

Holophyly and associated concepts if the unknown is unclassifiable

YEGOR SHISHKIN

Laboratory of Protistology & Experimental Zoology, Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, Universitetskaya Embankment 7/9, 199034, Saint Petersburg, Russia

E-mail: yegor.a.shishkin@gmail.com ORCID: orcid.org/0000-0003-0287-1718

Abstract.—In natural sciences, the key criterion for proving the existence of an object (like a physical particle of a previously unknown kind, an atom of a previously unknown element or an organism of a previously unknown species) is its observation and/or observation of traces of the existence of this object (radiation, remains of vital activity etc.). Only objects that meet this criterion can be classified (*e.g.* entered in the periodic table of elements or described as a species). Single unknown organisms or species cannot enter the classification. Despite this, the current system of phylogenetic terms (holo-/monophyly, paraphyly, and polyphyly as they are currently defined) is not adapted to the separation of known and unknown organisms (as well as populations, species, etc.) including ancestral ones. There are longstanding confusion and controversy regarding these “phyletic states”. There seem only two ways in such a situation. The first way is to somehow directly include unknown ancestors in taxa, describe species for them and unavoidably to introduce at least one paraphyletic subtaxon during dividing each taxon. The second way is do not include unknown ancestors in taxa directly and amend the system of concepts and terms. Here the second way was followed and the possible definitions of the main phylogenetic concepts for the views of dealing only with known group members were proposed. Inability to provide a concise definition of holophyly using the existing terms indicates the lack of more basic concepts. These concepts were also proposed here and holophyly was defined using them near the end of the paper. The intersection of four basic “phyly” (*enophyly*, *merophyly*, *kollitophyly*, and *schizophyly*) results in the unambiguous triad of *holophyly*, *paraphyly*, and *schizophyly*. The definitions of the terms in this triad are believed to be unambiguous unlike the widespread definitions of holo-(mono-), para- and “polyphyly”. Here, many terms were defined using others in order not to make the definitions too cumbersome. Nevertheless, the “primary phyly” seem not less useful in phylogenetic discussions than the phyly of the triad. So also is the terms *inprestor*, *rendestor*, *ancessure*, *drade* and *skade* as well as the more precisely defined term *clade*. The first two terms seem to be necessary and important regardless of the views on the classification of unknown organisms.

Keywords: *rendestor*; *inprestor*; *enophyletic*; *merophyletic*; *kollitophyletic*; *schizophyletic*; *holophyletic*; *paraphyletic*.

INTRODUCTION

A remarkable feature of biological systematics and phylogenetics is that known descendants inevitably imply the existence of their ancestors, even if the latter are unknown. But this implication is not enough to “establish the identity” of each of unknown ancestors. This can be compared to the following analogy from the judiciary. If the existence of some illegal organisation with a leadership management system is proven, then it certainly has a leader. However, this data is not enough to bring any person to trial. As in the case of unknown ancestors, a “vacancy” is known here, but not a specific

identity. Despite all this, non-classification of unknown organisms contradicts the current system of phyletic states and creates confusions in it.

Hennig (1950; 1965; 1966) was probably aware that only known organisms could be classified. At the same time, he considered the probability of finding an ancestor of any group so low that he did not stipulate the inclusion of ancestors in any way in his definitions of phyletic terms. For example, his monophyletic group was defined as all descendants of one ancestor. Ancestors *de facto* were excluded from the groups. This theoretical inaccuracy led to the rapid inclusion of the common ancestor by other authors in the definitions of the three phyly. The inclusion was direct and unconditional. This lack of differentiation between known and unknown organisms has led to confusion among taxonomists (see Podani, 2010 for review). Probably, both of the main types of the vision of phylogenetic trees shown by Podani (2010) are not completely correct and the consensus seems possible and necessary (compare Figure 1 in Podani, 2010 and Figure 1 here).

Currently, there is some inconsistency in the current phylogenetic practice. The phylogenetic terminology does not specifically stipulate the separation of known and unknown organisms. The definitions are widely used, but users often do not pay attention to unknown organisms. For example, they divide one clade into two subclades, supposedly without a remainder and use the wording “last common ancestor of X” disregarding the unknown organisms, which can be the part of X. If we assume that unknown organisms are not considered members of groups *a priori*, then arises the problem of demarcation of holophyly and paraphyly from enophyly and merophyly respectively (see below). Regardless of the user’s views, the terms “last common ancestor” and “progenitor” are used like there are no unknown organisms at all. Logically, the ancestor/progenitor of the group can not be a member of that group (see Figure 2).

To be consistent we should either to somehow directly include unknown ancestors in taxa, describe species for them and introduce at least one paraphyletic subtaxon per taxon or not to include them directly and amend the terminological system. The differentiation between known and unknown organisms makes it possible to accurately and unambiguously define a number of phylogenetic terms, to avoid the problem of description of the unknown, and to make the cladistic systematics possible. However, this differentiation makes the changes necessary as well as the amendments in terminology. The revised and supplemented system of terms with definitions and comments is proposed below.

BASIC CONCEPTS

PHYLA (/ˈfaɪlə/; from Greek *φῦλον* [*phylon*] - tribe) — the ancestor plus all its descendants, or the set of a sole member having no descendants.

ENOPHYLETIC group (/ɛnə(ʊ)ˈfaiˈlɛtɪk/; from Greek *ενότητα* [*enótita*] - unity) — a set of known organisms, for which at least one phyla exists, all known members of which they are.

MEROPHYLETIC group (/mɪrə(ʊ)ˈfaiˈlɛtɪk/; from Greek *μέρος* [*méros*] - part, portion) — a set of known organisms, which have at least one common ancestor, and for which no phyla exists, all known members of which they are.

POLYPHYLETIC group (/pɒlɪfaiˈlɛtɪk/; from Greek *πολύς* [*polús*] - many, much) — a set of known organisms, which have no common ancestor; descendants of different ancestors (an exclusively theoretical concept in the present state of knowledge of life in the universe).

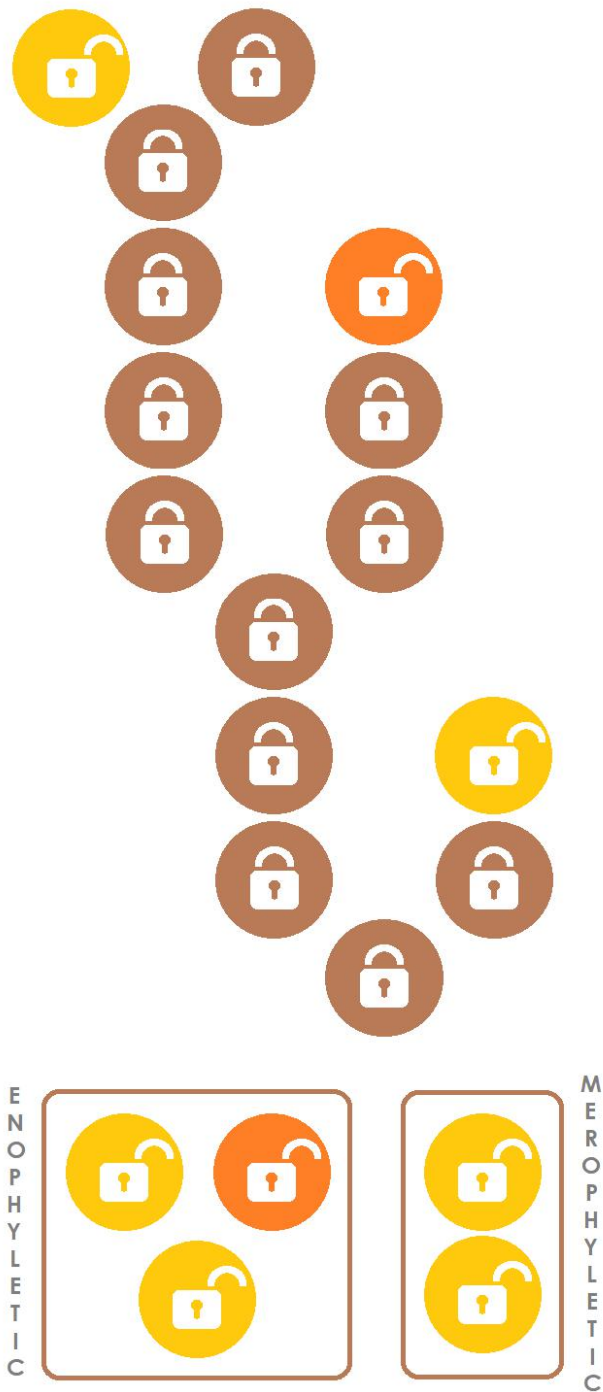


Figure 1. The tree of ancestor-descendant relationships. Circles with open locks represent known organisms, circles with closed locks represent unknown organisms.

There is the logical and semantical issue with both the wording “*last common ancestor*” and the word “*progenitor*” without adding “*of the known members of the group*”. The progenitor of a group can not be a member of that group. Likewise, an ancestor of a group can not be a member of that group. Therefore, they also should be *unable to be included* in the group (see Figure 2). The phrase “*My granny is the progenitor of my granny, my mother and me*” is not correct. So is this phrase if we put the word “*ancestor*” at the place of “*progenitor*”. The last common ancestor of Eukaryotes could not have a nucleus. Otherwise, it will be one of the eukaryotes but not their ancestor. The same is with the terms “*concestor*” (simple contraction from “*common ancestor*” coined by Nicky Warren and popularised by Dawkins (2004); see p. 7 in *op. cit.*) and “*cenancestor*” (originally coined by Fitch and Upper (1987)). This issue and the cumbersomeness of the wording “*the last common ancestor of the known the members of the X*” as well as un-pronounceability of possible abbreviation LCAKM were reasons for the more short term (*rendestor*) for it. The absence of the concept was the reason for the introduction of *inprestor*.

INPRESTOR of the group (/ˈɪnpɹɛstə/ from *in*; Greek *πρώτα* [*próta*] - first; and *ancestor*) — the first in history common ancestor of all known members of the group, which [the ancestor] is able to be included (or is included) in the given group.

RENDESTOR of the group (/ˈrɒndɛstə/; portmanteau from Fr. *rendezvous* - a meeting at an agreed time and place - and Eng. *ancestor*) — last common ancestor of all known members of the group except this ancestor itself if this ancestor is known.

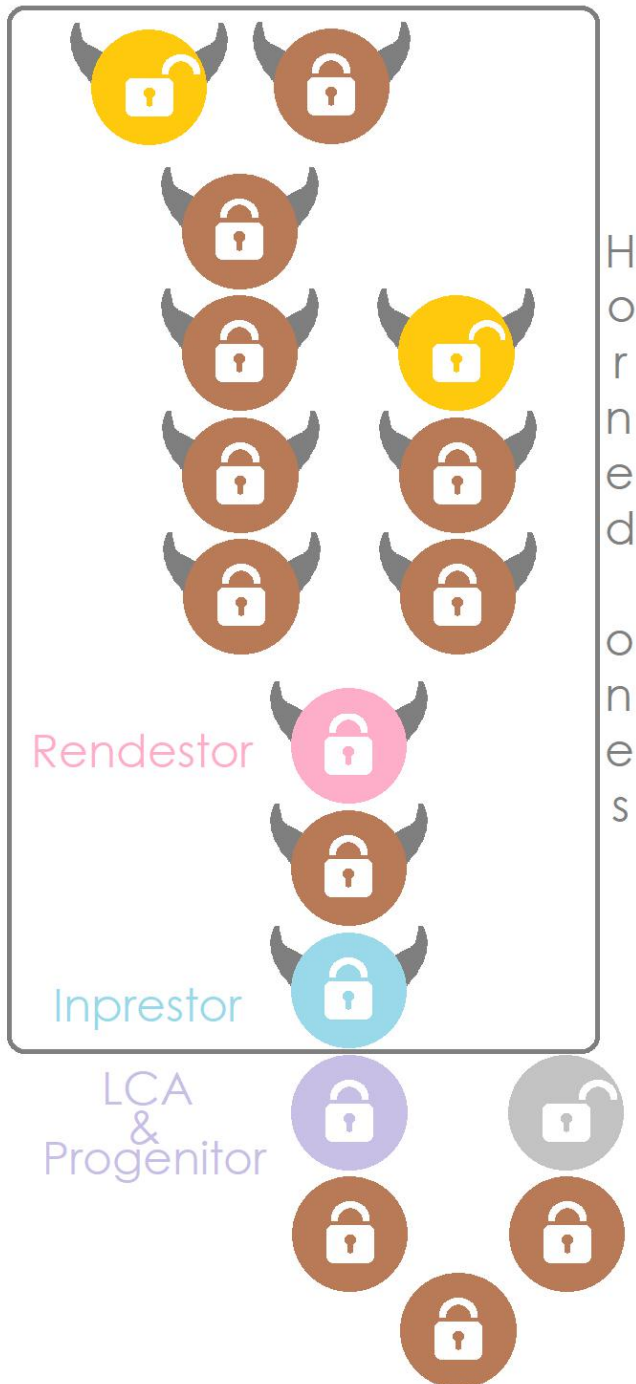


Figure 2. *The tree of ancestor-descendant relationships illustrating the different identities of rendestor, inprestor and last common ancestor (LCA) and progenitor. The grey frame is the borders of the group based on apomorphy of horns.*

As the reader could see, the basic criterion, which gives us *eno-* and *merophyly*, are not dealing with unknown organisms, including unknown ancestors. The latter are matter that can not be

completely divided into discrete units (*e.g.* could not be divided into a number of species). Although inside it some positions can be distinguished (such as the *rendestor*, the *inprestor* of a feature-based group (*e.g.* apomorphy-gainer), the *inprestor* of a max-clade). As in other cases the lineages of unknown ancestors can not be divided into discrete units and in each given moment of the time appear to humanity as solid functional units with the *rendestor/inprestor*, the common term seems highly suitable.

ANCESSURE of the group (/ansəs'juə/; portmanteau from *ancestor* and *commissure*) — the inprestor the group plus all inprestor's descendants, which also are ancestors of known members of the given group.

The word “commissure” was not used directly because, for example, the ancessure plus all members of the drade (see below) do not always represent the min-clade, *i.e.* it can contain “a tail” to the inprestor when the group have a definition other than that of min-clade or simple listing of the members (see Figures 2 and 3). An *ancessure* is about both known and/or unknown organisms. Using the term *ancessure*, it is possible to define the second basic dichotomy of the phyletic states.

KOLLITOPHYLETIC group (/kəlaɪtə(ʊ)flɪ'letɪk/; from Greek *κολλητός* [*kollitós*] glued) — a group of known organisms, which is able to include all members of its ancessure (according to the definition of the group).

SCHIZOPHYLETIC group (/skɪtsə(ʊ)flɪ'letɪk/; from Greek *σχίζω* [*skhízō*] - split) — a group of known organisms, which is unable to include all members of its ancessure (according to the definition of the group).

A kollitophyletic group has the potential to include unknown members of the *ancessure*, but does not include them before they are known. A kollitophyletic group is always able to include the *rendestor* (as it is always the part of the ancessure) and have the *inprestor*. Schizophyletic groups are unable to include the *rendestor* and have no single *inprestor*, although their parts have their own inprestors. Having the two basic phyletic dichotomies it is finally possible to define the duet of phyletic variants emerging at the intersection of kollitophyly with eno- and merophyly (see Figure 3). It is holophyly and paraphyly.

HOLOPHYLETIC group (/hɒlə(ʊ)flɪ'letɪk/ from Greek *ὅλος* [*hólos*] - whole) — a kollitophyletic group, the ancessure of which did not give rise to any known organism outside the group.

PARAPHYLETIC group (/pɛɪə(ʊ)flɪ'letɪk/; from Greek *παρά* [*pará*] - beside, near, alongside) — a kollitophyletic group, the ancessure of which gave rise to at least one known organism outside the group.

The term *holophyly* was preferred here, because of the number of meanings of the term *monophyly*, both definitional and etymological. Evolutionary taxonomists used it in the same sense as *kollitophyly*. Phylogenetic taxonomists used it in the stricter sense, probably the same as *holophyly*. As there was no differentiation between known and unknown organisms relatively to the phyletic terms before, the

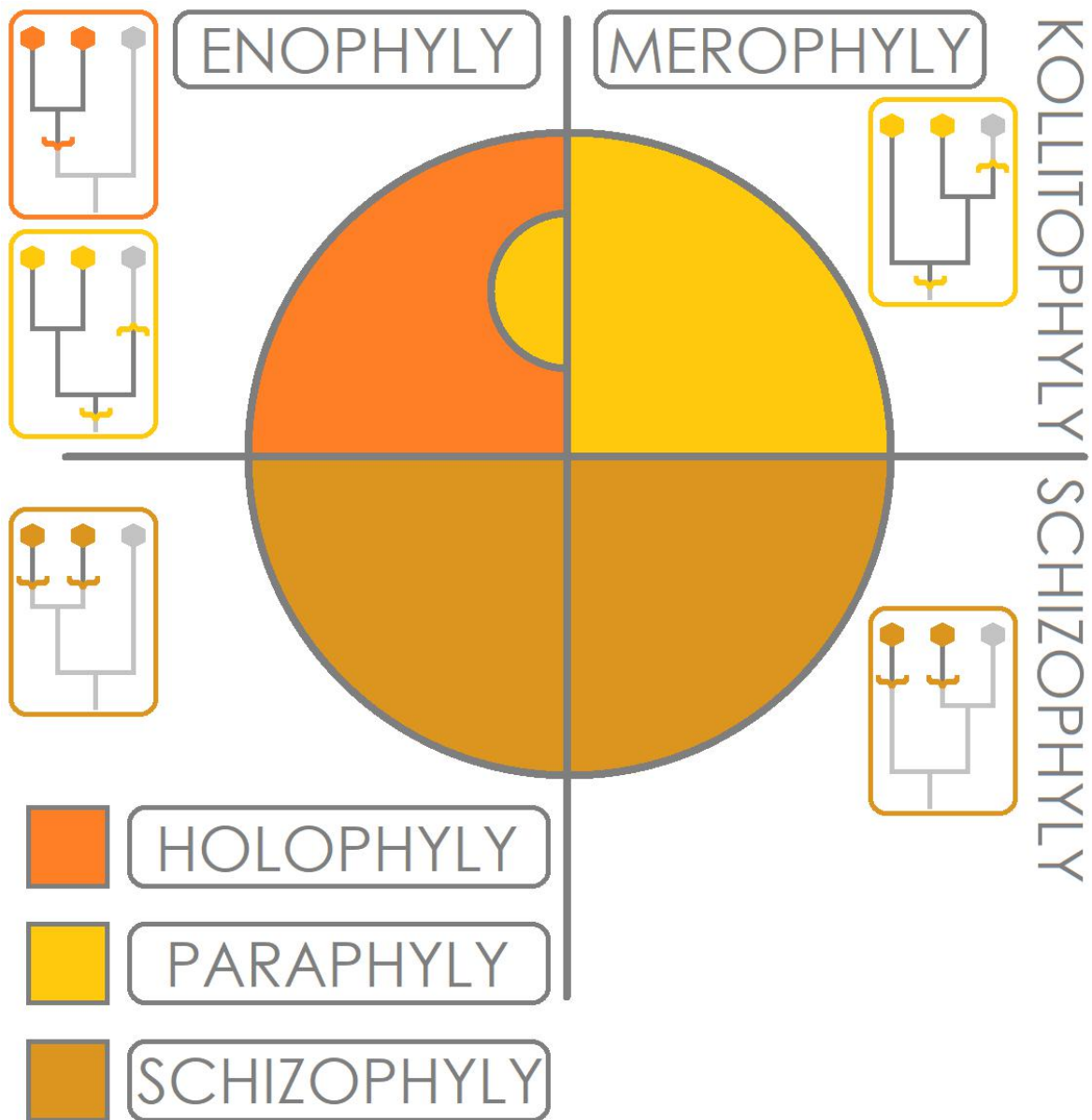


Figure 3. *The diagram of the relationships of “phyly” among themselves. Coloured hexagons represent known group members. Grey hexagons represent the known organisms outside the group. The groups: on all three trees on the left are enophyletic; on all two trees on the right are merophyletic; on all three trees above are kollitophyletic; on all two trees below are schizophyletic. The colour of the hexagons and borders around each tree matches the colour of the segment of the vertical half of the diagram.*

additional confusion arose. This confusion allows the term *monophyly* to be treated as a synonym of enophyly or to cover both eno- and merophyly. It is quite attractive to treat monophyly in the latter sense (“descendants of one ancestor”) in oppose to *polyphyly*, although it extremely reduces the term’s usefulness. Haeckel, who introduced the concepts *monophyly* and *polyphyly* (Haeckel, 1866), also used them as antonyms and, notably, did not reject the possibility of multiple origins of life (*e.g.* Haeckel, 1866 p. 198; 1868 p. 347; 1873 p. 371; 1894 p. 89; see Dayrat, 2003 and Rieppel, 2010 for review). In such views the terms *polyphyly*, as it was defined here, and *monophyly*, as its antonym, were applicable. Despite all this, it is possible to treat *monophyly* as the synonym of *holophyly* (as the latter was defined here).

The distinct entities of enophyletic and holophyletic groups raise the question of to what kind of groups the term *clade* should be ascribed. Although merophyletic groups now are not generally considered clades, there are still three options left of what the clade can be considered (see Figure 3): 1) an enophyletic group, 2) a kollito-enophyletic group or 3) a holophyletic group. As the etymology of the term *clade* (from Greek κλάδος [*kládos*] - shoot, branch) minds the inclusion of the ancessure and no breaks in the branch, here the term *clade* was applied only to holophyletic groups as they were defined here.

Such an interpretation of the term *clade* borns the issue with the groups formed by a simple listing of their members or subgroups. Such groups are unable to include their entire ancessures. For example, the wordings “Sar+Telonemia clade” or “Telonemia formed the clade with Sar” or “TSAR clade” will not be correct since the ancessure is not fully included here (see Strassert *et al.*, 2019 for the phylogeny). For such purposes, the short term for “*enophyletic* group” seems needed.

DRADE /dɹeɪd/ — an enophyletic group (from the two first letters of the semantic core in L. *polydrupa* - the type of fruit to which raspberries belong, bramble fruit; the type of aggregate fruit easily separable from the receptacle) and the ending *-ade*, like in the terms *grade* and *clade*).

CLADE /kleɪd/ — a holophyletic group (from Greek κλάδος [*kládos*] - shoot, branch).

The short term for paraphyletic groups also seems necessary there. The term [evolutionary] *grade* does not mean “a paraphyletic group”. Grades can also be holophyletic and schizophyletic as these terms were defined here (see Huxley 1957; 1958; 1959).

SKADE /skeɪd/ — a paraphyletic group (from Greek σκαλών [*skalón*] - flight of stairs).

CONCLUSIVE REMARKS

- 1) All known organisms are both descendants of one ancestor (universal rendestor) and descendants of different ancestors.
- 2) Regardless of the likelihood of knowing each specific ancestor, good terminology should be designed for the case when ancestors are found. The likelihood of this is very high near the terminals of the Tree of Life. However, the inclusion of unknown organisms (including unknown ancestors) cannot be direct and immediate.
- 3) To be holophyletic or paraphyletic, the group must be able to include the entire *ancessure* (see also Kwok (2010), who used the terms “connected group” and “disconnected group”, although he does not distinguish between known and unknown organisms). In other cases, we can only talk about *enophyly* or *merophyly*.
- 4) The demarcation between *enophyly* and *holophyly* is the Achilles’ heel of existing terminological systems.
- 5) The difference of holophyletic and paraphyletic groups from schizophyletic groups is the ability to include the entire *ancessure*.

6) The difference of holophyletic groups from paraphyletic groups is whether the *ancesture* of the group gave rise to any known organism outside the group.

7) Definitions of the term *clade* that do not use the separation of known and unknown organisms **do not allow any clade to be completely divided into subclades**—at least the *rendestor* will remain. This has been emphasised, for example, by Cavalier-Smith (1998 p. 211). Such definitions of the *clade* create problems in cladistic systematics. Cladistic systematics is the drive to constantly reduce non-holophyletic groups to only unknown organisms. However, such a definition of the clade (like the one given here for the *phyla*) is widespread and adopted, for example, by the *PhyloCode* (Article 2.1). Regardless of how acceptable we think paraphyletic groups are, cladism is possible only in the views adopted here and paraphyletic groups can be reduced to cases with known ancestors.

ORIGINS OF THE TERMS

The number of terms used here was introduced by other authors with the different definitions and/ or application. The terms *monophyly/monophyletic* and *polyphyly/polyphyletic* are originated from Haeckel (1866). The terms *paraphyly/paraphyletic* are originated from Hennig (1965). The terms *holophyly/holophyletic* are originated from Ashlock (1971). The terms *merophyly/merophyletic* are originated from Bernardi (1981). The term *clade* is originated from Cuénot (1940).

The terms *enophyly/enophyletic*, *kollitophyly/kollitophyletic*, *schizophyly/schizophyletic*, *drade*, *skade*, *inprestor*, *rendestor*, and *ancesture* are believed by the author to be new.

COMPETING INTERESTS

The author has declared that no competing interests exist.

REFERENCES

Ashlock, P. D. (1971). Monophyly and Associated Terms, *Systematic Biology*, 20(1), 63–69.

doi.org/10.1093/sysbio/20.1.63

Bernardi, N. (1981). Parentesco filogenético, grupo monofilético e conceitos correlatos: novas definições. *Revista Brasileira de Entomologia*, 25(4), 323–326.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biological Reviews*, 73, 203–266.

doi.org/10.1111/j.1469-185X.1998.tb00030.x

Cuénot, L. (1940). Remarques sur un essai d'arbre généalogique du règne animal. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 210, 23–27.

gallica.bnf.fr/ark:/12148/bpt6k31623/f24.item

Dawkins, R. (2004). *The Ancestor's Tale: A Pilgrimage to the Dawn of Life*. Boston: Houghton Mifflin.

Dayrat, B. (2003). The Roots of Phylogeny: How Did Haeckel Build His Trees? *Systematic Biology*, 52(4), 515–527. doi.org/10.1080/10635150390218277

Fitch, W. M. & Upper, K. (1987). The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology*, 52, 759–767. doi.org/10.1101/sqb.1987.052.01.085

Haeckel, E. (1866). *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie*. Erster Band: *Allgemeine Anatomie der Organismen*. Verlag von G. Reimer, Berlin. doi.org/10.5962/bhl.title.3953

Haeckel, E. (1868). *Natürliche Schöpfungsgeschichte. Gemeinverständliche wissenschaftliche Vorträge über die Entwicklungslehre im Allgemeinen und diejenige von Darwin, Goethe und Lamarck, im Besonderen über die Anwendung derselben auf den Ursprung des Menschen und andere damit zusammenhängende Grundfragen der Naturwissenschaft*. Berlin. Verlag von Georg Reimer.

Haeckel, E. (1873). *Natürliche Schöpfungsgeschichte. Gemeinverständliche wissenschaftliche Vorträge über die Entwicklungslehre im Allgemeinen und diejenige von Darwin, Goethe und Lamarck und Besonderen*. Vierte verbesserte Auflage. Berlin, Verlag von Georg Reimer. doi.org/10.5962/bhl.title.15259

Haeckel, E. (1894). *Systematische Phylogenie. Entwurf eines Natürlichen Systems der Organismen auf Grund ihrer Stammesgeschichte*. Erster Theil: *Systematische Phylogenie der Protisten und Pflanzen*. Berlin, Verlag von Georg Reimer. doi.org/10.5962/bhl.title.3947

Hennig, W. (1950). *Grundzüge einer Theorie der Phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin.

Hennig, W. (1965). Phylogenetic systematics. *Annual Review of Entomology*, 10, 97–116. doi.org/10.1146/annurev.en.10.010165.000525

Hennig, W. (1966). *Phylogenetic systematics*. University of Illinois Press, Urbana.

Huxley, J. S. (1957). The three types of evolutionary process. *Nature*, 180(4584), 454–455. doi.org/10.1038/2F180454a0

Huxley, J. S. (1958). *Evolutionary processes and taxonomy with special reference to grades*. pp. 21–39, in Hedberg, O. (ed). *Systematics Today*. Uppsala Universitets Arsskrift, Sweden.

- Huxley, J. S. (1959). *Clades and grades*. pp. 21–22, in Cain, A. J. (ed). *Function and Taxonomic Importance*. The Systematics Association, London.
- Kwok, R. B. H. (2010). Phylogeny, genealogy and the Linnaean hierarchy: a logical analysis. *Journal of Mathematical Biology*, 63(1), 73–108. doi.org/10.1007/s00285-010-0364-6
- Podani, J. (2010). Monophyly and paraphyly: A discourse without end? *Taxon*, 59(4), 1011–1015. doi.org/10.2307/20773972
- de Queiroz, K. & Cantino, P. D. (2020). *International Code of Phylogenetic Nomenclature (PhyloCode)*. CRS Press. doi.org/10.1201/9780429446320
- Rieppel, O. (2010). Ernst Haeckel (1834-1919) and the monophyly of life. *Journal of Zoological Systematics and Evolutionary Research*, 49(1), 1–5. doi.org/10.1111/j.1439-0469.2010.00580.x
- Strassert, J. F. H., Jamy, M., Mylnikov, A. P., Tikhonenkov, D. V. & Burki, F. (2019). New Phylogenomic Analysis of the Enigmatic Phylum Telonemia Further Resolves the Eukaryote Tree of Life. *Molecular Biology & Evolution*, 36(4), 757–765. doi.org/10.1093/molbev/msz012