Interactive Hierarchical Visualization for Exploring the Neotoma Paleoecological Taxonomy

Yanbing Chen*
University of Wisconsin-Madison
Jonathan K. Nelson†
University of Wisconsin-Madison
Simon J. Goring‡
University of Wisconsin-Madison
Qianheng Zhang§
University of Wisconsin-Madison
Jessica L. Blois¶
University of California, Merced
John W. Williams‖
University of Wisconsin-Madison

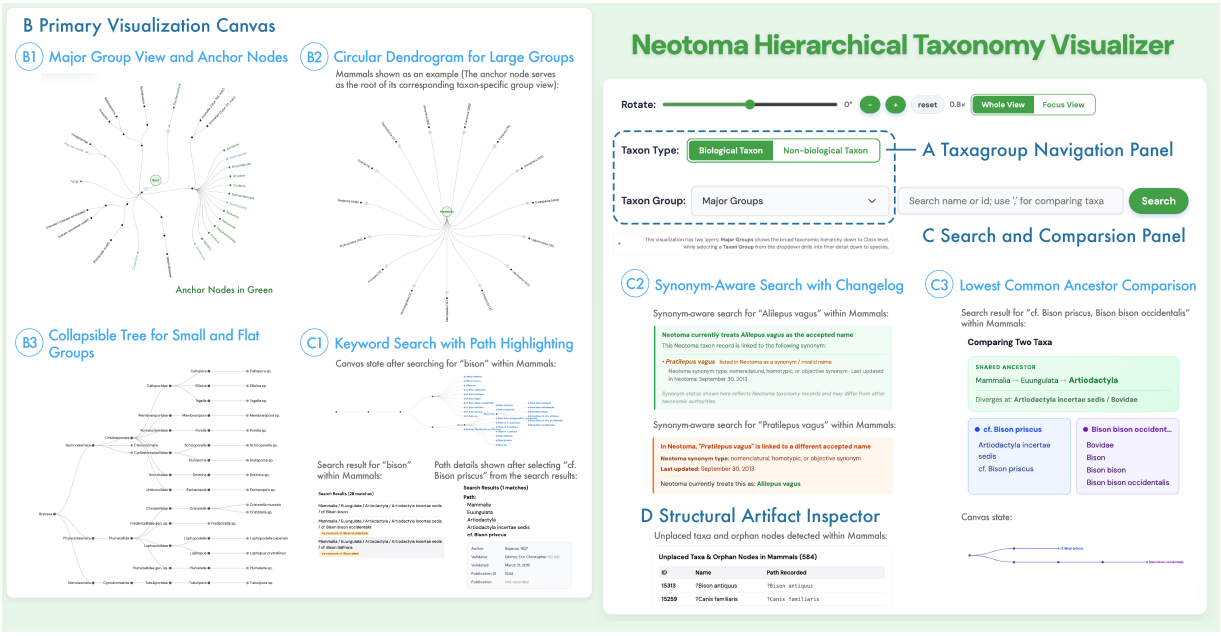

Figure 1: The Neotoma Hierarchical Taxonomy Visualizer integrates four coordinated components: (A) Taxagroup Navigation Panel organizing 52 taxon groups into Biological and Non-biological categories; (B) Primary Visualization Canvas consisting of (B1) anchor nodes rendered in green marking structural interventions in non-Linnaean paths; (B2) circular dendrogram for large biological groups; (B3) collapsible tree for small or flat taxagroups; (C) Search and Comparison Panel consisting of (C1) keyword search displaying full root-to-leaf paths of all matches; (C2) Synonym Panel with an orange box for invalid queried names and a green box for valid names listing historical synonyms with changelog; (C3) Lowest Common Ancestor widget rendering the shared ancestor path and divergence point of two queried taxa, and (D) Structural Artifact Inspector itemizing unplaced orphan taxa. The interface is publicly available at https://yanbingchen.site/Neotoma-Visualizer/.

Abstract

Paleoecological researchers rely on large community-curated databases such as Neotoma to locate, interpret, and compare taxa across geologic timescales. Existing query interfaces assume users already know what they are looking for and offer little support for exploring the hierarchical taxonomic structure itself. Neotoma's taxonomy compounds this problem: its 67,885 taxa span 52 taxagroups that vary from a handful of nodes to nearly twenty thousand, follow a non-Linnaean parent-child structure without fixed rank assignments, and carry historical naming ambiguities accumulated over a century of curation. We present a two-part formative design study focused on detailing our process of, and key takeaways from, developing the Hierarchical Taxonomy Visualizer to address these challenges.

Index Terms: Hierarchy Visualization, Paleoecology Taxonomy, Circular Dendrogram, Collapsible Tree

1 Introduction

Paleoecological and paleoenvironmental researchers rely on taxonomy to study interactions between organisms and their environments across geologic timescales. Taxonomy provides the foundational hierarchical framework for accurately identifying and classifying fossil organisms, enabling the reconstruction of past environments and unlocking new insights into how ecosystems have changed over time. For paleoecologists, taxonomy is both a scientific practice and a data infrastructure: researchers classify fossil and proxy evidence, link names across time periods, and align their work with international nomenclature codes and community standards. Yet, these practices are complicated by evolving classifications, synonym histories, and deep hierarchical structures that

*e-mail: yanbing.chen@wisc.edu
†e-mail:jknelson3@wisc.edu corresponding author
‡e-mail: goring@wisc.edu
§e-mail: qianheng.zhang@wisc.edu
¶e-mail: jblois@ucmerced.edu
‖e-mail: jwwilliams1@wisc.edu

make navigation difficult. Taxonomy work is not a simple lookup task, and existing data interfaces often reduce taxonomy to static tables or search results, obscuring provenance, hierarchy, and the sociotechnical context in which taxonomic decisions are made.

Visualization tools can help support paleoecologists in establishing species relationships and detecting evolutionary trends and noteworthy events in the fossil record [23, 12, 9]. However, creating an effective hierarchical visual representation of taxonomic data remains challenging due to data volume and information overload, structural intricacies and misalignment, complex relationships among entities, and the need to communicate uncertainties [11, 5, 7]. Here, we report on a formative design study focused on developing the Hierarchical Taxonomy Visualizer (HTV) for the Neotoma Paleoecology Database. Neotoma is an open software ecosystem whose mission is to advance understanding of global-change processes by providing an expert-curated data resource; the database contains over 11 million observations from 30,000+ datasets across 20,000+ sites, representing biotic and abiotic proxies of environmental and ecological conditions of the past [29]. Neotoma serves an international audience of paleoecologists, biogeographers, and global change scientists to support research, teaching, and outreach initiatives. This study was motivated in part from a systematic user evaluation of the broader Neotoma software ecosystem that identified a key need for collaboration among Neotoma data stewards and developers in creating tools specifically for taxonomic harmonization and sensemaking [19].

2 Relevant Work

2.1 Interactive Exploration of Biological and Geoscience Databases

Existing interactive tools for large biological and ecological databases fall into three categories. The first treats the database as a record store with minimal structural visualization, supporting metadata-driven dataset discovery and tag-based search [24, 15, 3, 1, 2]. The second adds geospatial and temporal browsing of occurrence records, including PaleoBioDB Navigator [21], SEAD [28], and Neotoma Explorer [18], which allow users to query records by location, time, or taxon name but do not visualize the underlying taxonomic structure. The third category, and the most directly relevant to our work, addresses the hierarchical taxonomic structure itself. GBIF [10] pairs occurrence maps with a sunburst-style taxonomic distribution view spanning kingdoms to orders. Open Tree of Life [20] renders a synthesized phylogeny for interactive lineage tracing. WoL [31] provides a scalable web-based phylogenetic tree viewer capable of handling trees with over 500,000 nodes. These hierarchical tools demonstrate the value of interactive structural visualization for hierarchical taxonomy, but they assume a clean phylogenetic structure: rooted, monophyletic, and consistent in rank assignment. GBIF's sunburst, for instance, requires a Linnaean backbone (kingdom, phylum, class, order) to allocate angular space. Open Tree of Life and WoL assume each internal node has a defined biological ancestor. However, Neotoma's taxa span biotic and non-biotic proxies, include polyphyletic operational groupings, and are linked through a parent-child field that does not encode Linnaean ranks.

2.2 Hierarchical and Circular Visualization

Beyond domain-specific applications, general hierarchical visualization research has produced a range of layout strategies for representing tree structure at scale. Classical techniques include node-link diagrams, treemaps[25], sunburst charts [26], and circular dendrograms, each suited to different combinations of hierarchy depth and quantitative encoding. Recent work extends these classics through alternative layouts. Bubble Treemaps [11] use force-based circle-packing with nested contour arcs. NGL Tree [5] compares multiple hierarchical representations, including pythagoras tree, icicle plot, galaxy plot, circular treemap, and others. Circular dendrograms have become a common choice for large hierarchies because they distribute screen space across angular sectors, scaling to thousands of nodes within a single view [22]. They are also familiar to paleoecologists, who routinely encounter them in both static figures [30, 27] and interactive phylogenetic visualizations [13, 4]. Collapsible tree layouts complement this technique by supporting drill-down navigation: users expand and collapse subtrees on demand, focusing attention on a chosen branch while keeping the surrounding context available [6, 8]. We draw on this body of work in selecting circular dendrogram and collapsible tree as the two primary layouts in our design, while extending standard implementations with adaptive layout selection, synthetic node markers, and explicit anomaly handling appropriate to the heterogeneous structure of paleoecological data.

3 Data and User-Centered Design Approach

Neotoma currently comprises 67,885 taxa organized across 52 taxon groups, representing over a century of accumulated fossil and environmental records including extinct taxa. Taxonomy in this context is not a static lookup structure but an ongoing collaborative process, and the gap between what is recorded in the database and what researchers need to see is where this design study begins.

We adopted a two-stage user-centered design process. In stage 1, we conducted semi-structured needs-assessment interviews with seven faculty-level paleoecology researchers whose expertise spans multiple paleoecological subfields, including fossil insects, mammals, diatoms, pollen, micropalaeontology, and plant-related paleoecology. All the participants are data stewards from Neotoma and closely allied data resources, meaning they are both experts in particular proxy types and have the necessary permissions to upload data directly into Neotoma. Interviews lasted approximately one hour and focused on how participants search, verify, and update taxa, trace synonym histories, and maintain cross-dataset consistency. Interview transcripts were analyzed using open and axial coding following a needs-based Grounded Theory approach, then mapped using Activity Theory to situate findings within systemic relations among subjects, objects, tools, rules, and division of labor [16, 17]. This qualitative research methodology provides a sociotechnical framework for analyzing how subjects (i.e., paleoecology researchers and data stewards) use visualization tools to support a variety of taxonomic workflows. In Stage 2, following the development of the beta interface, five think-aloud sessions were conducted with participants drawn from the same cohort, each following a structured four-task protocol covering taxon location, genus comparison, synonym resolution, and free group exploration. Each session informed revisions to the interface, thus reflecting an iterative process to HTV development.

4 Need Assessment Analysis

4.1 Structural Tensions

Activity Theory analysis of the interview transcripts surfaced four structural tensions that motivated the design objectives and defined the tasks that our taxonomy visualization tool needed to support:

Subject ↔ Tools: Scale mismatch at the point of entry. Researchers approach the taxonomy as individual subjects seeking specific taxa, but existing tools present all taxagroups uniformly regardless of size or depth. Taxagroups range from a handful of nodes to nearly twenty thousand, and users have no way to form a scale-appropriate mental model before entering a group.

Division of Labor ↔ Rules: Unconnected curation paths. Data stewards and contributors share curation responsibilities, but database rules do not require contributors to specify intermediate ancestors when adding a taxon. When a taxon is added without connecting it to the expected parent chain, the HigherTaxonID field

encodes a broken path that the system stores without flagging, and neither researchers nor stewards can identify it without tracing the full path context.

Community ↔ Object: Nomenclature drift versus static records. The paleoecology community operates under evolving nomenclatural codes, but the database retains records at the time of submission. Historical and deprecated names accumulate alongside current valid names, and a failed search cannot distinguish between a taxon that is absent from Neotoma and one recorded under a superseded name.

Subject ↔ Tools: Retrieval without relational reasoning. Existing interfaces treat taxon lookup as the terminal step, returning a record without supporting the relational questions that follow. Researchers routinely need to assess a found taxon's position relative to phylogenetic neighbors or compare it against another taxon of interest, tasks that currently require manual reconstruction across separate lookups.

4.2 Design Objectives and Tasks

We then abstracted the four user-derived tensions in taxonomic workflows into four overarching design objectives, each of which defining a task that our visualization design aims to support:

(DO1) Provide a scale-appropriate overview of heterogeneous taxagroups: The design should help users move from a high-level overview to detailed exploration without losing orientation. Users should be able to navigate from taxon types to scale-appropriate group overviews, forming an initial understanding of a selected taxagroup's scale and organization before committing to detailed exploration *(T1)*.

(DO2) Support taxonomic lookup under naming ambiguity: The design should help users connect queried names to the corresponding taxa without hiding the naming ambiguity. Users should be able to search by a taxon name, identify synonym-based matches, and inspect the related taxonomic change log *(T2)*.

(DO3) Support interpretation of taxonomic relationships after lookup: The design should help users reason about shared taxonomic context and lineage relationships between taxa. Users should be able to search for taxa by keyword, inspect matching results with their path context, and compare two selected taxa by identifying their lowest common ancestor, and divergence point *(T3)*.

(DO4) Enable inspection of structural irregularities arising from Neotoma's non-Linnaean taxonomic organization: The design should preserve sufficient path context to help both researchers and data stewards locate where these irregularities occur and inspect the affected taxa. Users should be able to locate and inspect structural artifacts within their taxonomic path context, assess the reliability of a taxon's placement and identify where path inconsistencies originate *(T4)*.

In summary, these tasks aimed to operationalize the design objectives in addressing the four tensions and guided the development of HTV described in Section 5.

5 Visualization Design

Neotoma Hierarchical Taxonomy Visualizer (HTV) integrates four coordinated interface components: a Taxagroup Navigation Panel (A) for structured entry into the taxonomy, a Primary Visualization Canvas (B) that adapts its layout to the scale and structure of the selected group, a Structural Artifact Inspector (C) that surfaces irregular and anomalous paths, and a Search and Comparison Panel (D) supporting synonym resolution and pairwise relational queries. Figure 1 shows the complete interface with labeled callouts. Each component addresses one or more of the tasks defined in Section 4.2, and together they constitute a coordinated response to the four design objectives. The interface is implemented as a client-side web application using D3.js.

5.1 Taxagroup Navigation Panel (A)

The navigation panel is the primary entry point into the taxonomy and is designed to make the categorical heterogeneity of Neotoma taxagroups visible before the user commits to any specific group. It organizes the 52 taxagroup codes across two classification levels. At the first level, taxagroups are divided into two broad categories: Biological Taxa (37) and Non-Biological Taxa (15). At the second level, the panel presents taxagroup codes organized within each broad category, allowing users to navigate from a major group overview into a specific taxagroup of interest. Within the Biological Taxa category, groups are ordered by broad clade affinity, so that related lineages such as Mammals, Herpetofauna, Fish, and Vertebrates undiff. cluster together and reinforce the user's developing mental model of the taxonomy before any visualization is rendered (DO1).

5.2 Primary Visualization Canvas (B)

The canvas is the central interactive space of the interface, which evaluates the size and structural character of the selected group and selects a rendering strategy accordingly. This adaptive behavior aims to directly addresses T1 and DO1 through anchor node assignment and visualization treatment based on taxagroup structure and size.

5.2.1 Anchor Nodes (B1)

Neotoma's raw taxonomic paths do not follow fixed rank assignments, so we made two structural interventions explicit in the visualization. First, anchor nodes mark the point at which a coherent subtree begins for each taxagroup and remain visible across both the major-group overview (B1) and the taxagroup-specific layout (B2), serving as a persistent reference point when users move between overview and detailed exploration. Second, we introduced synthetic anchor nodes when taxa that needed to be displayed together did not share a non-leaf parent in the extracted path set and would otherwise appear as scattered entries in the overview. In these cases, the synthetic node intercepts the paths and groups them into a labeled cluster. Each such node is explicitly marked "Synthetic" to indicate that it is an author-defined interface construct for navigation and readability. For example, Mammalia serves as the anchor node for the Mammals view, while Fish is a synthetic anchor node that groups Chondrichthyes and Osteichthyes. This rule can transfer to other hierarchical datasets where meaningful domain groupings are not consistently encoded in the underlying parent-child structure.

5.2.2 Circular Dendrogram for Large Groups (B2)

HTV chooses between circular dendrogram and collapsible tree layouts based on group size and branching structure. We use fifty displayed nodes as a practical threshold from prototype testing: at or below this scale, labels and parent-child links remained readable in a collapsible tree, while larger groups quickly produced clutter and required an overview-oriented layout. Taxagroups with more than fifty nodes therefore render as a circular dendrogram, a layout suited to deeply branched evolutionary hierarchies in bounded screen space [14]. Sibling nodes are ordered phylogenetically, and a custom angular separation function widens gaps between distinct super groups and tightens them within sister clusters, keeping higher-order structure perceptible at full scale. For very large groups such as Mammals, Vascular plants, and Insects, the canvas initializes in an overview-collapsed state. Only the anchor node and its immediate descendants are rendered on load; deeper subtrees are stored in a hidden child property and expanded on demand. Toggle buttons at the innermost circular depth are persistently visible, while buttons at greater depths appear only on hover, reducing visual noise at scale.

5.2.3 Collapsible Tree for Small and Flat Groups (B3)

Taxagroups with fifty or fewer displayed nodes render as a left-to-right collapsible tree. Flat or weakly hierarchical groups with at most two levels also use the same layout regardless of node count, since they function more like enumerated lists. In this layout, siblings are ordered alphabetically to support scan-based lookup, expansion and collapse transitions use coordinate caching to preserve visual continuity during interaction.

5.3 Search and Comparison Panel (C)

The Search and Comparison Panel handles tasks that cross from navigation into analysis, specifically locating taxa under naming ambiguity (T2) and comparing two taxa (T3).

5.3.1 Keyword Search with Path Highlighting (C1)

The panel accepts a free-text query and highlights the complete root-to-leaf path of every matching taxon in the canvas, so users can see the phylogenetic distribution of matches at a glance. For large groups such as Mammals, Insects, Diatoms, and Vascular plants, a focus-view optimization drops unmatched nodes from the scene, isolating matched lineages and preventing the surrounding structure from overwhelming the result (DO2).

5.3.2 Synonym-Aware Search with Changelog (C2)

The interface handles synonym relationships directionally. Searching a valid taxon name highlights its canonical node and opens a green Synonym Box listing all historical names with a changelog recording the date and reason for each revision. Searching an invalid or outdated name highlights the canonical node and opens an orange Synonym Box showing the current valid name and the same changelog. A reverse lookup table maps each invalid name to its canonical counterpart, keeping the tree structure intact while the full historical name space remains searchable (DO2).

5.3.3 Lowest Common Ancestor Comparison Widget (C3)

A comma-separated query of two taxon names or identifiers triggers Lowest Common Ancestor (LCA) computation within the currently rendered subtree. A floating lineage widget displays the shared root-to-LCA path and marks the divergence point. Both lineages are highlighted concurrently from the divergence point to their respective leaves, letting users trace the full structural difference between the two taxa in a single view (DO3).

5.4 Structural Artifact Inspector (D)

The canvas handles anchor and synthetic-root interventions inline, while a separate Structural Artifact Inspector reports records excluded from the primary rendering (T4, DO4). HTV flags two types of excluded records relative to the active visual root: path-level anomalies, where a recorded lineage conflicts with the biological constraints of the selected taxagroup, and unplaced or orphan records, where a lineage fails to pass through the anchor or synthetic cut-point used to construct the current view. These records are listed in an expandable Anomaly Bar, with each entry showing the taxon name and recorded path. This makes excluded records inspectable without distorting the main tree, helping stewards identify records that require curation and helping researchers assess whether anomalies affect the branch relevant to their question.

6 Think Aloud Studies

With an initial prototype of Neotoma HTV in place, we then conducted five think-aloud sessions with paleoecology domain experts over video conference to diagnosis user experience issues and iteratively refine overall usability and utility. Participants were drawn from the same cohort of seven paleoecology researchers for the needs-assessment interviews. Each session lasted approximately one hour and followed a structured scenario protocol comprising four tasks: locating a familiar species-level taxon, comparing two related genera, resolving a synonym relationship using both a valid and an invalid name, and free exploration of a large group.

The sessions produced task-level evidence of how participants used HTV. All five participants completed the familiar-taxon locating and synonym-resolution tasks, while four of five completed the two-genera comparison task. First, regarding navigation and structural orientation (DO1, DO4), participants navigated the two-level panel with minimal guidance, although some difficulties involved interpreting expandable versus terminal nodes and recovering orientation after moving between major-group and taxon-group views. Second, regarding search and comparison discoverability (DO2, DO3), participants successfully located and used the search and comparison functions, and requested support for comparing more than two taxa simultaneously. During free exploration, they also used search, filtering, and hierarchical navigation to inspect large taxonomic groups and identify records requiring closer review. Third, regarding synonym resolution (DO2), highlighting received consistently positive responses, with one participant describing the directional changelog as directly useful for pre-upload data submission workflows; a second participant noted uneven synonym coverage across taxagroups and suggested linking canonical nodes to external nomenclatural authorities. In addition, participants' questions about feature purpose and workflow sequence pointed to the need for stronger onboarding, leading us to add in-interface guidance that explains each feature and its role in exploration. By exposing taxonomic issues through hierarchical paths and sibling context, HTV helped stewards detect disconnected, duplicated, or inconsistently linked records that required correction or further review. For example, a search for *"Chrysomela"* within Insects similarly showed five separated branches of matching taxa, where the steward immediately noticed inconsistently listed entries. Similarly, a text search for *"Bembidion"* returned 213 matches in the filtered tree view, allowing the steward to inspect related taxa at a glance without checking separate SQL records one by one.

7 Conclusion

Representing the taxonomic structure of large paleoecological databases in a way that reflects both scientific practice and data complexity is an open challenge. Although HTV builds on established visual forms, its contribution lies in integrating these forms into a domain-grounded workflow for exploring messy, provenance-rich hierarchical data. Our experience yields three transferable lessons for deep hierarchical data that is heterogeneous in scale, structure, and provenance. First, adaptive layout selection driven by group structure can accommodate uneven branch depth and group size better than a single canonical layout. Second, rendering structural irregularities as explicit, labeled interface elements can make missing or inconsistent parent-child links inspectable, helping users distinguish data-level inconsistencies from principled visualization interventions. Third, directional synonym disclosure with provenance metadata can help users trace naming histories and assess the validity of records encountered in the literature. Beyond Neotoma, these lessons may transfer to other messy, crowdsourced, or legacy scientific hierarchies where scale, incomplete lineage, and inconsistent naming are common. For future work, we plan to expand synonym coverage by linking canonical nodes to external authorities, develop audience-specific modes for exploratory users and data stewards, and to conduct a summative evaluation with a broader participant pool to further assess trade-offs in usability versus utility.

Acknowledgments

This work was supported by the U.S. National Science Foundation under Award #2410961. We thank the participating experts for their time and insights, and the reviewers for their constructive feedback.

References

[1] AlgaeBase: Online information on algae. https://www.algaebase.org/. Accessed 2026. 2

[2] DiatomBase. https://www.diatombase.org/. Accessed 2026. 2

[3] Mikrotax: Online taxonomy of microfossils. https://www.mikrotax.org/. Accessed 2026. 2

[4] F. Asnicar, G. Weingart, T. L. Tickle, C. Huttenhower, and N. Segata. Compact graphical representation of phylogenetic data and metadata with graphlan. *PeerJ*, 3:e1029, 2015. 2

[5] M. Burch, H. van de Wetering, and N. Klaassen. The power of interactively linked hierarchy visualizations. *Journal of Visualization*, 25(4):857–873, Jan. 2022. doi: 10.1007/s12650-021-00818-3 2

[6] R. Dachselt and J. Ebert. Collapsible cylindrical trees: a fast hierarchical navigation technique. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.*, pp. 79–86. IEEE, 2001. 2

[7] V. Filipov, A. Arleo, and S. Miksch. Are we there yet? a roadmap of network visualization from surveys to task taxonomies. In *Computer Graphics Forum*, vol. 42, p. e14794. Wiley Online Library, 2023. 2

[8] Y. Gao, R. Patro, and P. Jiang. Collapsible tree: interactive web app to present collapsible hierarchies. *Bioinformatics*, 40(11):btae645, 2024. 2

[9] A. K. George, R. E. Roth, S. Widell, and J. W. Williams. Range mapper: An adaptable process for making and using interactive, animated web maps of late-quaternary open paleoecological data. *Open Quaternary*, 9, 2023. 2

[10] Global Biodiversity Information Facility (GBIF). Gbif taxonomy browser. https://www.gbif.org/occurrence/taxonomy?occurrence_status=present, 2025. Accessed: 2025-10-13. 2

[11] J. Gortler, C. Schulz, D. Weiskopf, and O. Deussen. Bubble treemaps for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):719–728, Jan. 2018. doi: 10.1109/tvcg.2017.2743959 2

[12] G. S. Kandlikar, Z. J. Gold, M. C. Cowen, R. S. Meyer, A. C. Freise, N. J. Kraft, J. Moberg-Parker, J. Sprague, D. J. Kushner, and E. E. Curd. ranacapa: An r package and shiny web app to explore environmental dna data with exploratory statistics and interactive visualizations. *F1000Research*, 7:1734, Nov. 2018. doi: 10.12688/f1000research.16680.1 2

[13] I. Letunic and P. Bork. Interactive tree of life (itol) v4: recent updates and new developments. *Nucleic Acids Research*, 47(W1):W256–W259, July 2019. doi: 10.1093/nar/gkz239 2

[14] A. Macquisten, A. M. Smith, and S. J. Fernstad. Evaluation of hierarchical visualization for large and small hierarchies. In *2020 24th International Conference Information Visualisation (IV)*, pp. 166–173. IEEE, 2020. 3

[15] K. McHenry, M. Bobak, K. Coakley, D. Fils, L. Gatzke, B. Zhang, R. Kooper, S. Richard, D. Valentine, I. Zaslavsky, A. Shepherd, and E. Lingerfelt. Geocodes, 2021. Version 0.9.9, 2026-02-19. 2

[16] M. J. Muller and S. Kogan. Grounded theory method in human-computer interaction and computer-supported cooperative work. *The Human–Computer Interaction Handbook*, 3, 2012. 2

[17] B. Nardi. Context and consciousness: Activity theory and human-computer interaction, 1998. 2

[18] J. K. Nelson, M. Anderson, B. Bills, M. Stryker, S. Crawford, J. Blois, S. J. Goring, and J. W. Williams. Neotomadb/explorer: Neotoma explorer version 1.2.0. https://apps.neotomadb.org/explorer, 2023. doi: 10.5281/zenodo.7921067 2

[19] J. K. Nelson, J. Blois, S. Goring, J. W. Williams, and S. Dominguez. User-centered design and evaluation of the neotoma paleoecology open software ecosystem. *International Journal of Digital Earth*, 17(1), 2024. doi: 10.1080/17538947.2024.2378822 2

[20] Open Tree of Life Project. Open tree of life taxonomy browser. https://tree.opentreeoflife.org/opentree/argus/opentree15.1@ott93302, 2025. Accessed: 2025-10-13. 2

[21] Paleobiology Database. Paleobiodb navigator. https://paleobiodb.org/navigator/, 2025. Accessed: 2025-10-13. 2

[22] G. A. Pavlopoulos, T. G. Soldatos, A. Barbosa-Silva, and R. Schneider. A reference guide for tree analysis and visualization. *BioData mining*, 3(1):1, 2010. 2

[23] S. E. Peters and M. McClennen. The paleobiology database application programming interface. *Paleobiology*, 42(1):1–7, Dec. 2015. doi: 10.1017/pab.2015.39 2

[24] C. L. Schoch, S. Ciufo, M. Domrachev, C. L. Hotton, S. Kannan, R. Khovanskaya, D. Leipe, R. McVeigh, K. O'Neill, B. Robbertse, S. Sharma, V. Soussov, J. P. Sullivan, L. Sun, S. Turner, and I. Karsch-Mizrachi. Ncbi taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020:baaa062, 2020. doi: 10.1093/database/baaa062 2

[25] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1):92–99, 1992. 2

[26] J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International journal of human-computer studies*, 53(5):663–694, 2000. 2

[27] J. Stiller, S. Feng, A.-A. Chowdhury, I. Rivas-González, D. A. Duchêne, Q. Fang, Y. Deng, A. Kozlov, A. Stamatakis, S. Claramunt, et al. Complexity of avian evolution revealed by family-level genomes. *Nature*, 629(8013):851–860, 2024. 2

[28] Strategic Environmental Archaeology Database (SEAD). Sead taxonomy browser. https://browser.sead.se/, 2025. Accessed: 2025-10-13. 2

[29] J. W. Williams, E. C. Grimm, J. L. Blois, D. F. Charles, E. B. Davis, S. J. Goring, R. W. Graham, A. J. Smith, M. Anderson, J. Arroyo-Cabrales, A. C. Ashworth, J. L. Betancourt, B. W. Bills, R. K. Booth, P. I. Buckland, B. B. Curry, T. Giesecke, S. T. Jackson, C. Latorre, J. Nichols, T. Purdum, R. E. Roth, M. Stryker, and H. Takahara. The neotoma paleoecology database, a multiproxy, international, community-curated data resource. *Quaternary Research*, 89(1):156–177, Jan. 2018. doi: 10.1017/qua.2017.105 2

[30] D. Wu, P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, et al. A phylogeny-driven genomic encyclopaedia of bacteria and archaea. *Nature*, 462(7276):1056–1060, 2009. 2

[31] Q. Zhu, U. Mai, W. Pfeiffer, S. Janssen, F. Asnicar, J. G. Sanders, P. Belda-Ferre, G. A. Al-Ghalith, E. Kopylova, D. McDonald, et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. *Nature communications*, 10(1):5477, 2019. 2