

# Scalable, interactive and hierarchical visualization of virus taxonomic data

Kashyap Balakavi\*<sup>1</sup>, Rushitha Janga\*<sup>1</sup>, Ahmedur Rahman Shovon<sup>1</sup>, Don Dempsey<sup>2</sup>, Elliot Lefkowitz<sup>2</sup>, and Sidharth Kumar†<sup>1</sup>

<sup>1</sup>Department of Computer Science, <sup>2</sup>Department of Microbiology, University of Alabama at Birmingham

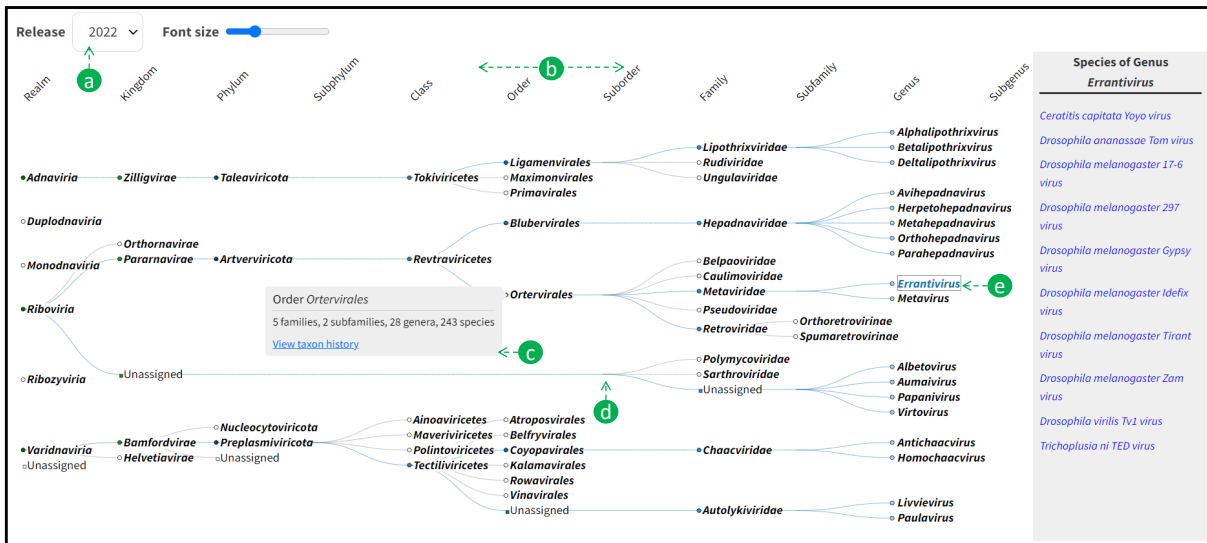


Figure 1: A screenshot of our *Visual taxonomy browser*. The visualization is decoupled into two views: the tree view which contains data till the penultimate level, and the species list view (the grey box on the right) which contains data for the last (leaf) level. Decoupling the visualization into these linked views enables us to scale our visualization to an increasingly high number of virus taxa. (a) lets users select the year from 1971 to 2022, (b) legend illustrating all the ranks present in the selected year (c) tool-tip that includes rank name, taxon name, number of taxa in the preceding lower ranks, and a URL redirecting to the taxon's history (d) showing our bifurcation feature, where the unassigned node is displayed only when there are assigned sibling nodes (e) the last selected node appears in blue color with a rectangular border.

## ABSTRACT

The International Committee on Taxonomy of Viruses (ICTV) (<https://ictv.global>) maintains a database of virus taxonomy, including records from 1971 up to the most recent release in 2022. Recent years have seen an increasing number of viruses, recording 11,273 distinct species in the latest report. Exploring this increasingly high number of species requires a custom data visualization approach that is both scalable and interactive. Our paper presents a structure-preserving, collapsible, node-link layout-based hierarchical visualization called the *Visual Taxonomy Browser* (<https://ictv.global/taxonomy/visual-browser>), for the virus taxonomy dataset. Developed in collaboration with microbiologists, our visualization improves over the existing *Taxonomy Browser* (<https://ictv.global/taxonomy>) which is a stacked bar-based approach that lacks scalability and does not preserve the underlying topological structure of the taxonomy. The *Visual Taxonomy Browser* has been deployed on the ICTV website and has been serving virologists from across the world.

\* authors contributed equally

†sid14@uab.edu

## 1 INTRODUCTION

Biological taxonomy is often referred to as the science of classifying and naming living organisms based on their characteristics and evolutionary relationships. It is a hierarchical system of categorization that groups organisms into increasingly broader and more inclusive categories, with each level reflecting the organism's shared traits and ancestry. The International Committee on Taxonomy of Viruses (ICTV) has the database of virus taxa from 1971 to the latest 2022 release [1]. ICTV maintains this authoritative database of virus taxonomy. This database provides information on the accurate placement of each virus taxon within the taxonomic hierarchy with a complete history of each virus taxon, dating back to its initial release in 1971. The ICTV virus taxonomy database is an essential resource for virologists, researchers, and students alike.

The ICTV provides a web-based visualization tool named the *Taxonomy Browser*, which presents virus taxa in a flattened tree structure as blocks, with a focus on showcasing the taxa lineage rather than its connections to other taxa of the same rank [1]. The visualization does not preserve the underlying structure of the hierarchical data; for example, the nine genera comes directly under the order level in Figure 2 (left), completely skipping the intermediate levels: suborder, family, and subfamily. The emphasis on lineage over relationships to other taxa at the same rank also limits the tool's ability to provide a comprehensive overview of virus taxonomy (see Figure 2 (left)). This implementation of *taxonomy browser* cannot effectively convey the hierarchical relationships between dif-

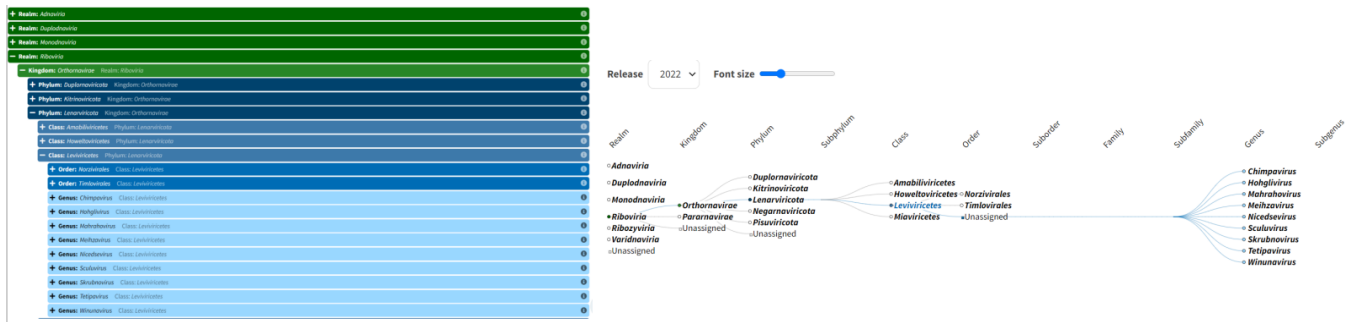


Figure 2: A comparison highlighting the differences between the existing *taxonomy browser* (on the left) and the proposed *visual taxonomy browser* (on the right) in terms of structure preservation. The former visualization *skips* levels which are not present, however, our visualization preserves the topological structure of the data. For example in our visualization we continue to see the three missing ranks (sub-phylum, suborder, family, and sub-family) Our visualization is also more space efficient.

ferent virus taxa, which can be better represented in a traditional hierarchical tree structure.

The data represented in the ICTV database can be effectively visualized using a hierarchical tree data structure. Our *visual taxonomy browser* (see Figure 2 (right)) visualizes virus taxonomy data in a tree visualization with a node-link-based connection, providing several improvements over the previous browser. The node-link-based layout for virus taxonomy makes it easy to see the relationships between different viruses and how they are classified. It allows the stakeholders to easily understand the evolutionary relationships between different viruses, identify patterns and similarities among viruses, and present complex data in an intuitive way. Additionally, with the increasing amount of virus taxonomy data being generated in each release, efficient visualization techniques like tree data visualization can help researchers navigate through the taxa easily and extract meaningful insights. Overall, visualizing virus taxonomy data in a tree data visualization with node-link layout with the facility to accommodate the leaf-level species taxon is critical for understanding the complex relationships among different viruses and their potential impact on public health. Following are the key contributions that make the *visual taxonomy browser* both interactive and scalable:

1. Developed a structure (topology)-preserving hierarchical visualization to handle irregular hierarchical structures—tree structures with missing nodes.
2. Enhanced the existing node-link hierarchical visualization technique to manage a large number of nodes by decoupling the data and visualization into two linked views. One view is for the node-link layout of the tree till the penultimate level, and another separate view is for the last level of the tree. The former view is zoomable and also uses pagination to deal with nodes with high degrees.

Note that we refer to the visualization developed as part of the paper as the *Visual Taxonomy browser* (following the website convention), and the previously existing visualization as the *Taxonomy Browser*. Our visualization has been developed in collaboration with key personnel of ICTV (comprising of virologists) and has been well-received by users for its ease of use, informativeness, and ability to quickly identify relationships between different taxa of the ICTV virus taxonomy.

## 2 RELATED WORK

Several visualization methods have been presented in the literature to explore hierarchical data. Popular examples include hierarchical trees [2], treemaps [3], sunburst charts [4], and dendrograms [5].

The hierarchical tree structure can be categorized into two categories [6] based on the layouts: connection layout and enclosure layout. Connection types of tree visualizations focus on the node-link representation where relationships between the nodes are drawn using edges, allowing for easier perception of the relational structure. Connection-based techniques include tree representation such as cone tree [7], hyperbolic tree [8], disk tree [9], dendrograms etc. The technique of enclosure involves the assignment of each node to a rectangular area that is then subdivided to illustrate the proportional size of its child nodes. Tree-maps are a popular example of this technique, typically arranged in a nested or adjacency layout [2, 3, 10]. Many other hierarchical visualization techniques such as circular tree maps [11], 3D tree maps [12], Pythagoras trees [13], and 3D step trees have been proposed to improve user experience. In particular, [14] used a combination of a 3D tree map and circular treemap for visualizing National Center for Biotechnology Information (NCBI) taxonomy data. The data set has 324,276 nodes spread over 42 levels. Although able to visualize the large volume of data, the system lacks exploratory ability as it does not support a search functionality. Though hundreds of broad strategies for visualizing single trees are presented in Schulz’s treevis survey [15], they fall short of solving our particular issue in the context of structure-preserving virus taxonomy tree visualization. Moreover, to enhance the visualization of hierarchical data for specific types of datasets, utilizing a blend of various tree visualization techniques can provide an optimal viewing experience [16].

The field of biological taxonomy provides an excellent example of a hierarchical taxonomy data structure that can greatly benefit from the use of hierarchical tree visualization techniques. Specifically, research in the area of phylogenetic tree visualization has demonstrated the effectiveness of utilizing a node-link layout for representing and exploring complex hierarchical relationships within biological taxonomic data [17–19]. Such advancements have not only enhanced the understanding of biological classification systems but also have broad applications in various fields that involve hierarchical data structures. The comprehensive database, which is maintained by ICTV, stores virus taxonomy, taxon names, and associated metadata and is recognized as one of the largest databases of biological taxonomy [20–22]. Owing to the scale (upto 14K nodes), it is not possible to directly adapt any existing visualization technique for the virus taxonomy dataset, we therefore develop a custom visualization that decouples the given data into two views both capable of scaling individually.

## 3 LIMITATIONS OF THE *Taxonomy Browser*

The *taxonomy browser* displayed taxa as blocks using a flattened tree structure, placing more emphasis on a taxa lineage than its rela-

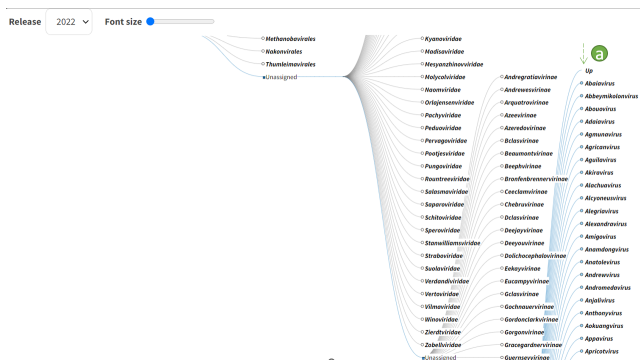


Figure 3: Pagination technique of our *visual taxonomy browser* handling large number of nodes on the penultimate level.

tionships with other taxa of the same rank. The existing taxonomy browser has several key limitations, including scalability, structure preservation, and interactivity issues.

**Scalability** The flattened tree structure was not designed to handle large and complex taxonomic data sets with many taxa, making it difficult for the users to navigate through the taxonomy hierarchies. As a result, users often encountered decreased usability and performance issues when dealing with large taxa, which limited the usefulness of the previous taxonomy browser. In our current implementation, we selected a collapsible tree to visualize the taxonomy. Also, to improve the user experience we implemented several features like pagination, auto-expand, and bifurcation which would be discussed in detail in the coming sections.

**Structure preservation** The existing taxonomy browser listed the taxa as drop-down lists in a linear fashion, which did not preserve the hierarchical structure. Different rank taxa were placed at the same level, making it challenging for users to discern the relationships and dependencies between different taxa. The taxonomy browser also needs to be changed with each release if any property of the taxa is changed in a revised version of the data. Moreover, the existing taxonomy browser cannot visualize the entire visualization at a single glance. As the user opens up additional taxa, the visualization continues to expand vertically, eventually requiring the user to scroll upwards to view the initially opened taxa. Thus, the existing taxonomy browser impedes users from using it effectively.

In the current visualization, the data is visualized in a hierarchical format where the nodes are displayed at their respective levels. The tree opens horizontally and vertically depending on the depth of each node, allowing the users to better understand the tree and its hierarchy. We also incorporated the legend as a part of the tree to give users a better sense of the ranks. The current visualization holds good even with the frequent changes in the data and can visualize the whole tree at a single glance.

**Interactivity** The taxonomy browser suffered from a lack of interactivity, which limited its ability to provide users with a flexible and dynamic browsing experience. In response to this limitation, new features should be implemented to allow users to zoom, pan, adjust the font size, and dynamically adjust the layout based on the ranks and spacing of taxa. Additionally, tooltips should be added to provide users with additional metadata and link to the previous years' releases of each taxon, further enhancing the interactivity and usefulness of the taxonomy browser. We have added features like a zoom slider, font slider, and assigned tooltip to each node which also displays the information of the node in our current visualization.

## 4 VISUAL TAXONOMY BROWSER

We design *Visual Taxonomy Browser* to mitigate the existing *taxonomy browser's* limitations mentioned in the previous section. Our visualization tool uses *D3.js*, a widely-used open-source JavaScript library [23]. We retrieve taxonomic data from existing ICTV web services in JSON format. Since there is vast taxon data and large number of ranks displaying it in hierarchical tree structure can display the parent child relation clearly compared to other structures like tree maps, sunburst and dendrograms. In tree maps data is displayed hierarchically in rectangles, the larger rectangle represents the major category and smaller rectangles represent subcategories. Since there are many ranks in this case it will be very difficult for the user to differentiate the hierarchy by looking at size of rectangles, also this structure drives visual intuition towards metric relation rather than parent child relation. Similarly dendrograms are node link structure that represents hierarchy on data in clusters based on the similarity and not exactly parent-child relation and sunburst chart are concrete circles that represent data hierarchically limiting the number of nodes that can displayed in one circle. We have developed the visualization as a modular package and seamlessly integrated it into the existing ICTV website as a Drupal module. In this section, we discuss the key features of our proposed visualization.

### 4.1 Linked views

The 2022 ICTV report on virus taxonomy includes over 14,000 taxa, with 11,265 listed at the species rank. Similarly, the species rank represents the majority of the taxonomic data across all ICTV reports. Due to the large amount of species data in each ICTV report, it is not feasible to visualize it using only the node-link layout of a hierarchical tree visualization. The idea behind creating two different panels was to manage the large number of internal nodes occurring at the genus level and the large number of leaf nodes at the species level. Table 1 shows the increasing number of species and genus counts over the years. For example, in 2022, there are 2,810 genera and 11,265 species. We note that the maximum degree (branches) at these two levels for that year is over 400. Visualizing these increasing numbers of species and genera in the same panel leads to a highly clustered, non-intuitive user experience. We partially tackle the scale issue by implementing pagination on the genus level; handling the species in the same way as the genus is not a user-friendly method since the visualization will be cluttered and the user has to juggle between pagination at two levels making the navigation difficult.

We address this challenge of developing a scalable visualization by decoupling the hierarchical dataset into two relatively smaller groups and developing a scalable visualization for each of them. Accordingly, we divide our visualization space into two panels. Our visualization tool displays the hierarchical tree with all ranks (tree levels) except species in the primary panel and the species of the selected taxon on an adjacent panel as a list structure. Both panels are integrated on a single view which can be seen in Figure 1. Separating the species from the tree allows for a cleaner and more organized display, enhancing the user's ability to focus on the hierarchical relationships between higher ranks while still being able to access detailed information about individual species. Furthermore, to accommodate larger taxa that exceed a specific threshold for non-species ranks, we implement a pagination technique on the penultimate level shown in Figure 3. By utilizing linked views and pagination techniques, our visualization can efficiently handle a large volume of taxonomic data, allowing users to seamlessly navigate through multiple levels of hierarchy and access detailed information on each taxon.

### 4.2 Structure preservation

The taxonomy data is highly irregular containing a significant amount of missing taxa. The circular nodes from the visualiza-

Ranks	Years							
	1979	1987	1995	2004	2015	2019	2022	
Realm						4	6	
Kingdom						9	10	
Phylum						16	17	
Subphylum						2	2	
Class						36	40	
Order			1	3	7	55	72	
Suborder						8	8	
Family	24	37	50	73	111	168	260	
Subfamily	8	8	9	11	27	103	182	
Genus	84	116	166	290	609	1421	2810	
Subgenus						68	84	
Species	1008	1275	2220	1832	3704	6590	11265	

Table 1: The table outlines how the taxa count varies across the years from 1971 to 2022. There has been a continuous increase in the species count over the years, and hence we decided to implement the split view.

tion represent nodes of a regular taxon using the taxon name and are called assigned nodes. In contrast, we use rectangular nodes to represent the missing data at each level and call them *Unassigned*. To enhance the user experience we implemented an auto-expand feature. The tree initially loads in a collapsed fashion and when the unassigned node is clicked, the auto-expand feature comes into play. It opens the tree till the next assigned node or when there is bifurcation. The situation where an unassigned node has an assigned sibling is called bifurcation. Additionally, the unassigned nodes are only displayed if they are on the highest level or when there is a bifurcation. This feature allows the user to expand and collapse the tree which has an unassigned node as a sibling at will. There is also a legend as a part of the tree, which allows the user to get a better understanding of the rank hierarchy and the node's rank at a glance. This legend moves dynamically with the tree and remains intact when users use the tree's features, such as zooming, panning, dragging, and dropping.

Figure 2 shows a comparison of the former *taxonomy browser* and our *Visual Taxonomy Browser* in terms of structure preservation. As the old implementation shows the taxa in a stacked view, it cannot preserve the topological structure of the taxa.

### 4.3 User interaction

We have implemented several interactive features to enhance the user experience. These include zooming and dragging, a font slider to adjust text size, path highlighting, and tooltips that display additional metadata for each taxon.

**Zoom and pan** We have integrated zoom and pan features into our visualization to enhance the user experience. Users can zoom in or out of the entire tree using the trackpad or mouse scroll via the zoom feature. Meanwhile, the drag functionality enables users to move the visualization by either clicking and holding the left mouse button or using the trackpad to drag the tree. These features promote interactivity, allowing users to explore and concentrate on specific areas of interest within the taxonomy tree with ease.

**Font slider and Zoom slider** Users can customize the font size and zoom level of the tree using the font slider and zoom slider respectively from the control panel on top. The visualization is made comfortable to read by allowing users to change the text size and tree size easily. The font slider and zoom slider gives customers the freedom to alter the visualization in accordance with their tastes, resulting in a more unique and pleasurable user experience.



(a) The taxonomy for *African swine fever virus* in the 2019's report



(b) The taxonomy for *African swine fever virus* in the 1998's report

Figure 4: The visual taxonomy browser maintains the structure of the virus taxonomy across different releases of ICTV reports. The figure shows different representations on different reports for the *African swine fever virus*.

**Path highlighting** The path-highlighting capability has been included to offer a seamless and consistent user experience across a variety of devices. The clicked taxon is highlighted alongside the edges of the path. This feature allows the users to perceive the selected taxon and compare their chosen path with other open paths, providing a more comprehensive understanding of the taxonomy.

**Tooltip** We have added a pop-up window that displays detailed information about each taxon, including the taxon name, rank name, number of nodes in the next levels, and a link that redirects to the taxa history over the years. Users can access this information by hovering over the taxon name or node of the respective taxon, providing quick and easy access to valuable insights about the taxon, including its release history. By including this feature, users can better understand the relationships between different taxa and make more informed decisions.

**Visualization export** We have integrated a versatile export feature into our application, empowering users with the ability to choose from three distinct image formats: JPEG, PNG, and PDF. This functionality can be accessed using a drop-down box positioned at the top of the user interface for convenience.

### 4.4 Dynamic layout adjustment

The layout of the tree is greatly influenced by the spacing between its nodes. To compute the positions of all child nodes relative to the root node, we rely on the *d3* library's layout method. During rendering, we calculate two critical gaps: the gap between taxa within the same rank ( $d_y$ ) and the gap between ranks ( $d_x$ ). These values ( $d_y$  and  $d_x$ ) are dynamically determined based on the dimensions of the SVG container, which, in turn, depend on the screen size.

The adaptive nature of the gap between ranks ( $d_x$ ) ensures that it adjusts to the number of ranks present in the given year. As a result, the spacing between ranks in Figure 4 for the years 1998 and 2019 may differ. This adaptability enables us to accommodate varying distances between ranks while displaying the tree effectively.

On the other hand, the spacing between nodes of the same rank ( $d_y$ ) is calculated based on the number of nodes visible on the screen. This calculation ensures a minimum gap between nodes to prevent overlapping. Notably, as the number of nodes on the screen changes, the distance between the nodes dynamically adjusts to maintain

a suitable gap. This feature is particularly beneficial when dealing with parent taxa that may have many child nodes, preventing overcrowding and maintaining a clear layout.

Our collapsible tree design is highly flexible, allowing it to handle diverse ICTV reports, each with different numbers of ranks. This flexibility is essential in providing a comprehensive layout for various datasets, such as the *African swine fever virus* in 1998 and 2019 releases, as depicted in Figure 4.

## 5 USER STUDY

Our *Visual Taxonomy Browser* is available on the ICTV website, and a group of domain experts was asked to evaluate it for design, interactivity, and scope. We provided users with access to the tool and conducted interviews to gather their assessments and feedback. Our study particularly centered on engaging virologists as a primary group of domain experts. This group, comprised of four virologists, brought a wealth of specialized knowledge in virus taxonomy through their affiliation with the International Committee on Taxonomy of Viruses (ICTV). This ensured that our *Visual Taxonomy Browser* received a comprehensive evaluation from experts deeply familiar with the intricacies of virus classification. In addition to the virologists, our study encompassed a diverse set of participants, including two postdoctoral researchers, two software engineers, and five graduate students. This multidisciplinary approach allowed us to gather feedback and insights from various angles, contributing to a well-rounded assessment of the tool’s design, interactivity, and scope.

We found out that our tool facilitated hypothesis-driven analysis by providing an intuitive way to explore and visualize the virus taxonomy data. Participants found the visualization easy to navigate and appreciated the ability to zoom in and out of specific sections of the tree. They also mentioned that the tool helped to identify patterns and relationships within the taxa, which was not possible using the existing taxonomy browser. According to their opinions, the *Visual Taxonomy Browser* offers an enhanced user experience and streamlined interaction with the hierarchical tree on a single view. Specifically, features such as zoom, pan, pagination, tooltip, and linked views provided users with greater flexibility and interactivity. They mentioned that the path-highlighting feature was particularly helpful, making it easier for users to navigate between different taxa. The experts also appreciated the year-switching menu, which allowed them to easily visualize data for different years, track the evolution of taxa over time, and switch between different reports. One of the experts commented on adding a zoom slider near the font slider at the top of the primary panel which is added in the beta release. Feedback from the test group was overwhelmingly positive, with all experts recommending our *Visual Taxonomy Browser* over the existing stack-based taxonomy browser.

The *Visual Taxonomy Browser* was introduced on April 13th, 2023, and it has made a significant impact on user engagement, surpassing the performance of the old taxonomy browser, as revealed by Google Analytics data which is shown in Table 2 and 3. According to Google Analytics, a *user* refers to a unique visitor to a website within a defined time frame. This uniqueness is determined by a combination of browser and device, tracked using cookies. During the period of January 1st, 2023, to April 12th, 2023, the website had a total of 410,039 views and 80,334 users, with an average engagement time of 2 minutes and 50 seconds. Within this timeframe, the Taxonomy Browser received 44,887 views, had 19,916 users, and an average engagement time of 2 minutes and 35 seconds. After the *Visual Taxonomy Browser*’s availability, from April 13th, 2023, to July 24th, 2023, the website received a total of 378,841 views and 73,445 users, with an improved average engagement time of 3 minutes and 14 seconds. The old *Taxonomy Browser*, during the same period, obtained 44,887 views, had 19,916 users, and an average engagement time of 2 minutes and 35 seconds.

	<i>Taxonomy Browser</i>	<i>Visual Taxonomy Browser</i>
<b>Views</b>	44,887	5,348
<b>Users</b>	19,916	3,376
<b>Returning users</b>	9,772	1,728
<b>Conversion rate</b>	44.36%	63.12%
<b>Avg. engagement time</b>	2mins 35secs	1min 37secs

Table 2: The table shows the web analytics for the *Taxonomy Browser* and the *Visual Taxonomy Browser* from April 13th to July 24th. The later has 18% higher conversion rate.

	Pre study analytics	Post study analytics
<b>Views</b>	410,039	37,8841
<b>Users</b>	80,334	73,445
<b>Returning users</b>	22,408	22,987
<b>Avg. engagement time</b>	2mins 50secs	3mins 14secs

Table 3: The table compares the overall website access before and after introducing the *Visual Taxonomy Browser*. It is clear that the average engagement time of users after the introduction of the *Visual Taxonomy Browser* is longer, and the number of returning users is higher than before.

However, the *Visual Taxonomy Browser* garnered 5,348 views, attracted 3,376 users, and boasted an impressive average engagement time of 1 minute and 37 seconds. This visually appealing and efficient tool outperformed the old taxonomy browser by achieving a remarkable 63.12% conversion rate from views to users. Furthermore, the website experienced a notable increase in returning users. Three months prior to the Visual Browser launch, the returning user rate stood at 27.89%. However, after 3 months of introducing the *Visual Taxonomy Browser*, the returning user rate rose to 31.29%, marking a 12% increase compared to the pre-launch rate. In conclusion, the *Visual Taxonomy Browser* has demonstrated its effectiveness by offering a visually engaging and efficient method for users to navigate through the website’s content. Its improved performance in views-to-user conversion and the significant boost in returning users showcase its value and appeal to the website’s audience.

## 6 FUTURE WORK

In the initial phase of the project, we chose to design a solution to the old taxonomy browser using the collapsible tree from the d3-JavaScript library as it was a more tangible solution to represent a hierarchical structure over the existing visualizations of the d3 library like tree map, dendrogram, and sunburst chart. The collapsible tree is a bit more flexible to modify and visualize the hierarchical data as a parent-child relationship. We developed the collapsible tree as one unit but then noticed that there were a lot of child nodes to some ranks. In the subsequent iteration, we changed our view to a two-window visualization where we visualized all the nodes till the penultimate rank in one window and we call it a taxa panel. The data of all the other ranks are in a smaller window on the same page, which we call a species panel. The next problem we faced was when there were a lot of child nodes in the taxa panel. To handle this we developed pagination where we limit the number of nodes on the screen at a point of time for a better user experience.

In the continuous improvement process, we hid the unassigned nodes, with some exceptions, and implemented the auto-expand feature to give the user a better understanding of the data. We however had to display the unassigned nodes at the highest level and at the bifurcation. We introduced a tooltip to each of the assigned nodes

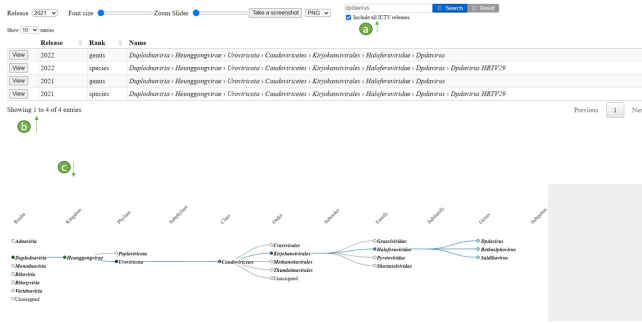


Figure 5: (a): search bar with option to include all releases, (b): table with search results with view button to open up the lineage with respect to the release year, (c): Tree opened up till searched taxa by clicking on the view button for 2021 release year

which displayed the information of the node and we provided a link that would navigate the user to a page that displays the history of the taxon. We provided a hyperlink to the species panel which would take the user to the history page on click. Throughout the development process, we added various visual features to our taxonomy browser, enhancing the user experience. Clicking on a node now highlights both the selected node and its path, making navigation more intuitive. A legend was incorporated into the tree visualization to illustrate the hierarchy of ranks better. Users can seamlessly swap between multiple years, observe changes over time, and adjust zoom and font size for a personalized experience. Additionally, we introduced a convenient screenshot feature for capturing snapshots of the tree in different formats. These enhancements make our taxonomy browser more interactive, user-friendly, and versatile.

As part of our ongoing research, we are planning on implementing the search functionality to the visualization. This feature displays the history of the searched taxa and its changes over several years in a tabular form, as depicted in Figure 5. Users can choose different years to observe the taxa’s evolution. When clicking the view button, the tree will automatically expand to display the searched taxa, with the path highlighted. This stand-out feature enhances user experience, facilitates user study, and boosts the tree’s overall interactivity. We anticipate the search functionality to be a valuable addition to our visualization.

We follow rapid prototyping techniques during the development and initial testing stages. After gathering the requirements, we deploy the test server with domain experts (virologists and ICTV officials), allowing us to gather feedback and make continuous improvements to refine the features. Once the features are fine-tuned, we deploy them on the production server for public use.

## 7 DISCUSSION AND CONCLUSION

In conclusion, our *Visual Taxonomy Browser* overcomes the limitations of the taxonomy browser by providing an interactive interface for users to explore and analyze taxonomy trees across different releases. By utilizing linked views and pagination techniques, our visualization can efficiently handle a large volume of taxonomic data, enabling users to seamlessly navigate through multiple levels of hierarchy and access detailed information on each taxon through interactivity. We believe that our approach can be extended to visualize other biological taxonomies and could lead to the development of a uniform taxonomy visualizer that provides an intuitive and interactive representation of taxon relations. In future releases, we plan to incorporate additional features, such as a zoom slider to adjust the level of detail displayed and a search functionality that automatically expands the tree to the searched node, providing further convenience

for users. We have a demo of our code with the upcoming features at <https://harp-lab.github.io/virus-taxonomy-D3/>.

## ACKNOWLEDGMENTS

This work was funded by the National Institute of Allergy and Infectious Diseases, U.S. National Institutes of Health, Award U24AI162625: Virus Taxonomy: A Community Knowledgebase Supporting Virus Research.

## REFERENCES

- [1] Elliot J Lefkowitz, Donald M Dempsey, Robert Curtis Hendrickson, Richard J Orton, Stuart G Siddell, and Donald B Smith. Virus taxonomy: the database of the international committee on taxonomy of viruses (ictv). *Nucleic acids research*, 46(D1):D708–D717, 2018.
- [2] Martin Graham and Jessie Kennedy. A survey of multiple tree visualization. *Information Visualization*, 9(4):235–252, 2010.
- [3] Brian Johnson and Ben Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91*, pages 284–291, 1991.
- [4] Georgios A Pavlopoulos, Theodoros G Soldatos, Adriano Barbosa-Silva, and Reinhard Schneider. A reference guide for tree analysis and visualization. *BioData mining*, 3:1–24, 2010.
- [5] I. Herman, G. Melancon, and M.S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, 2000.
- [6] Quang Vinh Nguyen and Mao Lin Huang. A space-optimized tree visualization. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, pages 85–92, 2002.
- [7] Andy Cockburn and Bruce McKenzie. An evaluation of cone trees. In *People and Computers XIV—Usability or Else! Proceedings of HCI 2000*, pages 425–436. Springer, 2000.
- [8] Yang Xiang, Michael Chau, Homa Atabakhsh, and Hsinchun Chen. Visualizing criminal relationships: Comparison of a hyperbolic tree and a hierarchical list. *Decision Support Systems*, 41(1):69–83, 2005.
- [9] Chang-Sung Jeong and Alex Pang. Reconfigurable disc trees for visualizing large hierarchical information space. In *Proceedings IEEE Symposium on Information Visualization (Cat. No. 98TB100258)*, pages 19–25. IEEE, 1998.
- [10] Ben Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11:92–99, 1992.
- [11] Fabian Fischer, Johannes Fuchs, and Florian Mansmann. Clockmap: Enhancing circular treemaps with temporal glyphs for time-series data. In *EuroVis (Short Papers)*, 2012.
- [12] Thomas Bladh, David A Carr, and Matjaz Kljun. The effect of animated transitions on user navigation in 3d tree-maps. In *Ninth International Conference on Information Visualisation (IV'05)*, pages 297–305. IEEE, 2005.
- [13] Fabian Beck, Michael Burch, Tanja Munz, Lorenzo Di Silvestro, and Daniel Weiskopf. Generalized pythagoras trees for visualizing hierarchies. In *2014 international conference on information visualization theory and applications (IVAPP)*, pages 17–28. IEEE, 2014.
- [14] Michael Burch, Willem Aerts, Daan Bon, Sean McCarren, Laurent Rothuizen, Olivier Smet, and Daan Wöltgens. Combining interactive hierarchy visualizations in a web-based application. In *VISIGRAPP (3: IVAPP)*, pages 191–198, 2019.
- [15] Hans-Jörg Schulz. Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications*, 31:11–15, 2011.
- [16] Aditeya Pandey, Uzma Haque Syeda, Chaitya Shah, John A Guerra-Gomez, and Michelle A Borkin. A state-of-the-art survey of tasks for tree design and evaluation with a curated task dataset. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3563–3584, 2021.
- [17] Daniel H. Huson, Daniel C. Richter, Christian Rausch, Tobias DeZulian, Markus Franz, and Regula Rupp. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, 8:460 – 460, 2007.
- [18] Daniel H Huson and Celine Scornavacca. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology*, 61(6):1061–1067, 2012.

- [19] Michael S. Horn, Matthew Tobiasz, and Chia Shen. Visualizing biodiversity with voronoi treemaps. In *2009 Sixth International Symposium on Voronoi Diagrams*, pages 265–270, 2009.
- [20] Elliot J Lefkowitz. Taxonomy and classification of viruses. *Manual of Clinical Microbiology*, pages 1262–1275, 2011.
- [21] Andrew MQ King, Michael J Adams, Eric B Carstens, and Elliot J Lefkowitz. Virus taxonomy. *Ninth report of the International Committee on Taxonomy of Viruses*, 9, 2012.
- [22] Marc HV van Regenmortel, Claude M Fauquet, David HL Bishop, EB Carstens, MK Estes, SM Lemon, J Maniloff, MA Mayo, DJ McGeoch, CR Pringle, et al. *Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses*. Academic Press, 2000.
- [23] Mike Bostock. D3.js - data-driven documents, 2012.