Evolution: Education and Outreach

RESEARCH

Open Access

Building trees by juggling information and following rules: an expert interview study on tree-building and phylogenetic inference

Steven Tyrrell[1*] and Philipp Schmiemann[1]

Abstract

Tree-thinking is a fundamental skill set for understanding evolutionary theory and, thus, part of biological and scientific literacy. Research on this topic is mostly directed towards tree-reading—the umbrella-term for all skills enabling a person to gather and infer information from a given tree. Tree-building or phylogenetic inference as the second complementary sub-skill-set, encompassing all skills which enable a person to build a phylogenetic tree from given data, is not understood as well. To understand this topic we conducted think-aloud-tasks with tree-building experts and conducted supplementary guided interviews with them. We used school-like character tables, as they are common in high schools for the experts to build trees and audio-recorded their speech while building the trees. Analyzing the transcripts of the tasks we could find a basic methodology for building trees and define a set of backbone-skills of tree-building. Those are based on an iterative cycle going through phases of organizing information, searching and setting taxa/characters, organizing and checking oneself. All experts used simple guidelines, either deploying maximum parsimony to arrive at a solution or relying heavily on their previous knowledge. From that, we were able to utilize our result to formulate a guideline and helpful suggestions especially for beginners and novices in the field of tree-building to develop a better understanding of this topic.

Keywords Tree-thinking, Tree-building, Think-aloud-tasks, Expert-interviews

Introduction

Roots and branches basics of cladograms

Cladograms are a specific type of phylogenetic tree (Pietsch 2012; Baum and Smith 2013). All phylogenetic trees show how biological entities are connected through common descent (Baum and Smith 2013). There are different diagrams with cladograms being one of the simpler ones, relaying only the most basic information. There are many phylogenetic trees, with chronograms, chronodendrograms and evograms just to name a few (Pietsch 2012), but in this paper, we will only focus on cladograms. A cladogram and its basic parts are shown in Fig. 1.

Time strictly flows from the root to the taxa with the terminal nodes, regardless of the diagram's rotation (see Fig. 2) and only statements about relations in sense of a before and an after can be made—provided the instances in question lay on the same branches.

Statements cannot be made about events on different branches separated by a split (Baum and Smith 2013). This kind of diagram seems to be simple, but there are strict rules on how information can be gathered, and inferences can be made from it. At first glance, the cladogram shows the branching pattern and the taxa at the

*Correspondence:
Steven Tyrrell
steven.tyrrell@uni-due.de
[1]Faculty of Biology, University of Duisburg-Essen, D-45117 Essen, Germany

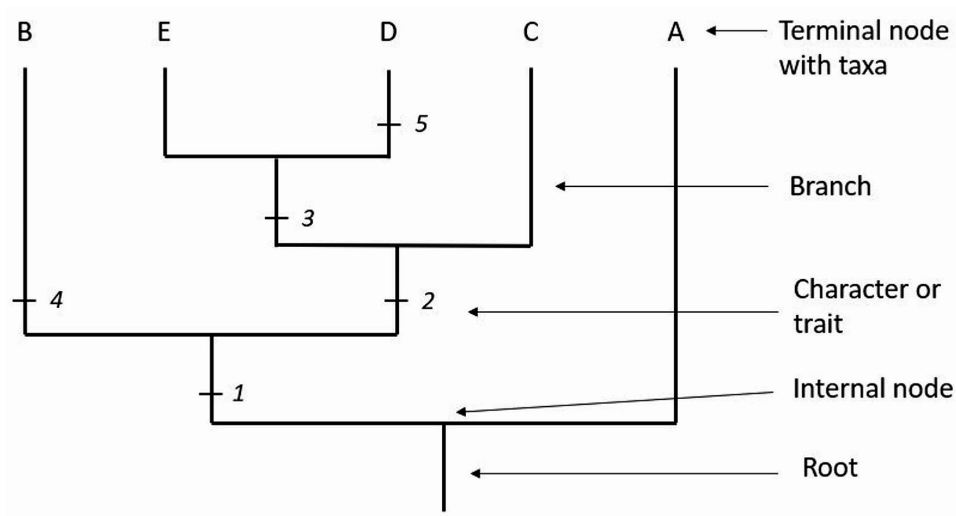

Fig. 1 Basic cladogram. Cladogram showing the taxa A-E on the terminal nodes and the characters 1–5 marked with dashes along the branches

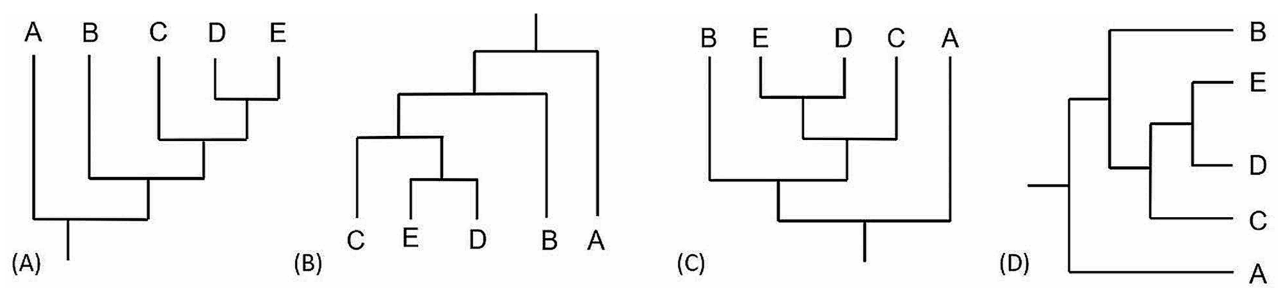

Fig. 2 Isomorphic trees. All trees in this diagram show the same information. Time flows from bottom to top (**A** and **C**), top to bottom (**B**) or left to right (**D**). The rotation of the tree or of the internal nodes does not change the topology of the tree and with that does not change the information conveyed

terminal nodes. It is common to indicate the gain and loss of traits or characters along the branches by dashes and annotations (Baum and Smith 2013). These dashes are meant to be read in an additive manner along a path—the taxa E and D in Fig. 1 have the traits 1, 2 and 3 and only D has 5 because 5 and E lie on different pathways read from the root to the terminal nodes. Trait 1 was acquired before traits 2, 3 and 5 and 2 was acquired before 3 and 5 and after 1, but no statements can be made about trait 4 in relation to 2, 3 and 5 because their position along different branches makes it impossible to confirm where they are positioned in relation to each other (Baum and Smith 2013). Position on the tips of the tree does not say anything about closeness of relationship, although it may suggest otherwise. The only informative feature in a cladogram regarding relationship is the branching pattern called the topology of the tree (Baum and Smith 2013). Moreover, trees can be rotated along the internal nodes without altering their informational value. All trees in Fig. 2 are isomorph, meaning they depict the same data.

Thinking about trees: tree-thinking and common misconceptions

A cladogram is a complex icon *(*Peirce 1903*)* with a high density of information, and a strict morphology must be learned to read it (Peirce 1903; Schnotz 2002). In general, all skills needed to handle phylogenetic diagrams could be subsumed under the term tree-thinking (O'Hara, 1997; Baum and Smith 2013; Catley et al. 2013). Tree-thinking is divided into two skill-subsets: tree-reading and tree-building. The former encompasses all skills necessary to read and infer information from a given tree, while the latter encompasses all skills necessary to build a tree from a given dataset (Halverson & Friedrichson, 2013; Schramm et al. 2019). Even with simpler tasks, this is an inference from data and the term phylogenetic inference is used to address this skillset (Kapli et al. 2020). Overall, there is a broader consensus about what skills are part of the tree-reading-set and an empirical tested skill-model exists (Schramm et al. 2019, 2021). There are elaborate theoretical considerations for skill models for tree-building. To our knowledge those models are not as empirically founded as those for tree-reading. Meir et al. (2007) named one tree-building-skill—reconstructing trees—which encompasses every activity when building a tree. Halverson (2011) suggested three skills for tree-building: Distinguishing, and using evidence, and communicating about a tree. Later Halverson and Friedrichsen (2013) formulated a hierarchical skill-model for tree-building with seven levels of expertise, starting with level 1 on which a person does not use any representation at all, level 2 on which a person paints pictures of evolutionary events and progresses to level 5 where they start to use hierarchical branching patterns to illustrate phylogenetic ideas. Level 6 and 7 are the stages on which a person can use representations in a scientific way and justifies and reasons with them. These models are quite helpful to distinguish between competent tree-builders but does not name certain activities when building a tree or imply in which order students should learn certain activities when learning to build phylogenetic trees or if there is a helpful order of progression at all.

Different problems can be described with tree-thinking (Gregory 2008; Meisel 2010; Phillipps et al., 2012). Novices like pupils and students have problems at certain points, and even teachers and experts can make incorrect assumptions about trees; there are many misconceptions about the informational value a given tree can provide (Lents et al. 2010; Catley et al. 2013). Most of them seem to be quite common. Gregory (2008) summed up ten mayor common misconceptions with tree-thinking but later research added more difficulties. The misconceptions Gregory (2008) compiled are for example misunderstanding recent taxa as ancestors of one another (for example: thinking of chimpanzees as ancestors of humans); thinking of evolution as an ever "higher"-evolving process (also called ladder-thinking, for example: thinking of birds as higher evolved than fishes); or confusing similarity in appearance with relatedness (for example: thinking of amphibians more closely related to fishes than mammals to fishes).

Schramm and Schmiemann (2019) added teleological thinking and essentialism as further obstacles when reading and building trees. Adler et al. (2022) highlighted that anthropomorphism and anthropocentrism are also highly problematic. More misconceptions are known and described. Additionally, there are hints that tree-thinking is influenced by many accompanying factors. Spatial intelligence and linguistic skills seem to be important as well (Catley et al. 2013; Schramm and Schmiemann 2019) and difficulties with cladograms may arise because they are very intricate forms of representations one must learn to use properly (O'Hara, 1997). Furthermore, evolutionary theory and cladograms are demanding subjects to learn. A novice must get acquainted with abstract and complex topics such as randomness and chance, heredity, variance within a species and much more. All these topics could be difficult to understand within themselves. As a complex concept, it is an obstacle for beginners and often trained users of cladograms, too (Adler et al. 2022).

Most studies regarding tree-thinking are targeted towards the United States. Although it is plausible to assume similar patterns in Australia, the UK or in the EU, it cannot be confirmed. The state of research on this topic could be more plentiful. Some researchers indicate that tree-reading should be taught initially (Halverson 2011). Others suggest that, first, tree-building should be taught when teaching novices (Julius and Schoenfuss 2006; Lents et al. 2010). Currently, this question does not have sufficient answers.

Relevance of tree-thinking

The relevance of cladograms and tree-thinking can be debated along three major lines of thinking: Phylogenetic trees, and especially cladograms, are an important type of diagram in modern biology and vastly used in schools and universities. They are not only used to gather and infer information of evolutionary events from a given tree, but also constructed from given datasets as a model of the understanding of the evolutionary descendance of a set of taxa (Baum and Smith 2013). Thus, phylogenetic trees are the most direct depiction of the core concepts of evolutionary theory (O'Hara, 1988; Baum et al. 2005) and with-it part of the backbone of modern biology (Dobzhansky 1973). Their importance cannot be overstated. The ability to adequately read information from cladograms and competently build them are considered as part of scientific literacy due to their close connection to the core of biology (Halverson and Friedrichsen 2013).

Consequently, cladograms and the abilities to handle them are given prominence in the school-curricula throughout the occidental world. Although cladograms and related skills are not directly mentioned for example in the NGSS (NGSS Lead States 2013), the ability to understand evolutionary data, derive meaning from it and find patterns are directly related to phylogenetic trees and the ability to understand them (Catley et al. 2013). Mostly on point seems "LS4A: Evidence of Common Ancestry and Diversity" (NGSS Lead States 2013; p. xxiv) and its iterations from pre-school to high school (NGSS Lead States 2013; p.187, 230, 272) which enforces the importance of evolutionary theory and understanding the branching patterns in descendance shown in cladograms. These skills are commonly taught in schools and universities using cladograms or other types of phylogenetic trees and they are found in most biology textbooks (Catley et al. 2013).

Moreover, the F-10-Curriculum in Australia (Australian Curriculum, Assessment and Reporting Authority (ACARA), 2010) states similar ideas, as does the National curriculum for England (Department of Education 2015).

This prominence emphases the value of understanding cladograms and building them in science education.

Aside from modern biology and the curricula of developed countries, there is a third line of thinking why the skills relating to phylogenetic trees are important. Topics adjacent to evolutionary theory are a part of regular discourse. The following examples illustrate where phylogenetic trees and the discussion of them are visible in daily life: In 2003, the remains of the Homo floresiensis and later, those of the Denisova human were discovered. Discussions around human evolution and the complex relationships in the human phylogenetic tree were present throughout mainstream media in the Western world. The cases of the "Linköping Doppelmord" and the "Golden-State-Killer" were both solved using phylogenetic methods and trees and generated a fair amount of attention. Moreover, daily in the last three years, phylogenetic trees, methods and interpretations based on inferences from phylogenetic trees were visible throughout the Sars-COV-2-pandemic.

Methods

Research question

Considering this, we want to elucidate the landscape of tree-thinking and focus on tree-building and the activities used to infer models from a specific dataset. We aim to obtain a better understanding of the basics of tree-building-activities. Therefore, the following research-questions are proposed:

RQ1: How do experts build phylogenetic trees from character tables?

RQ2: Which subskills of tree-building-activities can be derived from this process?

The findings may build the foundation for future research and give us more insights into this topic.

Participants, methodology and interviews

We used targeted experts and think-aloud-tasks with supplementary interviews (von Soest 2023; Kallio et al. 2016; Ericsson and Simon 1980) as the foundation because this method seems most appropriate to understand how experts think about a topic. Meuser and Nagel (1991), and von Soest (2023) argue, that the guided interview is the best methodology to raise data from experts. As even teachers seem to have difficulties with tree-thinking (Lents et al. 2010), we narrowed our sample to experts working at universities or similar institutions who specialized in phylogenetic disciplines. This adheres to the definition of expert by Meuser and Nagel (1991). They state that experts are defined by privileged access to knowledge and information and are often multipliers in educational processes. We wanted to observe experts while they were building trees to get a glimpse of their procedural knowledge. In addition, we explicitly asked experts for their advice for novices learning basic phylogenetic inference, to gather accompanying data.

We settled on the following two folded approach which can be seen in Fig. 3: We adapted the instrument from Schramm et al. (2021) and reverse engineered it into two character-tables for the two think-aloud-tasks (Eccles and Arsal 2017; Charters 2003; van Someren et al.,1994): One smaller task to get acquainted with the method and the other with more taxa and traits to gather the bulk of the data. The character tables were built to resemble tasks which are used in schools and the great clade race (Goldsmith 2003). Those tables usually do not include homoplasy, the outgroup is determined and polytomies are seldom used. Then we constructed a guided interview with nine questions regarding the topics building trees, typical problems while building a tree, and the best method to teach tree-building (Kallio et al. 2016; Castillo-Montoya 2016).

Moreover, we raised demographic data of our participants and gave them a number as a pseudonym (Direnga et al. 2016). Our interviews revolved mainly around how our participants build trees, in what order actions are taken and how they tackle difficulties. The interviews and the think-aloud-tasks can be seen in the additional files 1, 2 and 3. The additional files 2 and 3 show different think-aloud-tasks. This is because with our first think-aloud-tasks, we naively used the phylogenetic tree of arthropods underestimating how familiar the most experts would be with it. Strong influences of previous knowledge in our sample were visible and we reached a point of saturation after only three interviews. Hence, we changed the character tables. We replaced all names and characters (so that the emerging tree would be isomorph to the trees of Schramm et al. 2021) based on the local legends of the Elwedritsche and Wolpertinger—both fantastical creatures—to eliminate the influence of previous knowledge and place a stronger focus on the tree-building (Schmiemann et al. 2017; Schramm et al. 2022).

Afterwards, we advertised in professional magazines addressing experts in the field of phylogenetics and writing to some in person; five experts wanted to work with us. Their demographic data can be seen in Table 1.

The interviews were done via Zoom or in person. We recorded the audio and transcribed the think-aloud-tasks into GAT-transcripts (Selting et al. 2011) and the interviews into edited transcripts. The volume of the audio-files and transcripts are displayed in Fig. 3. Approximately 5 h of spoken language which translated to more than 100 pages of transcripts were gathered.

The think-aloud-tasks were coded using MAXQDA 2022 (Release 22.7.0). We inductively generated a code-system with nine codes and coded the transcripts line by

Fig. 3 Flowchart of approach

Table 1 Demographic data

Demographic data	
Participants	5
Gender	3 female, 2 male
Age	25–35, 36–45, 46–55, 46–55, 56–65
Highest qualification	All PhD, two qualified as professors
Job	Two research associates, two working as professors, one working in a museum
Studied Biology	5
Actively teaching tree-thinking related topics	3

Demographic data of the participants

Table 2 Codes, categories, and times of their coding

Codes	Group	Times coded
Finding the most common trait	Performative Skills	50
Setting taxa / setting apomorphie		46
Parting similar groups	Administrative Skills	9
Deleting known rows / coloums		9
Using previous knowledge		21
Overview/ strategize		33
Setting outgroup		14
Deleting autapomorphies		12
Checking	Inspective Skills	28

Used codes can be seen in the left column, the times we used them over all documents in the right column. The middle column shows the group we placed the codes in

line, while more than one code could to be attached to a line. After we constructed our code-system, we referenced the codes with skills which were suggested in literature to further validate our findings (Schramm et al. 2019; Halverson and Friedrichsen 2013). After that, we double-coded more than half of the think-aloud-tasks. We grouped the codes in three categories and analyzed patterns in their appearance, frequencies and connections to each other. Subsequently, we coded the interviews paragraph-wise and analyzed the transcripts for accompanying factors on how to build trees, problems with tree-building and how to teach tree-building.

Results

Table 2 shows the nine codes coded while we analyzed the think-aloud-tasks.

The first, second, and third columns indicate the name of the code, the code groups and how often the codes were used, respectively. Overall, we coded 222 passages. Approximately 60% of the think-aloud-tasks were double coded by two coders. After discussing discrepancies and coding again the Cohens kappa was raised to a good level at Cohens K=0,77 (Landis and Koch 1977; Altmann, 1990).

Checking for connections between the codes we generated a code-landscape using the MAXQDA-tool of the same name with codes adjacent in a range of one line. Though it is barely readable in its original form, Fig. 4 displays a simplified version.

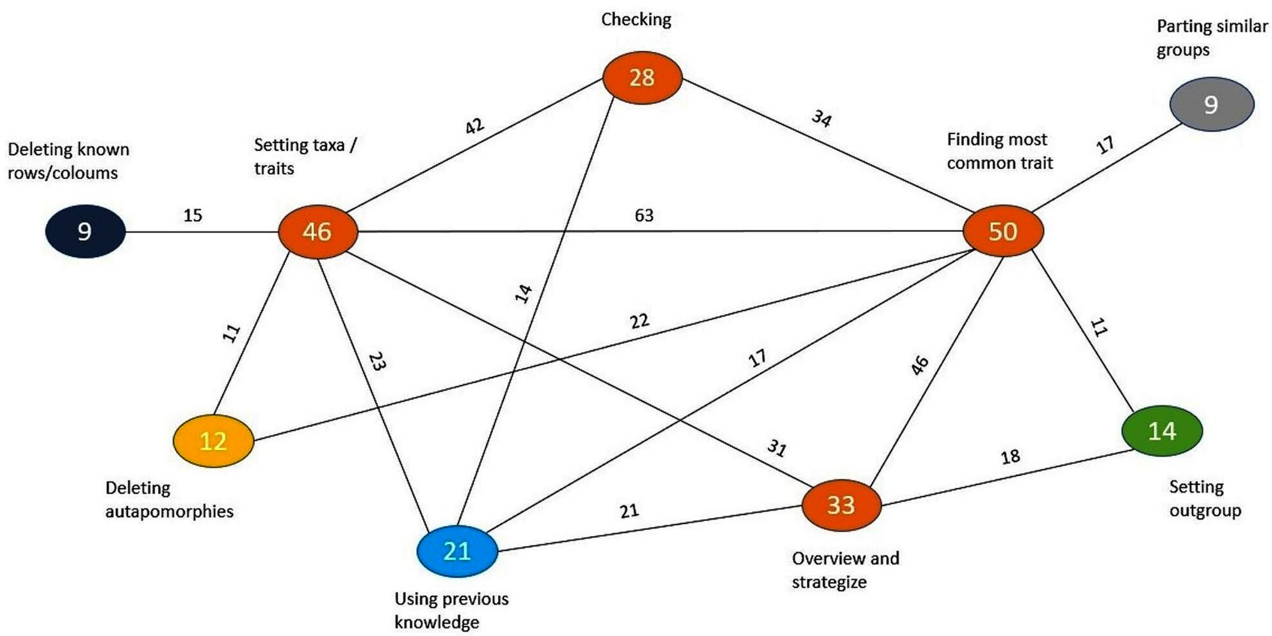

Fig. 4 Code-Landscape of the codes used during the coding with six clusters. Codes in the same color are in the same cluster. Numbers show the times they were coded. Numbers along the lines show the number of contacts of the segments within a range of one line

The map shows only the connections with a count of ten or higher within a range of one line. The frequencies of codes are shown within the bubbles and the number of connections is shown along the lines. The map is based on a cluster-analysis with six clusters. Clusters are marked in the same color. The original map alongside clustering with three to eight clusters are shown in code-landscapes. File 4 provides further details of the maps and the consistencies of the clusters.

Following this, we constructed a flowchart of the used skills by combing through the transcripts of the first and second think-aloud-tasks and drafting the order of skills used and combining them into one diagram. This diagram is shown in Fig. 5 with skills in black font and dark blue lined paths indicating those used by all experts. Skills in grey font and lighter lines show paths not used by all experts.

To get further insights into the performance of our experts, we generated a metricized document comparison chart shown in Fig. 6.

Discussion

Conclusions

The frequencies of codes and the code-landscape (Table 2; Fig. 4) both demonstrate a high frequency of our experts "searching for the most common character" and "setting taxa" accordingly. Both skills were used frequently by the participants and form a consistent cluster of up to seven clusters with "Overview and strategize." This does not seem surprising. To construct a phylogenetic tree, one must place taxa and characters in the tree. More surprising, all experts, except in one case, relied on finding the most common trait among the taxa as their instant approach. This means all experts used maximum parsimony to build their tree in the first try. This can be understood as an economical methodology used for the more simplified tasks. The one exception was found in GEHA092: The expert used the character table in a superficial way and relied on their previous knowledge about the cladogram of arthropods to place the taxa as shown in their code line (see Fig. 6, GEHA092). They underlined the first exclamation made when they saw the table and they lectured the interviewer during the construction of the tree about arthropods (see quotes 1 and 2 in quotes, additional file 6).

The code-landscape and the frequencies of codes in conjunction with the clustering hints at a backbone of four skills based on parsimonious principles: "Finding most common trait," "placing taxa," "overview and strategize," and "checking" which sits at the core of this tree-building task. By further analyzing the clustering of skills, "using previous knowledge" is a closely connected accompanying skill supporting that backbone. It falls in the same cluster as "Finding most common trait," "placing taxa," "overview and strategize," and "checking" up to five clusters. Examining the other skills, they do not seem to be as closely connected to the core loop of activities as the ones discussed above. With "deleting known rows/ columns" and "ignoring autapomorphies," both skills are devoted to organizing information. This might suggest that these skills are not as important, but we would argue otherwise. "Ignoring autapomorphies" was used by each

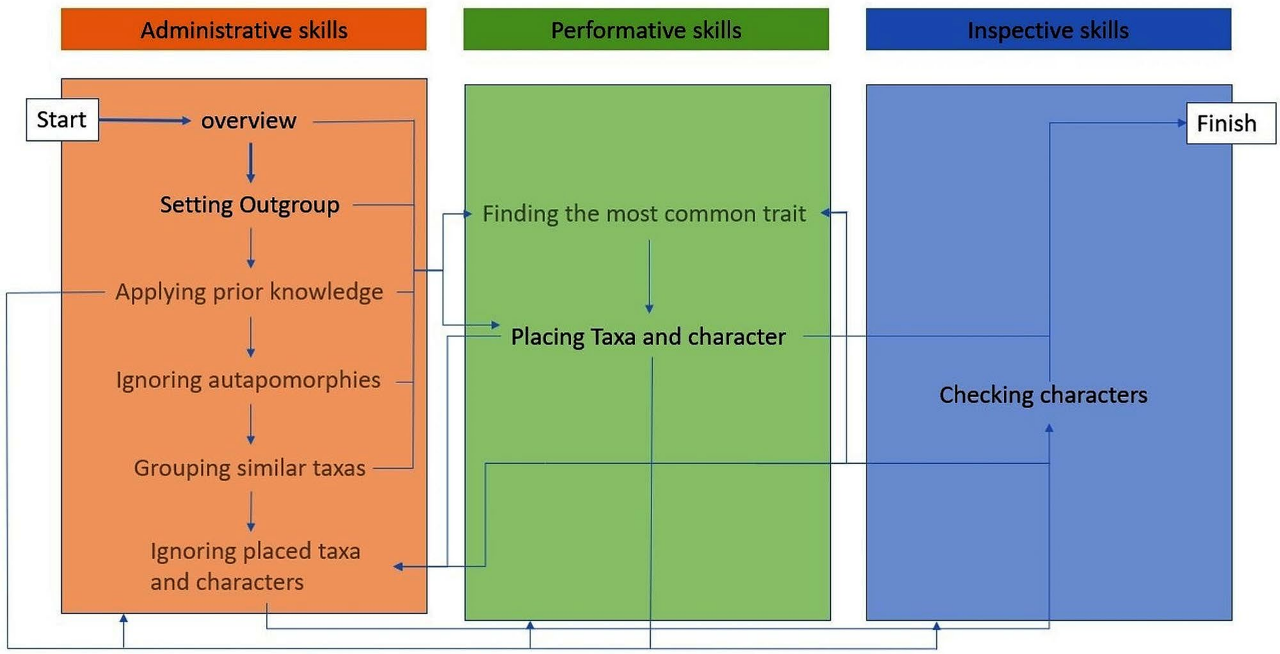

Fig. 5 Flowchart of the paths the experts used while solving the think aloud tasks. Black letters and darker arrows mark skills and paths all experts used while grey letters and lighter arrows mark paths not used by all experts. Arrows pointing to singular skills mark ways to this skill; arrows pointing at squares mark ways which can lead to any skill in that category. Lines can be read in any direction with blunt ends marking starting points and arrows marking endpoints

Fig. 6 Metricized document comparison table. The participants are shown on the left side (REWO07, ZOWO16, GEHA09, ADJO25, HEIJ19). The last number indicates the first or second think-aloud-task (1 or 2). The row above indicates the percentage of the documents. Red and yellow hues indicate administrative skill; green and blue hues indicate performative and inspective skills, respectively

participant in both tasks (except by GEHA09 as mentioned above) in the beginning or the middle of their task but was only used once or twice, diminishing its presence in the cluster. This skill may be used to reduce the informational load of the character-table with little time necessary. "Deleting known rows/columns" seems to work in a similar way: It is used by all experts in the second task (the more complex one) but was more oriented towards the middle or end of the process, suggesting it has some value in organizing oneself during the process of building the tree and/or checking it later. "Parting similar groups" is inconclusive. It was used by one expert in both attempts and one expert in the second task. The first expert dropped their attempt of neighbor-joining after a few, unsuccessful tries in favor of a maximum parsimonious approach. Later in the interview they stated quote 3 (in Quotes, additional file 6) explaining why they abandoned their approach. This implies that the approach is quite demanding in terms of time and handling informational load (Mayer 2014a; in: Mayer 2014b) and that they used their "intuition" to find the patterns connecting the taxa. We understand that as a nod to the underlying, subconscious routines based on experience and knowledge at work here. The second experts used it to get a grasp on the different taxa presented to them in the second task and to sort and organize them to reduce the number of taxa to be compared at a given time. This can be read as if "parting similar groups" can be used as an alternative approach to maximum parsimony or as a tool to reduce informational value – given the user has enough expertise to sort taxa appropriately.

The code-table and the flowchart (Table 2; Fig. 5) seem irregular. Most of the coded skills were devoted to organizing information and reducing the informational load one must handle simultaneously (administrative skills), while most activities where devoted to building the actual tree. The lack of a common path and order of operation and the interchangeability of at least some skills as shown above for "finding the most common trait" and "using previous knowledge" suggests that there is some kind of flexibility of approaches for building a tree. The many interconnections between the skills in the flowchart suggest the same conclusion.

A trend is visible when observing the document comparison table (Fig. 6). Usually the participants seem to start by mostly using the reddish administrative skill set, shifting more towards the greenish performative skills at the beginning and middle of the process and then towards blue inspective skills when finalizing their trees. Simultaneously, it can be observed that experts cycle through the different categories of skills while still maintaining the overall shift from administration to performance to inspective. Each expert cycles in the second task more through the categories than in the first, indicating that with growing complexity it is helpful to intersplice the (performative) building of the tree with dedicated phases of organizing information and checking. This is most clear when observing ADJO25's line in Fig. 6; it switches the most between the different skill-categories and it follows a highly iterative pattern. With this observation it seems advisable that with the growing complexity of a given tree-building-task a tree-builder should switch continuously between searching for and placing taxa, checking their placement, and organizing their information at hand.

(1) Always draw and read from rectangular cladograms. Everything else muddles the information and makes it harder for novices to identify the relevant information.

(2) Reduce available information on the character-table in a first step. Set your outgroup on your tree and delete it from the table, delete all autapomorphies from the table and mark characters in different colors.

(3) Find the most common character and set the taxa and characters accordingly.

(4) Check your placement. Read out loud what you placed while following the time flow from the root to the terminal knots.

(5) Delete all finally placed taxa and characters.

(6) Repeat step 3 to 5 till all taxa and characters are set finally and all columns and rows on the table are deleted.

(7) Go back and place autapomorphies.

(8) Read the tree out loud one last time.

(9) Finish your tree.

Fig. 7 Algorithm for novices when building a tree

Suggestions for practical work in schools

Being teachers, we want to reflect our findings back into schools and consider its impact on pupils, based on the assumption that children in school do not have any or very little knowledge about evolutionary theory, especially phylogenetic trees, and preconceptions about both are plentiful (Meisel 2010; Philipps et al. 2012; Adler et al. 2022). We would base our suggestions on three major guidelines derived from the think-aloud-tasks and the interviews: Use maximum parsimony when teaching pupils at school, reduce and organize information whenever possible and focus on the core skills—refined skills and algorithms can follow. While the first two guidelines are based on the observations we made during the think-aloud-tasks as mentioned above, the third one is made as a concession due to the complex nature of tree-building and the inferences needed and our limited pool of participants (Schnotz 2002; Catley et al. 2013). Therefore, we propose an algorithm for novices when building a tree shown in Fig. 7.

We attempted to translate the behavior of our experts into a simple guide. Although most points are self-evident after discussing our findings, some quotes are offered from the interviews to underline steps 1, 2 and 5. Quotes 4 and 5 in Quotes, additional file 6 are underlining especially step (1) in Fig. 7, quotes 6, 7 and 8 in additional file 6 step (2) and quotes 9 and 10 in additional file 6 steps (2) and (5). It is implied in different papers that rectangular cladograms are easier to understand for novices than other types (Catley et al. 2013; Schramm and Schmiemann 2019). A similar thought based on their experiences with teaching tree-building to students is posed by ADJO25 in the guided interview. Both points are forming the basis for step (1).

Notably, most of the experts working with us revert to their previous knowledge. While this alone is not surprising, it indicates the major thoughtfulness a teacher should dedicate to the material they want to use in their classroom when teaching phylogenetic inference. Well-thought-out character-tables and trees with unusual taxa or even imaginary and fantastical creatures and plants, unknown to pupils, could potentially help to focus on the relevant parts of the activity—an idea reiterated by Schmiemann et al. (2017). This can support building a tree without the hinderance of applying previous knowledge. Trimming down a tree and rotating it to defy preconceptions can be a valuable tool, and additionally, could be used to undermine superficial strategies.

Finally, we must stress a previous point: All the upper tips are for pupils and novices with little or no exposure to evolutionary theory or cladograms. In Germany, this means our guide is directed to pupils aged 16 to 18. It is only for those starting with the basics of tree-building. We are aware that this methodology may be helpful for beginners; however, it is flawed and will produce mistakes. It is necessary to exchange certain parts and fully replace them as novices learn more and understand the concept of evolutionary theory.

Limitations

This study has certain limitations. First, the number of experts was limited. Only a small group of experts could be gathered and our ability to draw general conclusions is limited in this regard. Our data seemed to reach a saturation because of the lack of new codes after the third participant even with our new instrument and all our conclusions are based on intersubjective findings; however, we strongly feel that further research could be fruitful to broaden our view and refine our conclusions.

The think-aloud-tasks were flawed in two ways. We designed them to suit a school-like scenario to make assumptions about how to teach in schools and used too well-known phylogenetic groups. The problems stemming from these two facts are varied. The experts were not acquainted with this kind of task. Although they were able to solve them, they do not work with character-tables like pupils do. At present, those tables are rarely used and other algorithms than maximum parsimony and computers are employed. The emerging tips for pupils are alluding to a simpler and older way to solve those tasks. Similarly, our task includes a given outgroup, no homoplasy, and chosen characters for the tree. This is quite common in schools but unrealistic in phylogenetic courses or real situations. As one of our participants stated: "Only few people recognized that selecting the relevant traits is the most important selection you can make" (quote 11 in quotes, additional file 6). Furthermore, in the same interview: "Which problems may occur and how to solve them - that's the selection of traits especially when such problems arise like 'don't have it.' For example, birds. I have the trait with a tail and without. And then you have the tail color. What do you do with birds without a tail? What do you write in the table?" (quote 12 in quotes, additional file 6). With the instrument used, we eliminated an important point in tree-building: Choosing the data and reviewing it. Thus, the results are necessarily limited. This step should be the subject for further research. On the other hand, we were not able to fully eliminate the role of previous knowledge. With the fifth expert, we changed the tasks by replacing names and characters but most of our data is tainted. Especially visible in GEHA092 where the character table is not used at all and the participant was just relying on their knowledge about arthropods, but we could find hints of previous-knowledge-usage throughout nearly every task. This is a further incentive to continue with our work and raise more data with more refined instruments.

Regarding the need to gather more data, the guideline for pupils must be tested. The guideline is based on our results, and we did not test it in the classroom, which must be done.

Our results pose one last question that we were not able to answer. A sensitive reader may have noticed how the expert built the tree and our algorithm emerging from that data were both superficial. The algorithm does not use biological content knowledge. How this or representational competence plays into tree-building-capabilities is still unclear but we are aware of this blind spot and agree that both topics should be investigated thoroughly. Although the guideline for novices suggests that content knowledge is unnecessary for building trees, that is not the case. As stated above, the selection of the characters is a crucial step in the whole process and could not be achieved without the proper expertise. Additionally, one cannot be called a competent tree-builder without the ability to select data for a tree. Moreover, after building the tree (or letting a computer do it), the interpretation of the tree is a part of tree-thinking which is one of the most demanding tasks. One of our participants stated those ideas and we want to end with their words because they highlight precisely how experts and novices are different from each other in this field of expertise: "That doesn't mean necessarily that the results are correct, the PC just helps you to calculate it – nothing more. This means I have to take the tree after construction and check it. Is this possible? Is there something in the tree after putting in the traits where I have to say 'This is impossible!'" (quote 13 in quotes, additional file 6). Later in the same interview, on the same topic: "That doesn't mean, 'That's it and everything else is wrong.' But based on the chosen traits and the way I have coded them and calculated them this is the most probable solution. Not more, not less." (quote 14 in quotes, additional file 6).

Conclusion

In conclusion, we want to add some final thoughts and a perspective on our plans:

We were able to sketch out a set of basic skills and a methodology which experts in phylogenetics used when they were building a phylogenetic tree from a given character table. Although the skill set and methodology are not complete due to the simplified tasks, they can be a starting point for further research. Additionally, we derived a limited, but helpful guide for pupils in schools which must be proven in the classroom. The guide can form a basis for formulating materials for novices to learn and better understand tree-building. This will give us the opportunity to test our findings in schools and classrooms, and provides a starting point for developing evidence-based learning materials for novices.

Further on, we want to observe if our new instrument can eliminate the role of previous knowledge and which data it will introduce. It may be necessary to differentiate our instrument so that we can elicit those supposed seldom-used skills. As stated earlier, this report is not the end of our research on this topic; it is merely the beginning, and we believe that there are many worthy insights to gather in the field of tree-building.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12052-024-00204-1.

Additional file 1

Additional file 2

Additional file 3

Additional file 4

Additional file 5

Additional file 6

Acknowledgements

We want to thank our participants for their support in our study and their time and also Jannik Joshua Böhnisch, student assistant, for coding and transcribing the interviews.

Author contributions

ST and PS conceptualized and designed the study, the think-aloud task and the interviews. ST led the interviews. ST and his student assistant Jannik Joshua Böhnisch transcribed the interviews, compiled the MAXQDA2022-Dataset and coded the Data. ST wrote the manuscript with input and revising from PS. Both authors read and approved of the final manuscript.

Funding

This article was funded by the Open Access Publication Fund of the University of Duisburg-Essen, Germany. The Interdisciplinary Centre of Educational Research at the University of Duisburg-Essen provided financial support for proofreading.

Open Access funding enabled and organized by Projekt DEAL.

Data availability

The data is available from the first author on any reasonable request. Inquiries should be sent to his official email address. All additional files mentioned above are available as additional material with this article.

Declarations

Competing interests

The authors declare no competing interests.

Received: 18 March 2024 / Accepted: 7 June 2024

Published online: 22 June 2024

References

Adler IK, Fiedler D, Harms U. Darwin's tales-A content analysis of how evolution is presented in children's books. PLoS ONE. 2022;17(7):e0269197. https://doi.org/10.1371/journal.pone.0269197.

Altman DG. Practical statistics for medical research. Chapman and Hall/CRC texts in Statistical Science. CRC Press LLC; 1990.

Australian, Curriculum. Assessment and Reporting Authority. Australian Curriculum. 2010.

Baum DA, Smith SD. Tree thinking: an introduction to phylogenetic biology. Roberts and Company; 2013.

Baum DA, Smith SD, Donovan SSS. Evolution. The tree-thinking challenge. Sci (New York N Y). 2005;310(5750):979–80. https://doi.org/10.1126/science.1117727.

Castillo-Montoya M. Preparing for interview research: the interview protocol refinement Framework. Qualitative Rep. 2016. Advance online publication.https://doi.org/10.46743/2160-3715/2016.2337.

Catley KM, Phillips BC, Novick LR. Snakes and eels and dogs! Oh, my! Evaluating High School Students' Tree-thinking skills: an Entry Point to understanding evolution. Res Sci Educ. 2013;43(6):2327–48. https://doi.org/10.1007/s11165-013-9359-9.

Charters E. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Educ J. 2003;12(2). https://doi.org/10.26522/brocked.v12i2.38.

Department of Education. National curriculum in England: science programmes of study. 2015.

Direnga J, Timmermann D, Lund J, Kautz C. Design and Application of Self-Generated Identification Codes (SGICs) for Matching Longitudinal Data. In 44th SEFI Annual Conference. 2016. https://www.researchgate.net/profile/julie-direnga/publication/324208940_design_and_application_of_self-generated_identification_codes_sgics_for_matching_longitudinal_data.

Dobzhansky T. Nothing in Biology makes sense except in the light of evolution. Am Biology Teacher. 1973;35(3):125–9. https://doi.org/10.2307/4444260.

Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qualitative Res Sport Exerc Health. 2017;9(4):514–31. https://doi.org/10.1080/2159676X.2017.1331501.

Ericsson KA, Simon HA. Verbal reports as data. Psychol Rev. 1980;87(3):215–51. https://doi.org/10.1037/0033-295x.87.3.215.

Goldsmith DW. The great clade race. Am Biology Teacher. 2003;65(9):679–82. https://doi.org/10.1662/0002-7685(2003065[0679:TGCR]2.0.CO;2).

Gregory TR. Underst Evolutionary Trees Evol. 2008;1(2):121–37.

Halverson KL. Improving tree-thinking one learnable skill at a time. Evolution. 2011;4(1):95–106. https://doi.org/10.1007/s12052-010-0307-0.

Halverson KL, Friedrichsen P. Learning Tree thinking: developing a New Framework of Representational competence. Multiple Representations Biol Educ. 2013;7:185–201. https://doi.org/10.1007/978-94-007-4192-8_11.

Julius ML, Schoenfuss H. (2006). Phylogenetic reconstruction as a broadly applicable teaching tool in the biology classroom. J Coll Sci Teach, 7(35).

Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs. 2016;72(12):2954–65. https://doi.org/10.1111/jan.13031.

Kapli P, Yang Z, Telford MJ. Phylogenetic tree building in the genomic age. Nat Rev Genet. 2020;21(7):428–44. https://doi.org/10.1038/s41576-020-0233-0.

Landis JR, Koch GG. The measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159. https://doi.org/10.2307/2529310.

Lents NH, Cifuentes OE, Carpi A. Teaching the process of molecular phylogeny and systematics: a multi-part inquiry-based exercise. CBE Life Sci Educ. 2010;9(4):513–23. https://doi.org/10.1187/cbe.09-10-0076.

Mayer RE. Cognitive Theory of Multimedia Learning. In R. Mayer, editor, The Cambridge Handbook of Multimedia Learning (pp. 43–71). Cambridge University Press. 2014a.https://doi.org/10.1017/CBO9781139547369.005.

Mayer R, editor. The Cambridge Handbook of Multimedia Learning. Cambridge University Press; 2014b. https://doi.org/10.1017/CBO9781139547369.

Meir E, Perry J, Herron JC, Kingsolver J. (2007). College Students' Misconceptions About Evolutionary Trees. The American Biology Teacher, 69(7), e71-e76. https://doi.org/10.1662/0002-7685(2007)69[71:CSMAET]2.0.CO;2.

Meisel RP. Teaching tree-thinking to Undergraduate Biology Students. Evolution. 2010;3(4):621–8. https://doi.org/10.1007/s12052-010-0254-9.

Meuser M, Nagel U. In: Kraimer K, editor. Qualitativ-empirische Sozialforschung: Konzepte, Methoden, Analysen. ExpertInneninterviews - Vielfach erprobt, wenig bedacht: ein Beitrag Zur Qualitativen Methodendiskussion. D. Garz; 1991.

NGSS Lead States. Next Generation Science standards: for States, by States. The National Academies; 2013.

O'Hara RJ. Homage to Clio, or, toward an historical philosophy for Evolutionary Biology. Syst Zool. 1988;37(2):142. https://doi.org/10.2307/2992272.

O'Hara RJ. Population thinking and tree thinking in systematics. Zoolog Scr. 1997;26(4):323–9. https://doi.org/10.1111/j.1463-6409.1997.tb00422.x.

Peirce CS. A Syllabus of Certain Topics of Logic. A. Mudge & Son, printers. 1903.

Phillips BC, Novick LR, Catley KM, Funk DJ. Teaching tree thinking to College students: it's not as Easy as you think. Evolution. 2012;5(4):595–602. https://doi.org/10.1007/s12052-012-0455-5.

Pietsch TW. Trees of life: a visual history of evolution. Johns Hopkins Univ; 2012.

Schmiemann P, Nehm RH, Tornabene RE. Assessment of Genetics understanding. Sci Educ. 2017;26(10):1161–91. https://doi.org/10.1007/s11191-017-9925-z.

Schnotz W. Towards an Integrated View of Learning from text and visual displays. Educational Psychol Rev. 2002;14(1):101–20. https://doi.org/10.1023/A:1013136727916.

Schramm T, Schmiemann P. Teleological pitfalls in reading evolutionary trees and ways to avoid them. Evolution. 2019;12(1):1–14. https://doi.org/10.1186/s12052-019-0112-3.

Schramm T, Schachtschneider Y, Schmiemann P. Understanding the tree of life: an overview of tree-reading skill frameworks. Evolution. 2019;12(1):1–13. https://doi.org/10.1186/s12052-019-0104-3.

Schramm T, Jose A, Schmiemann P. Modeling and measuring Tree-Reading skills in Undergraduate and Graduate students. CBE Life Sci Educ. 2021;20(3):ar32. https://doi.org/10.1187/cbe.20-06-0131.

Schramm T, Jose A, Schmiemann P. How to classify your dragons. Am Biology Teacher. 2022;84(5):308–11. https://doi.org/10.1525/abt.2022.84.5.308.

Selting M, Auer P, Barth-Weingarten D, Deppermann A, Hartung M. A system for transcribing talk-in-interaction: GAT 2. Gesprächsforschung: Online-Zeitschrift Zur Verbalen Interaktion. 2011;12:1–51. https://nbn-resolving.org/urn:nbn:de:bsz:mh39-1668.

van Someren MW, Barnard YF, Sandberg JAC. The think aloud method: a practical guide to modelling cognitive processes. Academic; 1994.

von Soest C. Why do we speak to experts? Reviving the strength of the Expert interview method. Perspect Politics. 2023;21(1):277–87. https://doi.org/10.1017/S1537592722001116.

Publisher's Note