Diaforá: A visualization tool for the comparison of biological taxonomies

Lilliana Sancho-Chavarría[1], Carlos Gómez-Soza[1], Fabian Beck[2], and Erick Mata-Montero[1]

[1] Escuela de Computación, Instituto Tecnológico de Costa Rica
lsancho@tec.ac.cr, carlos.gomezsoza@gmail.com, emata@tec.ac.cr
[2] paluno, University of Duisburg-Essen, Germany
fabian.beck@paluno.uni-due.de

Abstract. We address the problem of visualizing differences between two versions of a biological taxonomy. Given the dynamics of the taxonomic work, taxonomists are often faced with alternative versions of a taxonomy that need to be reconciled. Nevertheless, visual comparison of hierarchies is an open problem that involves several difficult challenges in Visual Analytics. First, how to display not one but two possibly large taxonomies on a fixed-size screen. Second, how to highlight all differences between the two taxonomies. We present *Diaforá*, an interactive tool that infers and visualizes the differences. Automatic inference is achieved by incorporating taxonomy rules to identify operations such as merging, splitting, and renaming of taxa, among others. Highlighting of differences is accomplished by using the *edge drawing* technique, which has been enhanced with a number of features suggested by users of a prototype version. *Diaforá* has been implemented and tested with real world taxonomies such as Bryozoa and Annelida as well as with artificial taxonomies.

Keywords: Biological taxonomies · Information visualization · Hierarchy comparison.

1 Introduction

Herbaria, museums, and biodiversity conservation initiatives maintain local, regional or global records of species, which are constantly updated due to taxonomic revisions, the discovery of new species, and the need to complete or correct the recorded information. Biological taxonomies are structures in which species are classified hierarchically according to the system proposed by Linnaeus in the 18th century [15], [24], where living organisms are classified into a hierarchical structure that includes the following taxonomic ranks: *domain*, *kingdom*, *phylum*, *class*, *order*, *family*, *genus*, and *species*. Living organisms are classified into groups, for example, birds. From a computing perspective, each group is represented by a node in the hierarchy, which in turn corresponds to a taxon (taxa in plural). For instance, the human being has been classified as species *Homo*

sapiens, which belongs to the genus *Homo*, to the family *Hominidae*, and to the order *Primates*. In this example, each group designated as *Homo sapiens*, *Homo*, *Hominidae*, and *Primates* correspond to a taxon. Taxonomists analyze the phenotypic characteristics of species given a set of criteria that they consider valid, they classify the species and describe them through scientific peer-reviewed publications. For more than two centuries, taxonomic information was only printed and scattered around the world. Consequently, before the digital revolution, integrating taxonomies developed world wide was not even feasible.

Given the dynamic nature of the biological taxonomies, it is common for taxonomists to come across different versions, which they can correct by applying comparisons. Since taxonomies can be large, the comparison becomes challenging. International initiatives such as Catalogue of Life have recorded approximately 1.8 million species of macro organisms, although many taxonomists believe that the planet's biodiversity is approximately six times that amount.[3]

It is important to mention that in this work we focus on biological taxonomies and not on phylogenetic trees. The latter are also hierarchical classifications of living organisms, but show the evolutionary relationships between species that have a common ancestry, and provide information regarding the evolution of species.

Visualization and comparison between hierarchies has been a prominent research topic in information visualization [8, 21]. However, despite these efforts, taxonomists do not yet have visual comparison tools readily available to facilitate the curation of taxonomies. The visualization of large individual hierarchies on a screen is in itself a complex problem because of the amount of taxa involved and limited screen space available. Consequently, comparing two hierarchies is an even more complex problem due to information overload and cluttering. When comparing two versions of a taxonomy, T_1 and T_2, taxonomists must perform several domain specific tasks [23] such as identification of taxa that in T_1 are shown as a single group and in T_2 appear divided (splits), or conversely, taxa that in T_1 appear separately and in T_2 are grouped under the same taxonomic concept (merges), identification of taxa that are located in a different place within the hierarchy (moved taxa), or that appear with a different name (renamed taxa), or that are not in a version of the taxonomy (excluded), or that have been added (added). In our research, we have worked with expert taxonomists from several countries and realized that some taxonomists work with very large groups of species whereas others with small groups. Also, when comparing two versions of a taxonomy, taxonomists might need a global view of the differences before focusing on a smaller group of species, or they might prefer to inspect directly a target group. They also value to have statistical information of the changes.

We present *Diaforá*[4], an interactive tool that automatically computes the differences between two versions of a taxonomy (see Fig. 1). *Diaforá* shows changes through explicit representations that make the visual recognition more efficient.

---

[3] The exact number of species of macro organisms is unknown because it is estimated that only about 20% of them have been identified [3].

[4] The word *diaforá* stands for *difference* in Greek.

It uses the *edge drawing* method for hierarchy comparison and color codes to explicitly represent the changes between the two versions of a taxonomy. It introduces the concept of *visual target synchronization* through which, if a taxon is the focus of interest of the user, the corresponding taxon in the other version of the taxonomy will be moved into the user's visual space so that both taxa can be visually compared side-by-side. It also allows the reorganization of data by users' demand in order to avoid cluttering, and provides visual summaries of the comparison to quickly get an overall sense of the magnitude of the differences. *Diaforá* also supports data cleaning tasks by highlighting, for example, undefined names and other naming errors. It also provides numerical summaries of the taxonomies and of the comparison. The code and sample data are publicly available at https://github.com/lsanchoc/Diafora.
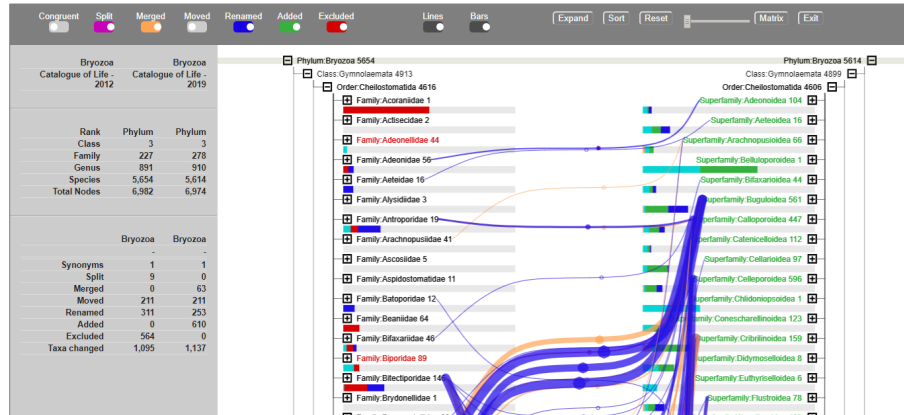

Fig. 1. An overview of Diaforá.

For testing, we first used artificial taxonomies in which we introduced representative cases of all types of changes in order to analyze alternative design features. We also tested with real public data from Catalogue of Life with taxonomies of up to approximately 15,000 species, which were displayed without noticeable lags.

This paper is structured as follows. Section 2 presents related work on hierarchy comparison and its application in the comparison of biological taxonomies. Section 3 introduces characteristics of the data. In Section 4 we describe the design requirements based on insights from previous research. Section5 presents the visualization design and interface. Section 6 explains some considerations of implementation and testing. In Section 7 we discuss results and lessons learned, and finally in Section 8 we present conclusions and future work.

2   Related work

The visual comparison of complex entities, whose complexity is due to multi-dimensionality and large number of components, is a common need in Visual Analytics [5]. It involves finding –visually– differences and similarities between objects of some domain and providing information for analysis. A comparison evidently involves a set of elements to be compared, which have specific characteristics that impose challenges; for example, the challenge of scalability. It also considers tasks of interest to the user, the strategies and methods to facilitate the comparison, and the selected visual design that allows an adequate visualization of the comparison.

Comparisons between hierarchies seek to find differences and similarities between information sets structured as trees. Differences can occur in the topologies and in the data associated with each node. Hierarchies can be represented in many ways [12]; however, not all of them are suitable for comparison. Hierarchy visualization can also consider multiple views [25]. Comparisons can take place between two trees or among multiple trees. Graham and Kennedy [8] extensively studied the visualization of multiple trees and summarized methods for comparing two hierarchies into five categories: *edge drawing*, *animation*, *coloring*, *matrix representation*, and *agglomeration*. On the other hand, Gleicher [6], defines three main types of comparison layouts, namely, *juxtaposition*, *superposition*, and *explicit encoding*.

The InfoVis 2003 contest focused on the visualization and pairwise comparison of trees [18]. From this contest TreeJuxtaposer [17] compares large phylogenetic trees and introduced an accordion-like distortion technique to support the concept of guaranteed visibility. Zoomology [27] took advantage of zoom techniques as well as overview and detail techniques to visualize the comparison. Further hierarchy comparison works have been reported in domains such as software evolution [11, 2, 20] and budget comparison [9]. In Biology, tree comparison has been directed to both phylogenetic trees [17, 19, 16] and biological taxonomies [27, 7, 14, 4]. Tasks for the comparison of biological taxonomies have been characterized [22] and show that taxonomists are interested in the identification of cases where taxa have been involved in splits, merges, moves or renames, or has either been added or removed from a version of a taxonomy. A study on methods for visualizing comparison and performing tasks for biological taxonomies comparison [21] indicates that taxonomists prefer the *edge drawing* representation method over the other techniques defined by Graham and Kennedy [8].

Edge drawing has been an issue in graph visualization because of cluttering when graphs are large [1]. Since trees are a special case of the graphs, they share similar concerns. Hierarchical Edge Bundles (HEB) [10] is a technique that nicely lays out edges while trying to reduce cluttering; however, it shows limitations [11] when it is required to distinguish individual relations among nodes.

3   Data

In simple terms, biological taxonomies are lists of taxa organized hierarchically where each inner taxon represents a category and each lower level taxon in the hierarchy represents a species. Global initiatives such as the Biodiversity Information Standards (TDWG)[5] and the Global Biodiversity Information Facility (GBIF)[6] make great efforts to standardize information in the databases and to promote sharing biodiversity knowledge. However, standardization remains an ongoing issue, because of the dispersion of information throughout databases in the world. Catalogue of Life (COL)[7] holds a comprehensive list of taxonomic information. It gathers about 1.8 million of species from about 168 databases, generates monthly and yearly versions of the lists, and provides open access to the information through JSON/XML/PHP-based web services.

The JSON format is lightweight and facilitates data exchange; however, since it uses textual labels, files could end up being heavy, and this can bring up to memory issues when working with large taxonomies. We reduced the label names to a one-character label (for instance, n stands for *name* and s for *synonyms*) and the labels are still easy to understand.

We downloaded taxonomies from COL of various sizes and of different years, so that we could compare different year versions. For each taxon, we obtained taxon name, taxonomic rank, source or author(s), date of publication, access date, the list of synonyms, and the list of descendants. These data fields are fundamental to run the inference algorithms and automatically recognize the differences between the two versions T_1 and T_2 of the taxonomy. It is not just enough to compare the taxa names when looking for differences. We can infer that two taxa refer to the same concept when the name of a taxon, the authors, and the year of publication are the same in both versions of the taxonomy. Synonyms play an indispensable role in the identification of changes because they link a taxon to its previous version. For example, we can recognize that a species x in the version T_1 of the taxonomy was split into three species p, q, and r, if x appears in T_2 as synonym of those three species. The more complete and accurate are the databases, the more precise would be the automatic inference of changes.

4   Design requirements

The problem to be solved is the visualization of differences between two versions of a biological taxonomy. It is necessary not only to recognize general-type changes (e.g., difference in size between the two hierarchies) but also what are the types of changes that occurred (e.g., if there were splits) as well as the specific changes (e.g., taxon x was split into p, q, and r). We consider insights obtained from previous research [21] as a framework of reference for the design of the tool.

---

[5] https://www.tdwg.org/
[6] https://www.gbif.org/
[7] http://www.catalogueoflife.org/

We synthesize the design requirements in terms of six aspects: the representation of the hierarchies, the comparison layout, the explicit representation of changes, multiple views, visual and numerical summaries, and efficiency.

- **Hierarchy representation.** Taxa names legibility is mandatory for taxonomists to analyze the taxonomies and to understand their differences. Thus, taxa names should be readable and visible at all times and the hierarchical representation should be such as to facilitate the reading of names. Compact representations of hierarchies, such as matrices, treemaps and icicle plots make an efficient use of space because nodes can be represented through a few pixels, so that large hierarchies can be displayed in a small area. They provide overview information as well as the possibility to identify patterns in changes. However, the space left to display the labels (that is, the taxa names) is so small that they are difficult to read or cannot be shown at all. The requirement of legibility of names lead us to consider indented lists as a design alternative.

- **Comparison layout.** Our framework of reference indicates that taxonomists preferred the *edge drawing* method over *matrix representation*, *animation* and *agglomeration*. Therefore, we consider a juxtaposed design that uses *edge drawing* as the central method to visualize the comparison. Taxonomies should be placed separately, side by side, to ease the comparison.

  In addition to representation, the role of interaction is key in information visualization. Interaction is also key in visual comparison [13] and it encompasses techniques such as: select, explore, encode, filter, connect and abstract/elaborate [26]. Given the potential number of relationships that could be visualized between the two taxonomies, selecting is necessary for users to study a type of change or a taxon of interest. By exploring, users should be able to examine a subset of the data (e.g., a family or a genus). Coding visual information into numerical information can allow users to quantify changes. Filtering is required to search for information that meets certain conditions (for instance, to know which species have been published by the same author). In the case of taxonomy comparison through the *edge drawing method*, connecting can be naturally implemented by the relationships among taxa, i.e., edges, which highlight changes between the two versions of the taxonomy.

- **Explicit representation of differences.** It is important for taxonomists to be able to recognize differences quickly and to clearly spot the *origin* and *destination* of changes. The *edge drawing* method fulfills this requirement, where edges take the leading role in the explicit representation of changes. In the main view, changes can be represented explicitly by colored lines that go from taxa in T_1 to taxa in T_2. The use of color for the explicit representation of changes is also very useful. All types of changes and all changes should be visualized by means of distinctive colors. In our design

color is used as following: pink for splits, orange for merges, brown for moves, blue for renames, red for exclusions, and green for added taxa.

In spite that edges make relations easy to understand, cluttering might be a problem. The hierarchical edge bundling technique was introduced to represent hierarchical graphs and reduce clutter [10], and then applied to the visual comparison of hierarchies [11]. Hierarchical edge bundling nicely packs together edges and, at the same time, the resulting visualizations provide overview information on changes. However, individual edges are hard to distinguish and, since the visual comparison of biological taxonomies requires clear recognition of origin and destination, the bundles should be such that this relationship does not get lost. The solution should reduce edge congestion and crossings but edges should clearly communicate individual relations, showing origin and destination.

- **Multiple views.** While the main method we propose for comparing taxonomies is edge drawing, in a previous work we found that expert taxonomists consider that comparison using matrices works better than edge drawing when it comes to globally overviewing changes and, that by combining several methods, the disadvantages of one method could be outweighed by the advantages of another. We propose a design in which global comparison is accomplished through a matrix representation and, by selecting an area of interest in the matrix, users can navigate to the edge drawing view where changes are visualized in more detail.

- **Visual and numerical summaries.** Visually identifying differences between higher level corresponding taxonomic groups is more complex than identifying differences at the species level because it requires more mental effort for users to summarize what happens at lower levels. Through summaries (visual and numerical) users could obtain information on the magnitude of changes at each taxonomic level.

- **Efficiency.** This involves effective and quick identification of differences as well as good performance in the visualization of large taxonomies. The first one is transversely addressed by the other five design criteria discussed here; for example, the more legible are the taxonomies and the more explicit the changes, the more efficient will be the identification of changes. The second one refers to how quickly taxonomies area loaded and to the response time during navigation.

5 Interface design

Fig. 2 illustrates the visual design of *Diaforá*. The window is divided into three panes. Pane 1 contains the main menu, pane 2 displays the numerical summaries of the comparison, and pane 3 is reserved for visualizing the comparison of the taxonomies.
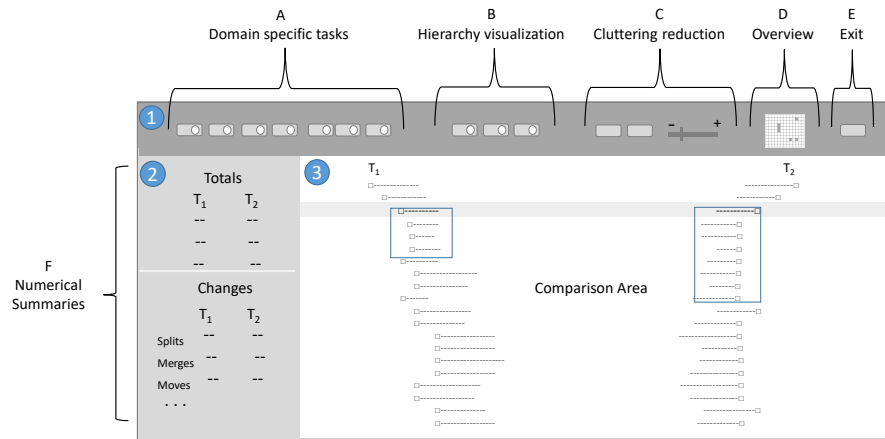
Fig. 2. The interface layout of Diaforá.

The main menu (pane 1) is divided into five parts. Part A contains seven toggle buttons to perform the domain specific tasks for the identification of similarities and differences; that is, for the identification of congruent taxa, splits, merges, moves, renames, added taxa, and excluded taxa. The toggle buttons allow users to visualize either one specific type of change or all changes. Part B contains buttons to control the hierarchy look; that is, a button to add or remove the hierarchy lines, a button to expand all sub-trees of a selected node down to the leaf level, and a button that would display the visual summary of changes next to each high-level taxon. Part C of the main menu includes functions to manage edge cluttering. It contains functions designed to reorganize the taxa within T_2, as close as possible to their related taxa in order to reduce edge cluttering. The menu also considers a reset button to return to the initial state of the taxonomies when they were first loaded. Another edge-managing function in this part of the menu is a slider control used to separate the edges when they appear too close together, making it easier to distinguish individual relations. Part D of the menu is reserved to navigate to an alternative overview, implemented through a matrix representation, that will be added to the tool in future work. Finally, part E for the menu is for exiting the system.

Pane 2 presents numerical summaries. The top part of the pane provides a summary on the structure of the taxonomies for each taxonomic rank; that is, it shows the amount of species, genera, families, orders, etc. of the two taxonomies being compared. The bottom part of the pane provides statistics on the amount of changes of the selected taxon; that is, it indicates the amount of splits, merges, renames, etc. that the comparison found.

Visual comparison takes place in pane 3, where hierarchies are placed juxtaposed, in a mirrored arrangement. Hierarchies are represented by indented lists where hierarchical relations are highlighted through edges. The edges are

optional to keep the comparison area as clear as possible. The mirrored arrangement is also a strategy to avoid crossings between the lines of the hierarchy and the edges. A drawback of node-link representations is the limitation on the number of nodes that can be displayed on the screen. To counteract, we make use of collapse/expand and zoom in/out mechanisms, which make it possible to enlarge and reduce the number of nodes in the visual space through interaction.

We synchronize the user selected taxon with the corresponding taxon in the alternative taxonomy; the alternative taxonomy moves either up or down, so that the two compared taxa are placed next to each other, highlighted by a horizontal grey line, in order to ease comparison (see Fig. 2). Taxonomists might want to understand what happened to either a specific species or to a group of species and might require to do visual searches, so navigation throughout the different levels in the taxonomy should be fluid. *Diaforá* lets users fully expand a branch of their interest. As users expand a selected branch, the corresponding changes of its descendants are refreshed.

The *edge drawing* method can clearly communicate the changes between the two versions of the taxonomy. The distinctive colors make it easy to recognize the different types of changes. Users can interact with the visualization and inspect changes that call their attention. Changes at the species level can be noticed by individual links between the involved taxa. Changes between higher-level groups (for example, between two versions of a genus or between two versions of a family) are displayed in two ways. One, for a higher level collapsed taxon, the amount of displayed links depicts the amount of differences detected between the two groups. The resulting thickness of the accumulated edges between the two compared taxa provides a cue on the magnitude of the changes for that group; although it has been limited to the height of the text. Second, changes between higher-level groups are summarized by bars that indicate the amount of changes proportional to the size of the group. This satisfies the design requirements on visual summaries (see Fig. 3). Visual summaries of changes are shown encoded as colored bars next to each higher level taxon. The assigned colors correspond to the previously discussed color coding for each type of change. The example illustrates the summary view when performing a comparison at *class*-level for the class *Clitellanata*. The pink color in the bar on the left refers to the amount of splitted taxa found in the *Clitellanata* group in T_1, and the green color in the bar on the right provides the amount of added taxa to the *Clitellanata* group in T_2. Additionally, the amount of species of each group is displayed next to each taxon name.

Cluttering of the edges is reduced in *Diaforá* in two ways: by grouping the edges and also by ordering the taxa. Edges are grouped by using a density-based spatial clustering algorithm (DBScan). We calculate a common central point where nearby edges are grouped. They are not bundled as in hierarchical edge bundling [10, ?] because in taxonomy comparison we always need to have the notion of origin and destination. The bundling we use can be controlled through the slider on the main menu. Users can also order the taxa through a commutation function (i.e, *Sort* function on the top panel). Commutation does

Fig. 3. Visual summary of the comparison between higher-level groups.

not alter the hierarchical relationships but makes edges that go to the same neighborhood to be grouped together in order to provide a cleaner visual space.

Computational efficiency when dealing with large taxonomies is approached by implementing a paging mechanism that loads into memory only the visible area of the screen, which contributes to the efficient management of pairs of large taxonomies.

6   Implementation and testing

Fig. 4 presents the implementation of *Diaforá*'s main window. It shows the comparison of versions 2012 and 2018 of the phylum *Cnidaria*. Notice how the thickness of edges provides a cue on magnitude of changes. In this case, the visualization indicates that many taxa were renamed (blue edges), also that new taxa was added (green names), some taxa were excluded (red names) and few taxa were merged (orange edge). The tool was implemented in Processing 3.4, HTML, database MongoDB, server Node.js and data files in JSON format.

**Fig. 4.** Example containing *Diaforá*'s main window.

We tested *Diaforá* with three pairs of taxonomies: Lycopodiopsida 2012 (158 species) VS Lycopodiopsida 2018 (1,415 species) for a total of 1,573 loaded, Marchantiophyta 2012 (773 species) VS Marchantiophyta 2018 (7,433 species) for a total of 8,206 species, and Annelida 2012 (12,635 species) VS Annelida 2019 (15,016 species) for a total of 27,651 species. All sets of taxonomies were loaded easily, smoothly visualized, and navigation and interaction were fluid. Further, we did profiling tests to know *Diaforá*'s usage of memory and CPU, and rendering times as taxonomies size increased. Our goal was to find out and extrapolate the impact of size on the tool's performance. We used the Google Chrome profiler tool. Testing results are presented in Fig. 5. Results indicate low increase on CPU and rendering times as taxonomy size increases. The displaying time is almost constant in all three cases; this is explained by the paging strategy where only the expanded taxa located within the boundaries of the screen are displayed. RAM presented the highest variation. Notice that, when comparing the Lycopodiopsida versions (1,573 species total), memory usage was 59.1 MB, for Marchantiophyta (8,206 species total) memory usage was 59.7 MB, and for Annelida (27,651 species total) memory usage increased to 140 MB. If we assume that the tool uses a constant amount of RAM for the browser and code, we can estimate that the increase from Lycopodiopsida to Annelida was 26.078 species and 81 MB, which represents a memory increase of 3.18 KB per specie approximately. We can use this number to extrapolate to any taxonomy size; for instance, a pair of taxonomies that sum up 100,000 species would require approximately 370 MB of RAM (that is, 59 MB + 3.1 KB * 100,000), which is a reasonable number. In other words, testing indicates that *Diaforá* tool is expected to perform well as taxonomies size increases.

7   Discussion

Diaforá contributes to taxonomists work in identifying differences and similarities between two versions of a taxonomy. Our work concentrates on the comparison of two taxonomies given that it is more likely that taxonomists perform pairwise comparison when looking for differences between a version that is familiar to them and a reference version. *Diaforá* was tested with pairs of taxonomies that together summed up to 27,651 species. It provides visual information for users to quickly recognize changes and it also provides numerical summaries on changes, which give information about the magnitude of the differences.

The data is a very important aspect to consider. We chose COL data because they provide access to taxonomic lists from several years, which makes it easier to have data in order to compare two versions of the same taxonomy of different years. Through the web services the complete taxonomy can be downloaded or the taxonomy can be obtained from a selected taxon, just as we did. As the COL database is constantly being updated and improved, it is likely that some detected changes are due to one of two reasons: either because the taxonomists made a redefinition or a new discovery of the species or because the new version of the database was completed with information not previously registered (for
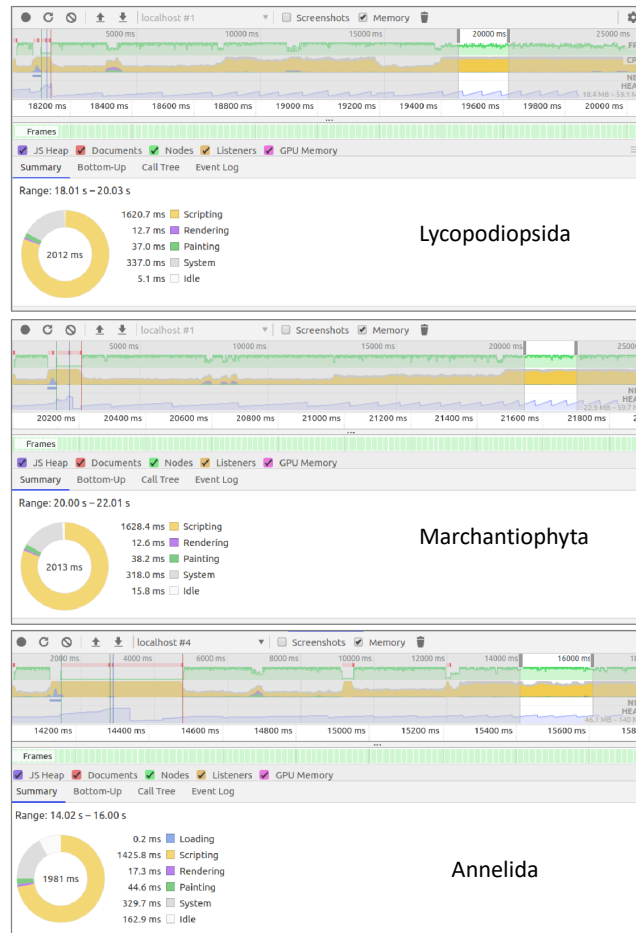
Fig. 5. Profiling tests.

example, a missing author was added or a name that was misspelled). In other words, the change detected could be due to the improvement of the database and not to a taxonomic reason. As a species is identified by the triplet (name, authors, year) a change in any of these data could be interpreted as a *rename*, despite what happened was an update of the data. Nonetheless, comparison is a way to discover inconsistencies and assist with data cleansing. It is also important to highlight that the precision of the inference algorithms for the identification of differences relies on the completeness and accuracy of the data.

Previous works on the comparison of biological taxonomies differ from *Diaforá* specially on the types of differences that the tool is capable of visualizing. We base our work on a set of required tasks for the curation of the taxonomies (i.e., identify splits, identify merges, etc.) whereas other works focus on the vi-

sualization of structural differences between the taxonomies [17], ancestor or descendants differences [14] or genus-corresponding species within other taxonomies [7]. We also provide visual cues on the magnitude of changes on higher-level taxa as well as numerical summaries of the comparison. We believe that the possibility to have both numerical and graphical information to depict summaries gives users valuable information for a quick identification of changes at an overview level. On the other side, the reorganization of taxa as well as the algorithm of grouping of edges makes the visualization cleaner for comparison.

8   Conclusions and future work

In this work we described a set of design requirements and proposed a visualization tool to solve the problem of visualizing and identifying differences between two versions of a biological taxonomy. We presented a tool that automatically infers the differences and highlights them through direct and explicit representation of changes. Both the visual representation of changes and the numerical summaries provide quick and valuable information to users.

Future work will be directed to perform several user studies and usability tests. We would like to test *Diaforá* with taxonomies that come from different organizations (different origins) and experiment with the data in order to get insights for further work. We also expect to add edit functionality to the tool, so that taxonomists would not only be able to quickly visualize the differences but also would decide which changes keep and which not, in order to support the curation process of a taxonomy.

9   Acknowledgments

References

1. Bach, B., Henry-Riche, N., Hurter, C., Marriott, K., Dwyer, T.: Towards Unambiguous Edge Bundling: Investigating Confluent Drawings for Network Visualization. IEEE Transactions on Visualization and Computer Graphics **23**(1), 541–550 (2017)
2. Beck, F., Wiszniewsky, F.J., Burch, M., Diehl, S., Weiskopf, D.: Asymmetric Visual Hierarchy Comparison with Nested Icicle Plots. In: Joint Proceedings of the Fourth International Workshop on Euler Diagrams and the First International Workshop on Graph Visualization in Practice. pp. 53–62. GraphVIP (2014)
3. CensusOfMarineLife: How many species on Earth? About 8.7 million, new estimate says. (2011), https://www.sciencedaily.com/releases/2011/08/110823180459.htm

4. Dang, T., Franz, N., Ludscher, B., Forbes, A.G.: ProvenanceMatrix: A visualization tool for multi-taxonomy alignments. In: Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data. CEUR Workshop Proceedings, vol. 1456, p. 13. CEUR-WS.org (2015)
5. Gleicher, M.: Considerations for Visualizing Comparison. IEEE Transactions on Visualization and Computer Graphics **24**(1), 413–423 (2018)
6. Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C.D., Roberts, J.C.: Visual comparison for information visualization. Information Visualization **10**(4), 289–309 (2011)
7. Graham, M., Craig, P., Kennedy, J.: Visualisation to Aid Biodiversity Studies through Accurate Taxonomic Reconciliation. In: Sharing Data, Information and Knowledge, pp. 280–291. Springer Berlin Heidelberg (2008)
8. Graham, M., Kennedy, J.: A survey of multiple tree visualisation. Information Visualization **9**(4), 235–252 (2010)
9. Guerra-Gómez, J.A., Buck-Coleman, A., Plaisant, C., Shneiderman, B.: TreeVersity: Interactive Visualizations for Comparing Two Trees with Structure and Node Value Changes. In: Proceedings Conference of the Design Research Society. vol. 1, p. 10 (2012)
10. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. In: IEEE Transactions on Visualization and Computer Graphics. vol. 12, pp. 741–748 (2006)
11. Holten, D., van Wijk, J.J.: Visual Comparison of Hierarchically Organized Data. Computer Graphics Forum **27**(3), 759–766 (2008)
12. Jürgensmann, S., Schulz, H.J.: A Visual Survey of Tree Visualization. Poster at Proceedings of IEEE Information Visualization, IEEE (2010)
13. von Landesberger, T.: Insights by Visual Comparison: The State and Challenges. IEEE Computer Graphics and Applications **38**(3), 140–148 (2018)
14. Lin, C., Qiao, H., Wang, J., Ji, L.: The Taxonomic Tree Tool, http://ttt.biodinfo.org/TF/
15. Linn, C., Gmelin, J.F.: Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis., vol. Tomus 1. Lipsiae: impensis Georg Emanuel Beer (1767)
16. Liu, Z., Zhan, S.H., Munzner, T.: Aggregated Dendrograms for Visual Comparison Between Many Phylogenetic Trees. IEEE Transactions on Visualization and Computer Graphics pp. 1–1 (2019)
17. Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L., Zhou, Y.: TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. ACM Transactions on Graphics **22**(3), 453–462 (jul 2003)
18. Plaisant, C., Fekete, J.D., Grinstein, G.: Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository. IEEE Transactions on Visualization and Computer Graphics **14**(1), 120–134 (2008)
19. Robinson, O., Dylus, D., Dessimoz, C.: Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. Molecular Biology and Evolution **33**(8), 2163–2166 (2016)
20. Rufiange, S., Melançon, G.: AniMatrix: A matrix-based visualization of software evolution. In: Proceedings - 2nd IEEE Working Conference on Software Visualization, VISSOFT 2014. pp. 137–146 (2014)
21. Sancho-Chavarria, L., Beck, F., Mata-Montero, E.: An expert study on hierarchy comparison methods applied to biological taxonomies curation. PeerJ preprints 7:e27903v1 https://doi.org/10.7287/peerj.preprints.27903v1 (2019)

22. Sancho-Chavarria, L., Beck, F., Mata-Montero, E., Weiskopf, D.: Visual Comparison of Biological Taxonomies: A Task Characterization. In: Poster submitted and presented at EuroVis2016 (2016)
23. Sancho-Chavarria, L., Beck, F., Weiskopf, D., Mata-Montero, E.: Task-based assessment of visualization tools for the comparison of biological taxonomies **4**, e25742 (2018)
24. Schuh, R.T.: Biological systematics: principles and applications. Cornell University Press (2000)
25. Teoh, S.T.: A study on multiple views for tree visualization. vol. 6495, p. 64950B. International Society for Optics and Photonics (2007)
26. Yi, J.S., ah Kang, Y., Stasko, J.: Toward a Deeper Understanding of the Role of Interaction in Information Visualization. IEEE Transactions on Visualization and Computer Graphics **13**(6), 1224–1231 (2007)
27. Young Hong, J., Richman, M., Westfall, M.: Zoomology: Comparing Two Large Hierarchical Trees. Tech. rep., Georgia Institute of Technology (2003)