides more details about the B-cell lineage inferred tree, Figure 1E. Users can explore the simplest tree to examine intraclonal diversities. ViCloD provides four types of representations: an interactive tree, a circular bar chart (mutational profile), a CDR3 logo, and a multiple sequence alignment.

Interactive lineage tree. We kept the most pruned tree with at least $N = 10$ most abundant subclones for an easier exploration. The coloured circles indicate observed subclones,

```
Name    fwr1                                                                    cdr1                              fwr2
        10    20    30    40    50    60    70    80    90    100   110   120   130   140
HAUS  GAGGTGCAGCTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCCCTGAGACTCTCCTGTGCAGCCTCT GGATTCACCTTTAGCAGCTATGCC ATGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTG
C1-1  .........T.............AT..............................G...................................
C1-2  ....T.............AT..................T..G..A....G.....
C1-3  ....T.............AT..................G.....
C1-4  .........C.......................T....................A.....
C1-5  ....T.............AT..................G.....
C1-38 ....................T........................
C1-8  ....T.............AT..................T..G..A....G.....
```

```
Name    cdr2                        fwr3
        150   160   170   180   190   200   210   220   230   240   250   260   270   280
HAUS  GGTCTCAGCT ATTAGTGGTAGTGGTGGTAGCACA TACTACGCAGACTCCGTGAAGGGCCGGTTCACCATCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCCGAGGACACGGCCGTAT
C1-1  ......AG....A...G.....CA..TT....................G.....................T..T..A.........G..........C.........A.G.
C1-2  ......AG.................CA..T....................A.....G.....................T..T..A.........G..........C.........A.G.
C1-3  ......AG....A...G.....CA..TT....................G.....................T..T..A.........G..........C.........A.G.
C1-4  ......AG..........C.....CA..TTTG....................A.....................T..A.....C.G..........C.........A.G.
C1-5  ......AG....A...G.....CA..TT....................G.....................T..T..A.........G..........C.........A.G.
C1-38 ......AG...............CA..TTC....................A.........................T..A.........G..........C.........A.G.
C1-8  ......AG...............CA..T.....A.....G.....................T..T..A.........G..........C.........A.G.
```

```
Name    cdr3                        fwr4
        290   300   310   320   330   340   350   360   370
HAUS  ATTACTGT GTGCGATATTGTAGTAGTCATGATTGGGAATCTTATTACTTTGACTAC TGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAG
C1-1  .C..T......GA..C...............C.C......C...........T.................
C1-2  .C..T.......A..C.......C.......G.C.C.G....C...........
C1-3  .C..T......GA..C...C.............C.C......C...........T.................
C1-4  .C..T........A..C...........C.................
C1-5  .C..T......GA..C...A.......C.C......C...........T.................
C1-38 ....T........A..C...C..................
C1-8  .C..T........A..C...C..C......G.C.C.G....C...........
```

Figure 2. Multiple sequence alignment of representative sequences of clonal lineage C1 (Patient 1). Sequence alignment of seven subclones of clonal lineage C1: the five most abundant (C1-1 to C1-5), the closest (C1-38) and the most distant (C1-8) to the HAUS. The alignment reveals conserved and mutated residues of subclone sequences compared to the HAUS. The framework regions (frw) fw1, fw2, fw3 and fw4 regions are displayed in yellow, while the complementary determining regions (cdr1, cdr2, cdr3) in violet. IGHD positions are underlined, only mutated nucleotide are displayed, and a dot represents conserved ones.

while the white circles represent unobserved nodes, which reflect unobserved sequences that were not sampled or eliminated during affinity maturation process; see ClonalTree algorithm (46). The HAUS is represented by a triangle. The branch length depicts the number of SHM between the representative sequences of connected subclones or the HAUS. We designate subclones by a number associated with their abundance, so the subclone 1 is the most abundant in the B-cell lineage. In the tree, nodes representing the five largest subclones have a bold border and are numbered from 1 to 5. When the user moves the cursor over nodes, subclone details (identifier, abundance, and functionality) appear. The option 'display abundance' displays each subclone abundance by increasing the node's diameter. In this visualization mode, nodes are labeled with their abundance (%) rather than a sequential number. Each subclone may contain groups of identical sequences (clonotypes). Choosing 'Display proportion of clonotypes' will turn each node into a pie chart, illustrating the numerical proportion of identical sequences within the subclone. A pie chart for each node can be viewed on a separate page by right-clicking on the node and selecting 'See the distribution of clonotypes' Supplementary (Figure S3).

Circular bar chart - mutational profile. The circular bar chart, also termed as 'Mutational profile', shows the cumulative path distances for a set of selected subclones. By default, ViCloD displays the five most abundant subclones, but any subclone can be included and/or removed from the plot; a maximum of eight items are displayed. Each colored section is related to a subclone and represents the number of point mutations observed between a given subclone and

its parent in the tree. Users must hover the mouse over the desired subclone identifiers to highlight the tree path from a subclone to the root (HAUS). To highlight a branch in the tree, one can hover the mouse over each section of the circular bar chart. The number of mutations between a pair of connected subclones are also displayed.

Multiple sequence alignment. To illustrate the conservation/mutations between the representative sequence of each subclone and the HAUS, we build a multiple sequence alignment (MSA) with the MUSCLE program (52). For each subclone, we display in a table: its identifier, the percentage and number of reads, the divergence rate (number of mutations) between the subclone representative sequence and the HAUS, and the percent identity of IGHV sequence when compared to the HAUS. Only mutated nucleotides of the subclone representative sequence are shown in the alignment, whereas a dot represents conserved positions. CDR and framework sections are highlighted with different colors, and the IGHD gene is underlined in each sequence; see an example in Figure 2. Users can sort out sequences in the table based on each column. The table can be downloaded into a CSV file by clicking on the download button. Users can also select one, multiple, or all sequences and send them directly to the IMGT/HighV-quest web server for additional information and more graphical visualization. A button on the top-right allows users to visualize a simplified MSA that can make its interpretability easier. A second button displays the complete alignment containing all subclones.

CDR3 logo. Sequence logos provide a more detailed and precise description of sequence similarities and are

Figure 3. CDR3 logo. The logo was constructed from the multiple sequence alignment of CDR3 subclone representative sequences and the HAUS. The upper section depicts the most frequent nucleotides that differ from the hypothetical ancestral unmutated sequence (HAUS), while the lower section shows the corresponding HAUS residues. In the upper section, a stack of residues represents each aligned position. The relative sizes of each letter indicate its frequency in the alignment, while the total height of each stack is proportional to the information content and computed in bits. In the lower section, the HAUS residues are displayed as a mirror, and their height corresponds to the total height of the stack in the upper section.

frequently used to unravel significant features difficult to visualize in the multiple sequence alignment. In the 'Lineage architecture' page, users can generate a sequence logo of the CDR3 region by clicking on the 'CDR3 Logo' button. We consider the multiple sequence alignment of CDR3 subclone representative sequences and the HAUS to construct the logo. It is split into two sections; the upper one shows the most frequent amino acids that differ from the HAUS, while the lower section shows the corresponding residues in the HAUS. Note that in the upper zone, the height of letters indicates their frequency in the alignment, while the total height of each stack is proportional to the information content computed in bits. In the lower section, the HAUS residues are displayed as a mirror, and their height corresponds to the total height of the stack in the upper section. Higher residues indicate more conserved positions, while short ones are highly mutated; see an example in Figure 3.

Implementation

The server side of the ViCloD website was implemented in Python (version 3), Bash script and PHP (version 7.4.3, http://php.net/) that was used for interactive data processing. We also used the MySql database (mysql.com/) to control the user's pending jobs. The client-side web interface was implemented using HTML5, CSS, and JavaScript (javascript.com/) along with several libraries, including D3.js (https://d3js.org/), FileSaver.js (github.com/eligrey/FileSaver.js/), saveSVGAsPNG (github.com/exupero/saveSvgAsPng), PDFKit (pdfkit.org/), and JSZip (stuk.github.io/jszip/). Users submit their input files and can provide an email address to receive their results. We first check that the input file contains all the required information. Then, data are processed by the ViCloD Data processing tools, and all analyses are directly available or by clicking on the link sent by email. Users' files (inputs and outputs) are available for download (for ten days) and can be retrieved during this time. Users can also download JSON files (json.org) that can be afterwards uploaded into the ViCloD web server. It enables data to be stored locally and re-analyzed without time constraints. Users can also run ViCloD locally, using a stand-alone version implemented in Python and bash scripts, see Section "Uploading input".

Tutorials

The 'Tutorial' section of the website includes a variety of text-based tutorials to help users getting started with ViCloD. In particular, we provided two tutorials to assist users in generating ViCloD inputs from IMGT/HighV-QUEST and iReceptor repository. Moreover, multiple analyzed data sets are provided on the 'Example' page to assist users step by step; all input files can be downloaded and inputed within ViCloD web server. Several sections under the 'Overview tab' provide additional information, such as references, algorithms used, and the server's versions.

Experimental data

IGH-VDJ rearrangement sequences were obtained from patients with B cell malignancies during standard routine diagnostic or prognostic procedures at the Pitié-Salpêtrière Hospital. Informed consent was obtained from patients according to local regulations. Genomic DNA was extracted from peripheral blood or tissue biopsy, and IGH-VDJ next-generation sequencing was performed as previously published (53). Four patients with either follicular lymphoma or chronic lymphocytic leukemia were selected based on the findings of intraclonal diversity in their tumor clonal IGH-VDJ gene rearrangements. The main clinical characteristics of these patients are provided in Supplementary Table S2.

RESULTS

BCR immune repertoire studies have a significant impact in a variety of research and clinical areas. Several works have reported a wide range of intraclonal diversity when analyzing BCR repertoires of patients with different B-cell malignancies (54–56). In the following section, we applied ViCloD to BCR repertoires from two B-cell malignancies, illustrating different levels of intraclonal diversities. The clinical properties of all analyzed patients are summarized in Supplementary Table S2.

Clonal heterogeneity in a case of follicular lymphoma

Follicular lymphoma (FL) is one of the most frequent types of non-Hodgkin lymphomas, and is often a relatively indolent disease, although it can transform into an aggressive

form. It is a tumor of mature B lymphocytes, typically developing in lymph nodes, and originates from germinal centers, where most of SHM occur. Hence, lymphomatous FL cells express surface IG, which carry high levels of SHM in their IG variable regions, comparable with that of normal germinal center B cells (54). The process of SHM may continue in lymphomatous cells leading to clonal evolution and the emergence of subclones over time (57,58).

We used the IGH repertoire of a patient with FL (labeled patient 1) to demonstrate all types of analyses provided by ViCloD, see http://www.lcqb.upmc.fr/viclod/Examples/example1. Around 186K sequences from the patient were uploaded, but only 46K ($\approx 25\%$) were analyzed after quality control and filtering ($t = 0, 005\%$); see Section "Uploading input". Figure 4A shows the repertoire composition of patient 1, with the presence of dominant clonal lineage accounting for more than 94% of the repertoire, annotated with IGHV3-23 and IGHJ4 genes. Such clonal proliferation is expected in this type of B-cell malignancy. Minor unrelated clonal lineages, each at a frequency <0.2% were also detected, probably originating from the non-tumoral polyclonal background of B cells, see Supplementary Table S3. ViCloD allows users to examine the five most abundant clonal proliferation, labeled in Figure 4A.

Users can click on a given clonal lineage to inspect its composition, see Section "Clonal composition". Figure 4B shows subclones for the most predominant clonal lineage C1. For this example, we labeled only six subclones, with an abundance greater than 1%, but 87 were detected, indicating a high level of intraclonal diversity for C1 lineage. Sequences of all subclones are productive. Subclones C1-1 and C1-2 are the most abundant of the C1 lineage, grouping around 60% and 22% of IGH sequences, respectively. We also observed intraclonal diversity for other clonal lineages; more details are available on the 'Examples tab' at ViCloD website, see the URL above.

ViCloD also provides lineage architecture details for a given clonal lineage. Figure 5 shows the B-cell lineage tree and mutational profile for the clonal lineage C1. In the tree, Figure 5A, the HAUS is marked by a triangle, while the most relevant subclones appear as circles. The circle area is proportional to the sub-clonal abundance, and pie charts indicate the sub-clonal composition with the proportion of clonotypes. For instance, the subclone C1-1 is composed of 536 different clonotypes, while C1-2 by 255; see clonotype distributions for the five most abundant subclones in Supplementary Figure S3. The tree topology confirms the high intraclonal diversity of lineage C1. The reconstructed tree has several levels, revealing the intricate evolutionary relationships of the most abundant/relevant subclones. Accordingly, subclone C1-1 gave rise to seven subclones, and C1-2 evolved from a minor subclone, C1-6. At the top of the tree, subclone C1-38, which accounts for <0.05% of the clonal lineage C1, is the most conserved, with a lower mutation rate (19 nucleotide differences) compared to the HAUS. The longest path in the tree is until subclones C1-8, C1-30 and C1-39, which group 0.49%, 0.07% and 0.05% of the lineage, respectively, and harbor the highest number of mutations ($n = 47$). Interestingly, the subclone C1-4 presents several descendants, but none of them is among the most abundant.

Figure 5B shows a circular plot with cumulative distances for the seven most abundant/relevant subclones of the C1 lineage. Each colored arc represents a subclone, and the arc length corresponds to the number of observed mutations when comparing the subclone and its direct parent. Numbers in the circular plot represent the cumulative mutations from a given subclone until the HAUS. For instance, C1-4 presents 30 cumulative mutations, 21 mutations between C1-4 and C1-38, and 19 mutations between C1-38 and the HAUS. On the other hand, C1-8 presents 47 cumulative mutations, the sum of mutations between subclones and their parents along the path from C1-8 to the root; see Figure 5B. The circular plot (mutational profile) is interactive, allowing users to add/remove subclones by clicking on nodes in the tree; a maximum of eight subclones can be represented simultaneously. Moving the mouse over a subclone on the circular plot shows its evolutionary path until the root (highlighted in red in the interactive tree).

Cumulative mutations can be further examined in Figures 2 and 3: the former shows multiple sequence alignment (MSA) along the entire IGH-VDJ rearrangement, while the later displays a sequence logo centered on the CDR3 region. Parts of sequences in MSA are colored according to their structural framework regions (FR, yellow) and complementary determining regions (CDR, violet). Conserved positions are indicated by dots, while the presence of a nucleotide indicates a mutation in that position. The IGHD gene is underlined within the CDR3, which is the most diverse part of IGH sequences. The sequence logo helps examine mutations in that region, Figure 2.

ViCloD identifies different types of clonal lineage architectures

Chronic lymphocytic leukemia (CLL) is another frequent B-cell malignancy, characterized by accumulation in the blood of mature B clonal B cells. Sequencing of IG expressed by the leukemic cells has revealed two subgroups of CLL depending on the mutation load of the IGHV genes. In addition, this parameter constitutes a strong prognostic indicator as patients with no or few mutations have a more aggressive disease (55). Recent high-throughput IGH repertoire sequencing experiments have revealed a much higher intraclonal diversity than anticipated, although the exact mechanisms leading to this ongoing SHM activity remain elusive (56). We selected three patients with CLL to demonstrate how ViCloD can analyze repertoires having different levels of intraclonal diversity. We selected the major clonal lineage from the repertoire of patients 2, 3 and 4 that correspond to 96%, 99% and 99% of the repertoire, respectively.

Figure 6 shows B-cell lineage tree topologies and mutational profiles for selected individuals' most abundant clonal lineages. When comparing the clonal trees of the three patients (Figure 6-A,C,E), we observe different topologies. The clonal tree of patient 2 is very branched; the most abundant subclone C1-1 descends from C1-2, the second most abundant. C1-2 presented eight cumulative mutations, while C1-1 14; we observed six point mutations between them. We can also see the high intraclonal diversity of patient 2 in the mutational profile shown in Figure 6B. Subclone C1-33 is the closest to the HAUS with seven
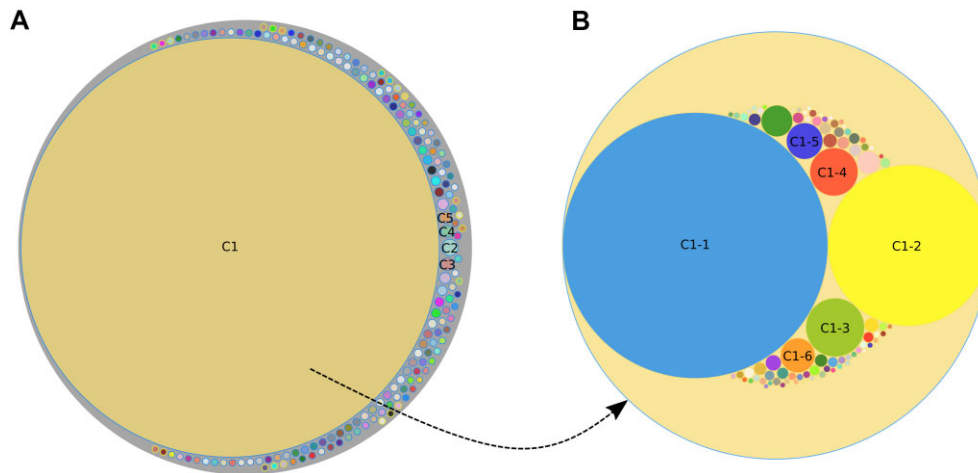
Figure 4. Repertoire composition of a patient with follicular lymphoma (Patient 1). (**A**) the outer circle represents the entire repertoire, while the inner circles represent B-cell clonal lineages inferred by MobiLLe algorithm. The area of inner circles correlates with clonal lineage abundance; only the five most abundant B-cell lineages (C1 to C5) are labeled. (**B**) Expanded view of C1 clonal lineage (the most predominant within the repertoire). Each circle represents a subclone, a group of sequences annotated with the same IGHV and IGHJ genes and identical CDR3 amino acid sequences. Subclone circle areas are proportional to their abundance. Only subclones with an abundance >1% are labeled.
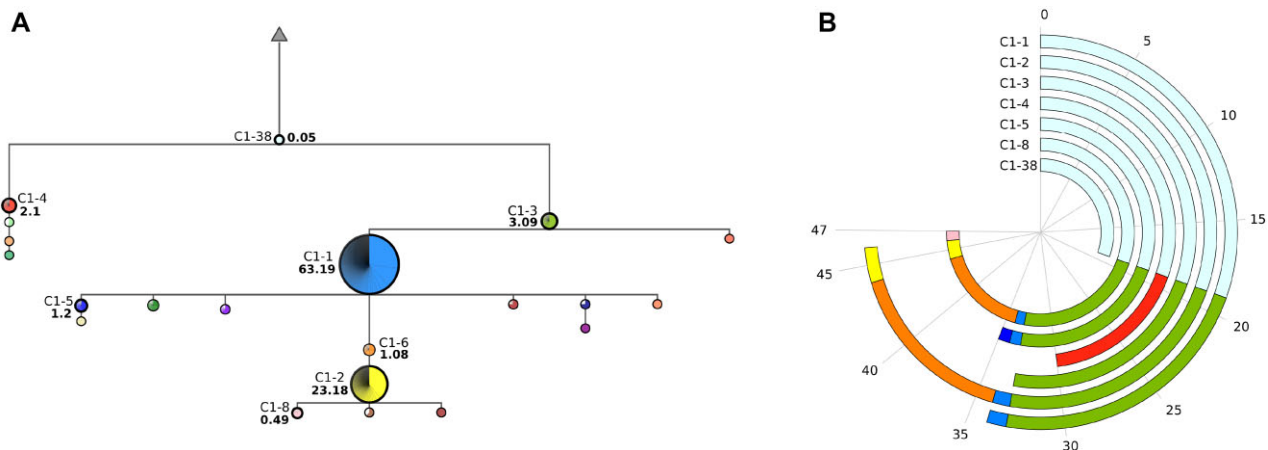
Figure 5. Lineage architecture (Patient 1). (**A**) Simplified representation of the evolutionary relationships between subclones within the dominant clonal lineage (C1) of patient 1. The triangle represents the hypothetical ancestral unmutated sequence (HAUS), whereas circles represent the subclones. The size of the circles is proportional to their sub-clonal abundance, and the pie charts depict the clonotype composition with a subclone. The reconstructed tree topology of clonal lineage C1 displays high intraclonal diversity, with several levels suggesting the intricate evolutionary links among the most abundant subclones. (**B**) Circular plot showing cumulative distances (number of mutations) for the five most abundant subclones, the most conserved (C1-38) and the more divergent subclone (C1-8). Each colored arc represents a subclone; its length corresponds to the number of mutations between the subclone and its direct parent. Numbers in the circular plot represent the cumulative mutations from a given subclone until the HAUS. For instance, C1-8 presents 47 cumulative mutations, the sum of mutations between subclones and their parents along the path from C1-8 to the root.

nucleotide differences, while C1-51 has the most mutated sequence with 27 cumulative mutations. The third most abundant subclone, C1-3, is also highly mutated, with 22 cumulative mutations. C1-3 descends from a less abundant subclone C1-5 that originates from C1-6.

In patient 3, we also observed some intraclonal diversity, but with a different tree topology and mutational profile, when compared to patient 2; compare Figure 6-AB and Figure 6C, D. The most abundant subclone of patient 3 (C1-1) grouped 61.93% of clonal lineage sequences and has 13 cumulative mutations. subclone C1-2, with an abundance of 35.69%, presented five cumulative mutations; between C1-1 and C1-2 we observed ten nucleotide differences. Both subclones share a few point mutations, which is why the ClonalTree algorithm puts them in different tree branches. The intraclonal diversity is also evident in the mutational profile of patient 3, Figure 6D shows a minimum of four (C1-50) and a maximum of 19 (C1-53) cumulative mutations.

On the other hand, patient 4 shows lower intraclonal diversity, which is clearly observed in its tree topology and mutational profile, Figure 6E, F. This patient's repertoire has one predominant subclone (C1-1), representing more than 95% of both clonal lineage and entire repertoire. It has just one cumulative mutation, while the subclone C1-50 has zero. We also observed a tree with few levels (only three), see Figure 6E. A total of 86 subclones were detected. All subclones (except C1-50) descend from C1-1, and all C1-1 descendants present just two cumulative mutations, and thus,
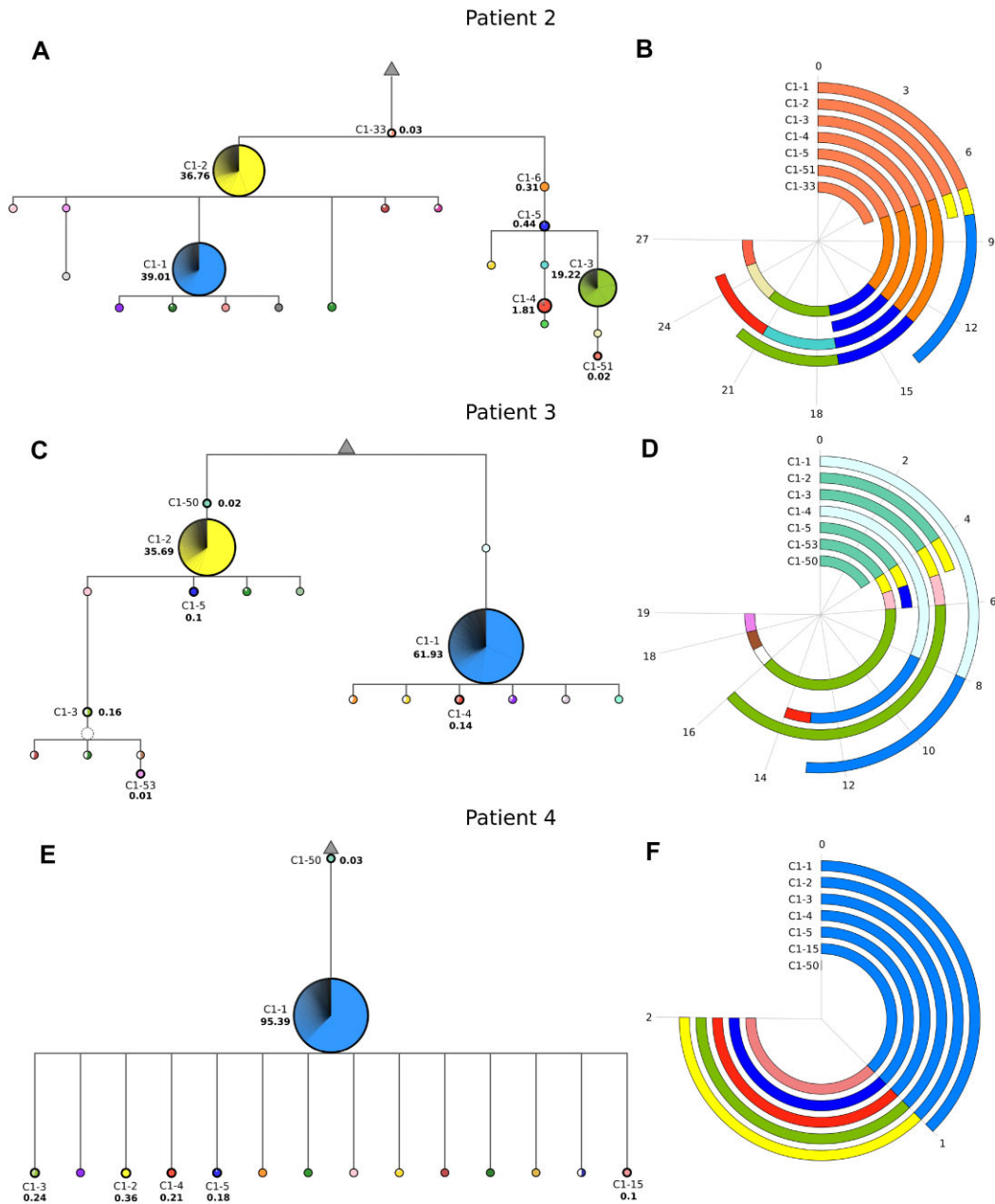
Figure 6. Clonal lineage architectures for three patients with CLL. B-cell lineage tree topologies (**A, C, E**) and mutational profiles (**B, D, F**) of three CLL patients' most abundant clonal lineage. Patient 2 has the most branched tree and many subclones with high mutation rates, presenting the highest intraclonal variability. Patient 3 is an intermediate case; the tree is also branched but with fewer ramifications when compared to the Patient 2 tree. In contrast, patient 4 has lower intraclonal diversity, with a predominant subclone and fewer ramifications in the tree. All mutations are found in the CDR3 regions and predominantly involve 'G' nucleotides.

just one more than C1-1. Note that all mutations were observed in the CDR3 regions, and interestingly, they involve predominantly 'G' and to a lesser extent 'C' nucleotides.

DISCUSSION

ViClod provides a unique environment for intraclonal diversity analyses

High-throughput sequencing analysis has made possible quantitative studies in immunology, where AIRR-seq data

can contain millions of different receptor gene variants. Individual repertoire analysis is now routinely possible and affordable, resulting in many practical applications in biology and medical research. Several web servers devoted to analyzing AIRR-seq data have been recently reported, most of them for pre-processing, V(D)J gene annotation, clonal grouping, repertoire diversity, and mutation analyzes (25–36). However, only some tools permit the analysis of intraclonal diversity, the characterization of SHM and clonal evolution within a specific clonal lineage (35,37,39,40). Most existing tools require some computational skills to

Table 1. Comparing ViCloD with others AIRR-seq visualization tools. Summary of ViCloD features compared to other tools that provide some evolutionary analyzes for clonal lineages such as tree reconstructions and/or intraclonal diversity. Note that most tools were not designed for intraclonal diversity analyses

	ViCloD	AncesTree	GlaMST	Alakazam	IgIDivA
Input and Data-processing					
Input file format	AIRR	Output files of external tools	FASTA	AIRR	Output files of external tools
V(D)J assignment	no	no	no	yes	no
Clonal grouping	yes	no	no	yes[1]	yes[a]
Evolutionary inference	yes	no	yes	yes	yes
Visualization					
Entire repertoire	yes	no	no	no	no
Clonal lineages	yes	no	no	yes	yes
Intraclonal diversity (tree/graphs)	yes	yes	yes	yes	yes
Mutation profile	yes	yes	no	yes	yes
Interactive plots	yes	yes	no	no	no
Diversity measures	no	no	no	yes	no
Clonal/sub-clonal abundance	yes	no	no	yes	yes
Other features					
Adoption of community standards	yes	no	no	yes	no
Web server	yes	no	no	no	no
Computational experience	no	yes	yes	yes	yes
Programming language	Python, Php, JS	Java	Matlab	R	R

[a]Not directly.

manage installations and resolve dependencies. Their usage is then limited to research environments since, most of the time, clinical environments have software installation constraints. Moreover, any tool provides an interactive, versatile, anywhere-access environment for intraclonal diversity analyses, with a complete view of the repertoire composition, where users can easily navigate among the most relevant clonal lineages and explore their evolutionary events. To fill this gap, we propose ViCloD, an interactive web server for visualizing B cell repertoire composition and analyzing intraclonal diversities. ViCloD offers several interactive plots, including B-cell lineage trees and mutation profiles. Our web server can process and analyze data coming directly from IMGT/HighV-QUEST, the international standard web software for annotating IG repertoire sequence data. However, it is not restricted to IMGT/HighV-QUEST outputs since it adopts the AIRR file format (49). In principle, the output of any V(D)J annotation tool and AIRR files downloaded from repertoire repositories such as IReceptor (59) could be used since it produces the necessary data, see http://www.lcqb.upmc.fr/viclod/Format.

ViCloD calculations are performed in real time, and the results are shown immediately on the screen (or, eventually, a link is sent by email). Therefore, the user can analyze complex data very efficiently and quickly. The web server is user-friendly and versatile, providing several views of BCR repertoire, which can help users to extract important features from their analyses. Moreover, all data and figures are easily downloaded and exported to other applications for further inspection. ViCloD is available to all users without requiring registration or any computational knowledge. Users' data are stored temporarily but can be downloaded in a predefined format that allows posteriorly re-analyzing.

Comparison with other tools

ViClod is an original tool to visualize repertoires and analyze intraclonal diversities. To our knowledge, existing tools

do not offer the same set of features, making direct comparisons difficult. Table 1 shows only the main differences between ViCloD and the other AIRR-seq tools. Only tools providing evolutionary analysis for clonal lineage, tree construction and/or intraclonal diversity visualizations were considered. Note that most tools were not designed for intraclonal diversity analyses. Ancestree is essentially a graphical user interface; although it provides a series of interactive plots, users must pre-process data and utilize external tools to investigate intraclonal diversity. GlaMST is a tool for reconstructing lineage trees; it runs in Matlab and requires input sequences in FASTA format, where the first sequence represents the HAUS (root tree). Users must reconstruct the HAUS manually, which is impractical for large-scale studies. Like Ancestree, GlaMST does not process the raw data, and the analysis of intraclonal diversity requires computational experience to process previous tasks. Furthermore, it does not provide interactive plots or mutational profiles to assist users in their analyses.

Alakazam (39) is part of the Immcantation framework, which allows the execution of several tasks, such as preprocessing, population structure determination, and repertoire analysis. To analyze intraclonal diversity, users must first run Change-O (32) to detect clonal groups and then select a clonal lineage of interest, such as the most abundant one. From this, Alakazam can infer a clonal lineage tree and provide a set of trees with different layouts. The trees are not interactive, and they do not display clonal abundance. Moreover, an overview of the repertoire composition is not direct but can be obtained by running other tools from the Immcantation framework such as Sumrep (35). Although Immcantation offers several useful functionalities for a wide repertoire analysis, including intraclonal diversity, it demands programming knowledge from the user.

IgIDivA (40) is a tool dedicated to intraclonal analysis; it first identifies all nucleotide variants (here termed as clonotypes) inside a subclone. Second, it constructs a network/graph centered on the most abundant clonotype

rather than a tree from the HAUS. Two clonotypes are connected only if their connection is consistent with the progressive acquisition of SHMs. Since IgIDivA analyzes clonotypes instead of subclones, its outputs are not directly comparable to ViCloD ones. Moreover, it does not adopt the AIRR community standards input files, making its usage more difficult with V(D)J assignment tools other than IMGT/HighV-QUEST. However, IgIDivA computes a set of graph network metrics, such as average degree and maximal path length in the network, among others, that are useful to quantify the intraclonal diversity of a given clonal lineage.

ViCloD helps visualize intraclonal diversity in repertoires of patients with different B-cell malignancies

We demonstrated the ViCloD usefulness by analyzing several BCR repertoires from four patients with different B-cell malignancies, some known for bearing intraclonal diversification in their IG variable regions. This clinical context was chosen since the B cell population is dominated by a single clonal IGH-VDJ rearrangement (or two in case of biallelic rearrangements), which helps visualize intraclonal diversity due to a large number of available sequences. ViCloD has constructed different B-cell lineage tree topologies and mutation profiles for the four analyzed repertoires. We have observed trees with higher depths and more cumulative mutations in patients with higher intraclonal diversity. These study cases demonstrated how ViCloD could analyze this type of data in detail, helping to visualize their intraclonal diversity. Furthermore, as it does require specific bioinformatics expertise, it can easily be implemented in clinical practice and could help investigate the clonal architecture of B cells tumors and address the question of how antigen selection influences the tumor development. Although we have applied ViCloD to B-cell malignancies, the server can be applied to other clinical conditions (60), aiming to investigate B-cell clonal expansions in various clinical or physiological contexts with at least one dominant clonal lineage.

Limitations and perspectives

ViCloD offers several interactive plots, including B-cell lineage trees and mutation profiles, allowing users to interact with their data and potentially gain insights from their analysis. However, the web server has some limitations, and additional functionalities will be implemented to provide a complete environment for studying intraclonal diversity and evolution. ViClod can only analyze IGH nucleotide sequences; we will extend it to perform analysis at the amino acid level. We will also incorporate IG light chain analysis in order to capture the entire BCR IG structure. To quantify the diversity of a repertoire, we will include several diversity measures, such as the species richness and the Shannon index, which account for clonal size distribution in the overall repertoire (35). Similarly, we shall provide measures that quantify the intraclonal diversity of a repertoire, computing, for instance, the average degree and maximal path length in the clonal lineage trees (40). As there is no consensus on what defines a clonal lineage, ViClod can be more adaptable, allowing users to use several definitions to visualize evolutionary events at all levels. Finally, we would like

to expand ViCloD to analyze several repertoires, allowing users to compare repertoires from the same individual at various time points or repertoires of different individuals. ViCloD's flexibility also offers the possibility of more deep extensions, including the estimation of B cell migration and differentiation (41).

CONCLUSION

We have produced a new AIRR bioinformatics tool, ViCloD, which can help process, analyze and visualize B cell receptor repertoire clonality and intraclonal diversity in B cell populations with some clonal expansion. ViCloD provides interactive modules for analyzing and visualizing the repertoire topology, clonal compositions, and clonal lineage architectures. Unlike the existing AIRR-Seq tools for intraclonal diversity analysis, ViCloD provides an anywhere-access graphical web interface for clinicians, immunologists, and researchers without strong computational backgrounds. It provides interactive plots and allows users to examine thoroughly their data, including sequences, plots, and figures, which are easily exportable. Finally, ViCloD run-time and memory requirements have been minimized in order to be compatible with clinical applications. With the growing importance of immune repertoire research, ViCloD will play an important role in identifying specific features within immune repertoires, helping decipher the high-dimensional complexity of the adaptive immune system.

DATA AVAILABILITY

ViCloD web server is available at http://www.lcqb.upmc.fr/viclod/, where one can also find user manual and repertoire examples. The stand-a-lone version is available at https://zenodo.org/record/7994894, DOI: 10.5281/zenodo.7994894.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

REFERENCES

1. Schatz,D.G. and Ji,Y. (2011) Recombination centres and the orchestration of V (D) J recombination. *Nat. Rev. Immunol.*, **11**, 251–263.
2. Tonegawa,S. (1983) Somatic generation of antibody diversity. *Nature*, **302**, 575–581.
3. Neuberger,M.S., Ehrenstein,M.R., Rada,C., Sale,J., Batista,F.D., Williams,G. and Milstein,C. (2000) Memory in the B–cell compartment: antibody affinity maturation. *Phil. Trans. Roy. Soc. Lon.. Series B: Biol. Sci.*, **355**, 357–360.
4. Odegard,V.H. and Schatz,D.G. (2006) Targeting of somatic hypermutation. *Nat. Rev. Immunol.*, **6**, 573–583.
5. Jacob,J., Kelsoe,G., Rajewsky,K. and Weiss,U. (1991) Intraclonal generation of antibody mutants in germinal centres. *Nature*, **354**, 389–392.
6. De Silva,N.S. and Klein,U. (2015) Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.*, **15**, 137–148.

7. Schroeder,H.W. (2006) Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev. Comp. Immunol.*, **30**, 119–135.

8. Elhanati,Y., Sethna,Z., Marcou,Q., Callan,C.G., Mora,T. and Walczak,A.M. (2015) Inferring processes underlying B-cell repertoire diversity. *Phil. Trans. Roy. Soc. Lon.. Series B, Biol. Sci.*, **370**, 20140243.

9. Hoehn,K.B., Fowler,A., Lunter,G. and Pybus,O.G. (2016) The diversity and molecular evolution of B-cell receptors during infection. *Mol. Biol. Evol.*, **33**, 1147.

10. Breden,F. and Watson,C.T. (2017) Using high-throughput sequencing to characterize the development of the antibody repertoire during Infections: a case study of HIV-1. In: *Recombinant Antibodies for Infectious Diseases*. Springer, pp. 245–263.

11. Zuckerman,N.S., Hazanov,H., Barak,M., Edelman,H., Hess,S., Shcolnik,H., Dunn-Walters,D. and Mehr,R. (2010) Somatic hypermutation and antigen-driven selection of B cells are altered in autoimmune diseases. *J. Autoimmun.*, **35**, 325–335.

12. Von Büdingen,H.C., Kuo,T.C., Sirota,M., Van Belle,C.J., Apeltsin,L., Glanville,J., Cree,B.A., Gourraud,P.A., Schwartzburg,A., Huerta,G. *et al.* (2012) B cell exchange across the blood-brain barrier in multiple sclerosis. *J. Clin. Invest.*, **122**, 4533–4543.

13. Palanichamy,A., Apeltsin,L., Kuo,T.C., Sirota,M., Wang,S., Pitts,S.J., Sundar,P.D., Telman,D., Zhao,L.Z., Derstine,M. *et al.* (2014) Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis. *Sci. Transl. Med.*, **6**, 248ra106.

14. Wu,Y.C.B., James,L.K., Vander Heiden,J.A., Uduman,M., Durham,S.R., Kleinstein,S.H., Kipling,D. and Gould,H.J. (2014) Influence of seasonal exposure to grass pollen on local and peripheral blood IgE repertoires in patients with allergic rhinitis. *J. Allergy Clin. Immunol.*, **134**, 604–612.

15. Patil,S.U., Ogunniyi,A.O., Calatroni,A., Tadigotla,V.R., Ruiter,B., Ma,A., Moon,J., Love,J.C. and Shreffler,W.G. (2015) Peanut oral immunotherapy transiently expands circulating Ara h 2-specific B cells with a homologous repertoire in unrelated subjects. *J. Allergy Clin. Immunol.*, **136**, 125–134.

16. Hoh,R.A., Joshi,S.A., Liu,Y., Wang,C., Roskin,K.M., Lee,J.Y., Pham,T., Looney,T.J., Jackson,K.J., Dixit,V.P. *et al.* (2016) Single B-cell deconvolution of peanut-specific antibody responses in allergic patients. *J. Allergy Clin. Immunol.*, **137**, 157–167.

17. Wu,Y.-C.B., Kipling,D. and Dunn-Walters,D.K. (2012) Age-related changes in human peripheral blood IGH repertoire following vaccination. *Front. Immunol.*, **3**, 193.

18. Ademokun,A., Wu,Y.-C., Martin,V., Mitra,R., Sack,U., Baxendale,H., Kipling,D. and Dunn-Walters,D.K. (2011) Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging cell*, **10**, 922–930.

19. Dunn-Walters,D.K. and Ademokun,A.A. (2010) B cell repertoire and ageing. *Curr. Opin. Immunol.*, **22**, 514–520.

20. Küppers,R. (2005) Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer*, **5**, 251–262.

21. Klein,U., Goossens,T., Fischer,M., Kanzler,H., Braeuninger,A., Rajewsky,K. and Küppers,R. (1998) Somatic hypermutation in normal and transformed human B cells. *Immunol. Rev.*, **162**, 261–280.

22. Küppers,R. and Stevenson,F.K. (2018) Critical influences on the pathogenesis of follicular lymphoma. *Blood*, **131**, 2297–2306.

23. Sutton,L.A., Agathangelidis,A., Belessi,C., Darzentas,N., Davi,F., Ghia,P., Rosenquist,R. and Stamatopoulos,K. (2013) Antigen selection in B-cell lymphomas–tracing the evidence. *Semin. Cancer Biol.*, **23**, 399–409.

24. Stamatopoulos,K., Agathangelidis,A., Rosenquist,R. and Ghia,P. (2017) Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia*, **31**, 282–291.

25. López-Santibáñez-Jácome,L., Avendaño-Vázquez,S.E. and Flores-Jasso,C.F. (2019) The pipeline repertoire for Ig-Seq analysis. *Front. Immunol.*, **10**, 899.

26. Yermanos,A., Greiff,V., Krautler,N.J., Menzel,U., Dounas,A., Miho,E., Oxenius,A., Stadler,T. and Reddy,S.T. (2017) Comparison of methods for phylogenetic B-cell lineage inference using time-resolved antibody repertoire simulations (AbSim). *Bioinformatics*, **33**, 3938–3946.

27. Yaari,G. and Kleinstein,S.H. (2015) Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.*, **7**, 121.

28. Duez,M., Giraud,M., Herbert,R., Rocher,T., Salson,M. and Thonier,F. (2016) Vidjil: a web platform for analysis of high-throughput repertoire sequencing. *PLoS One*, **11**, e0166126.

29. Cortina-Ceballos,B., Godoy-Lozano,E.E., Sámano-Sánchez,H., Aguilar-Salgado,A., Velasco-Herrera,M.D.C., Vargas-Chávez,C., Velázquez-Ramírez,D., Romero,G., Moreno,J., Téllez-Sosa,J. *et al.* (2015) Reconstructing and mining the B cell repertoire with ImmunediveRsity. In: *MAbs*. Taylor & Francis, Vol. 7, pp. 516–524.

30. Abdollahi,N., Jeusset,L., De Septenville,A.L., Ripoche,H., Davi,F. and Bernardes,J.S. (2022) A multi-objective based clustering for inferring BCR clonal lineages from high-throughput B cell repertoire data. *PLoS Comput. Biol.*, **18**, e1010411.

31. Ralph,D.K. and Matsen IV,F.A. (2016) Likelihood-based inference of B cell clonal families. *PLoS Comput. Biol.*, **12**, e1005086.

32. Gupta,N.T., Vander Heiden,J.A., Uduman,M., Gadala-Maria,D., Yaari,G. and Kleinstein,S.H. (2015) Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*, **31**, 3356–3358.

33. Marcou,Q., Mora,T. and Walczak,A.M. (2018) High-throughput immune repertoire analysis with IGoR. *Nat. Commun.*, **9**, 561.

34. Greiff,V., Miho,E., Menzel,U. and Reddy,S.T. (2015) Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol.*, **36**, 738–749.

35. Olson,B.J., Moghimi,P., Schramm,C.A., Obraztsova,A., Ralph,D., Vander Heiden,J.A., Shugay,M., Shepherd,A.J., Lees,W. and Matsen,F.A. IV (2019) Sumrep: a summary statistic framework for immune receptor repertoire comparison and model validation. *Front. Immunol.*, **10**, 2533.

36. Waltari,E., Nafees,S., McCutcheon,K.M., Wong,J. and Pak,J.E. (2022) AIRRscape: an interactive tool for exploring B-cell receptor repertoires and antibody responses. *PLoS Comput. Biol.*, **18**, e1010052.

37. Foglierini,M., Pappas,L., Lanzavecchia,A., Corti,D. and Perez,L. (2020) AncesTree: an interactive immunoglobulin lineage tree visualizer. *PLoS Comput. Biol.*, **16**, e1007731.

38. Yang,X., Tipton,C.M., Woodruff,M.C., Zhou,E., Lee,F., Sanz,I. and Qiu,P. (2020) GLaMST: grow lineages along minimum spanning tree for B cell receptor sequencing data. *BMC Genomics*, **21**, 583.

39. Stern,J.N., Yaari,G., Vander Heiden,J.A., Church,G., Donahue,W.F., Hintzen,R.Q., Huttner,A.J., Laman,J.D., Nagra,R.M., Nylander,A. and et,al. (2014) B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci. Transl. Med.*, **6**, 248ra107.

40. Zaragoza-Infante,L., Junet,V., Pechlivanis,N., Fragkouli,S.-C., Amprachamian,S., Koletsa,T., Chatzidimitriou,A., Papaioannou,M., Stamatopoulos,K., Agathangelidis,A. *et al.* (2022) IgIDivA: immunoglobulin intraclonal diversification analysis. *Brief. Bioinform.*, **23**, bbac349.

41. Hoehn,K.B., Pybus,O.G. and Kleinstein,S.H. (2022) Phylogenetic analysis of migration, differentiation, and class switching in B cells. *PLoS Comput. Biol.*, **18**, e1009885.

42. Baum,B.R. (1989) PHYLIP: phylogeny inference package. Version 3.2.

43. Hoehn,K.B., Vander Heiden,J.A., Zhou,J.Q., Lunter,G., Pybus,O.G. and Kleinstein,S.H. (2019) Repertoire-wide phylogenetic models of B cell molecular evolution reveal evolutionary signatures of aging and vaccination. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 22664–22672.

44. Yaari,G., Uduman,M. and Kleinstein,S.H. (2012) Quantifying selection in high-throughput Immunoglobulin sequencing data sets. *Nucleic Acids Res.*, **40**, e134.

45. DeWitt III,W.S., Mesin,L., Victora,G.D., Minin,V.N. and Matsen IV,F.A. (2018) Using genotype abundance to improve phylogenetic inference. *Mol. Biol. Evol.*, **35**, 1253–1265.

46. Abdollahi,N., Jeusset,L., de Septenville,A., Davi,F. and Bernardes,J.S. (2023) Reconstructing B cell lineage trees with minimum spanning tree and genotype abundances. *BMC Bioinformatics*, **24**, 70.

47. Kotouza,M.T., Gemenetzi,K., Galigalidou,C., Vlachonikola,E., Pechlivanis,N., Agathangelidis,A., Sandaltzopoulos,R., Mitkas,P.A., Stamatopoulos,K., Chatzidimitriou,A. *et al.* (2020) TRIP-T cell receptor/immunoglobulin profiler. *BMC Bioinformatics*, **21**, 422.

48. Lefranc,M.-P., Duroux,P., Li,S., Giudicelli,V. and Alamyar,E. (2012) IMGT/highv-quest: the IMGT web portal for immunoglobulin (ig)

or antibody and t cell receptor (tr) analysis from ngs high throughput and deep sequencing. *Immunome Res.*, **8**, 26.

49. Vander Heiden,J.A., Marquez,S., Marthandan,N., Bukhari,S.A.C., Busse,C.E., Corrie,B., Hershberg,U., Kleinstein,S.H., Matsen,I., Frederick,A. *et al.* (2018) AIRR community standardized representations for annotated immune repertoires. *Front. Immunol.*, **9**, 2206.

50. Alamyar,E., Giudicelli,V., Li,S., Duroux,P. and Lefranc,M.-P. (2012) IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (Ig) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.*, **8**, 26.

51. Mehta,M., Rissanen,J. and Agrawal,R. (1995) MDL-based decision tree pruning. In: *KDD*. Vol. 21. pp. 216–221.

52. Edgar,R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

53. Langlois de Septenville,A., Boudjoghra,M., Bravetti,C., Armand,M., Salson,M., Giraud,M. and Davi,F. (2022) Immunoglobulin gene mutational status assessment by next generation sequencing in chronic lymphocytic. In: *Immunogenetics*. Springer, pp. 153–167.

54. Klien,U., Goasens,T., Fischer,M., Kanzler,H., Braeuninger,A., Rajewsky,K. and Küppers,R. (1998) Somatic hypermutation in normal and transformed human B cells. *Immunol. Rev.*, **162**, 261–280.

55. Sutton,L.-A., Hadzidimitriou,A., Baliakas,P., Agathangelidis,A., Langerak,A.W., Stilgenbauer,S., Pospisilova,S., Davis,Z., Forconi,F., Davi,F. *et al.* (2017) Immunoglobulin genes in chronic lymphocytic leukemia: key to understanding the disease and improving risk stratification. *Haematologica*, **102**, 968.

56. Bagnara,D., Tang,C., Brown,J.R., Kasar,S., Fernandes,S., Colombo,M., Vergani,S., Mazzarello,A.N., Ghiotto,F., Bruno,S. *et al.* (2021) Post-transformation IGHV-IGHD-IGHJ mutations in chronic lymphocytic leukemia B cells: implications for mutational mechanisms and impact on clinical course. *Front. Oncol.*, **11**, 1769.

57. Aarts,W.M., Bende,R.J., Bossenbroek,J.G., Pals,S.T. and van Noesel,C.J. (2001) Variable heavy-chain gene analysis of follicular lymphomas: subclone selection rather than clonal evolution over time. *Blood, J. Am. Soc. Hematol.*, **98**, 238–240.

58. Carlotti,E., Wrench,D., Rosignoli,G., Marzec,J., Sangaralingam,A., Hazanov,L., Michaeli,M., Hallam,S., Chaplin,T., Iqbal,S. *et al.* (2015) High throughput sequencing analysis of the immunoglobulin heavy chain gene from flow-sorted B cell sub-populations define the dynamics of follicular lymphoma clonal evolution. *PLoS One*, **10**, e0134833.

59. Corrie,B.D., Marthandan,N., Zimonja,B., Jaglale,J., Zhou,Y., Barr,E., Knoetze,N., Breden,F.M., Christley,S., Scott,J.K. *et al.* (2018) iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.*, **284**, 24–41.

60. Cizmeci,D., Lofano,G., Rossignol,E., Dugast,A.-S., Kim,D., Cavet,G., Nguyen,N., Tan,Y.C., Seaman,M.S., Alter,G. *et al.* (2021) Distinct clonal evolution of B-cells in HIV controllers with neutralizing antibody breadth. *Elife*, **10**, e62648.