Bionomina, **6**: 52–56 (2013)
www.mapress.com/bionomina/

Correspondence

ISSN 1179-7649 (print edition)
BIONOMINA
ISSN 1179-7657 (online edition)

http://dx.doi.org/10.11646/bionomina.6.1.4
http://zoobank.org/urn:lsid:zoobank.org:pub:8AC3C682-2767-4A8D-A1D9-9A015D5B5E03

The misleading use of the terms parent, child, ancestor and descendant in databases dealing with biological evolution and taxonomy

Alain DUBOIS & Mohamed BERKANI

Reptiles & Amphibiens, UMR 7205 OSEB, Département de Systématique & Evolution, Muséum national d'Histoire naturelle, CP 30, 25 rue Cuvier, 75005 Paris, France. <adubois@mnhn.fr>, <mohamed.berkani@gmail.com>.

The rapid burst of electronic databases in the recent decades has led to a boom in specialised terminology meant to describe the structure and way of functioning of these databases. Some of this terminology was borrowed from "common language" vocabulary but used in a specialized technical language. This decision to use existing terms in a new sense, rather than coining new terms for the new concepts, was a questionable one as in some cases this terminology may be misleading, at least in certain contexts. This is the case, in our opinion, with the use of terms expressing genealogical relationships to express hierarchical taxonomic relationships in databases.

There are various kinds of hierarchical relationships but a basic distinction is between inclusive or "whole-part" relationships and genealogical relationships. In an inclusive relationship, B is *included* in A, whereas in a genealogical relationship B was *produced* by A, two situations which have little in common.

In graph theory and other computer science contexts, in a directed graph A→B, A was soon designated as the "*parent*" and B as the "*child*". These terms are used in various contexts to indicate hierarchical relationships between items. After several decades of such a practice, many users feel that the meaning of these terms is "obvious" and clear to all readers. However, the indiscriminate use of the terms "parent" and "child" in all hierarchical databases tends to describe both kinds of hierarchical relationships as genealogical.

In hierarchical database models, data are organized following a tree-like structure. In this structure, each entity is related to several subordinate entities, which is known as one-to-many relationships (Kamfonas 1992; Groff & Weinberg 1999; Celko 2004). The descriptive term commonly used for such a structure using *1:N* mappings is "parent-child relationships", according to which each parent can have many children, but each child has only one parent. However, to any biologist, the latter sentence doubtless looks strange, because, in bisexual species at least, each child has two parents! But even more misleading is the fact that parent-child relationships (in the ordinary meaning) are genealogical relationships, which is not the case with hierarchical relationships as expressed in databases, even dealing with phylogeny (genealogy) or taxonomy (a conventional way of expressing genealogy). Besides taxonomic hierarchies, examples that can be mentioned include taxonomies of products, customer segmentations, organizational and location hierarchies (countries, states, departments, etc.), component breakdowns, etc.

A principle of "pseudo-inheritance" is used and well established in object-oriented databases (Garcia-Molina 2008). In such a context, each "piece" of data, or data object, is linked to subordinate data objects through attributes (or characteristics) and this results in a hierarchy of data. The analogy with phylogenetic systematics is obvious here, but it is misleading. It would appear then important in our opinion to use specific terms in databases dealing with taxonomic hierarchies in order to avoid semantic confusion.

Two pairs of terms are often used to express relationships of superordination and subordination in databases: *parent* and *child* refer to immediate such relations, whereas *ancestor* and *descendant* designate more remote relations (Su *et al.* 2006; Singh 2009). Both pairs of terms are misleading in evolutionary and taxonomic databases because they clearly evoke genealogical relationships although they do not refer to such relationships. The choice of the analogy with parent-child relationships to express such hierarchical relationships was doubtless a bad one. If "common language" terms had by all means to be used, a better choice would have been through an analogy with fruits or nuts, using "hull", "shell" or "capsule" for the superordinate item and "stone" or "core" for the subordinate one.

The inappropriate terminology is striking when the database at stake deals with evolutionary biology, phylogeny and taxonomy. Let us just take one example: in Alexey Shipunov's (2011) database *HHDB*, the term "*ancestry*" is used to provide "*data about the taxonomical position of the nominal taxon*". A nomen in this database appears under the format: "*Name: Scutiger Theobald, 1868; ancestry: Animalia: Chordata: Amphibia: Anura: Megophryidae; rank: genus*". The term "*ancestry*" in such a sequence refers to the descending hierarchy of less and less inclusive taxa, starting from the regnum **Animalia** and ending with the genus *Scutiger*. As this deals with biological taxa, this use is inappropriate. The taxon **Animalia** is not the "ancestor" of the taxon **Chordata**, but a taxon superordinate to it, i.e., that includes it. According to the point of view advocated by different authors, ancestors can be either species or, in a stricter approach, individuals, but they are not higher taxa, which have no offspring but which include entities like species and individuals that can have descendants. Hierarchical taxonomic relationships of successive inclusion are distinct from phylogenetic relationships of ancestry and descent.

In biological taxonomy, several patterns of *topotaxy*[1] can exist between taxa (see Fig. 1). In a taxonomic hierarchy, the relations of *superordination* and *subordination* occur in the topotaxic pattern of *peritaxy*[2] shown in Fig. 1c, in which the superordinate (including) taxon can be called *angiotaxon*[3] and the subordinate (included) ones *endotaxa*[4] (Dubois 2005).

We suggest that, in the terminology of hierarchical taxonomic databases, the terms *parent/child* and *ancestor/descendant* be replaced by the taxonomic terms *angiotaxon* and *endotaxon*. This would clarify the communication and avoid misunderstandings and confusions.

These relations of ordination can be immediate or proximate (the angiotaxon being directly superordinate to its endotaxa) or distant (the angiotaxon being separated from the endotaxa at stake by one or several intermediate endotaxa). This difference may be expressed by the use of the prefix *get-*[5] for immediate peritaxy—*parent/child* becoming *getangiotaxon/getendotaxon*—and the prefix *tel-*[6] for distant peritaxy—*ancestor/descendant* becoming *telangiotaxon/telendotaxon*. In Fig. 2, C1 is the getangiotaxon of its getendotaxa O1 and O2, whereas it is the telangiotaxon of all taxa included in O1 and O2, such as F1, F4, G3 or G7, which are its telendotaxa.

Whereas a given endotaxon has only one getangiotaxon, the number of getendotaxa of an angiotaxon is indeterminate. In a taxonomy based on a hypothesis of phylogenetic relationships (cladogram, phylogenetic tree), this depends on the degree of resolution of the tree, with pairs of sister-taxa or with polytomies. All the getendotaxa of the same angiotaxon are *parordinate*: they are immediately subordinate to the same superordinate taxon and, in a consistent taxonominal hierarchy, they must share the same level or nomenclatural rank (Dubois 2006*b*: 827, 2007: 33, 2008: 60; 2011: 89).

The use of these terms in taxonomic hierarchical databases would avoid terminological absurdities like those exemplified above in Shipunov's (2011) database. To be sure, it would not be appropriate in databases that do not deal with taxa. For such databases, the use of the new, more general, terms *angion* and *endon*, having the same etymology, respectively for superordinate (including) and subordinate (included) items would be appropriate[7]. It seems unlikely that these terms would be adopted easily by the community of

---

[1] From the Greek τόπος (*topos*), "place" and τάξις (*taxis*), "order, arrangement".

[2] From the Greek περί (*peri*), "around".

[3] From the Greek ἀγγεῖον (*aggeion*), "hull, capsule".

[4] From the Greek ἔνδον (*endon*), "inside of".

[5] From the Greek γείτων (*geiton*), "neighbour",

[6] From the Greek τῆλε (*tele*), "far from".

[7] A homonymous term *angion* is used in specialised medical literature to designate a segment of lymphatic vessel between two semilunar valves (Wittlinger & Wittlinger 2004), but confusion between the two terms is quite unlikely as they concern completely distinct fields.

computer scientists and database users, as the terms parent, child, ancestor and descendant have been in general use for several decades among them, and are well understood by all of them. But some of them at least might consider that the “private” use of a terminology within a community of “insiders” may be misleading for those who do not belong in this community, especially when this terminology is borrowed from common language—which is not the case when this community uses specific terms coined especially with the purpose of designating specific concepts.

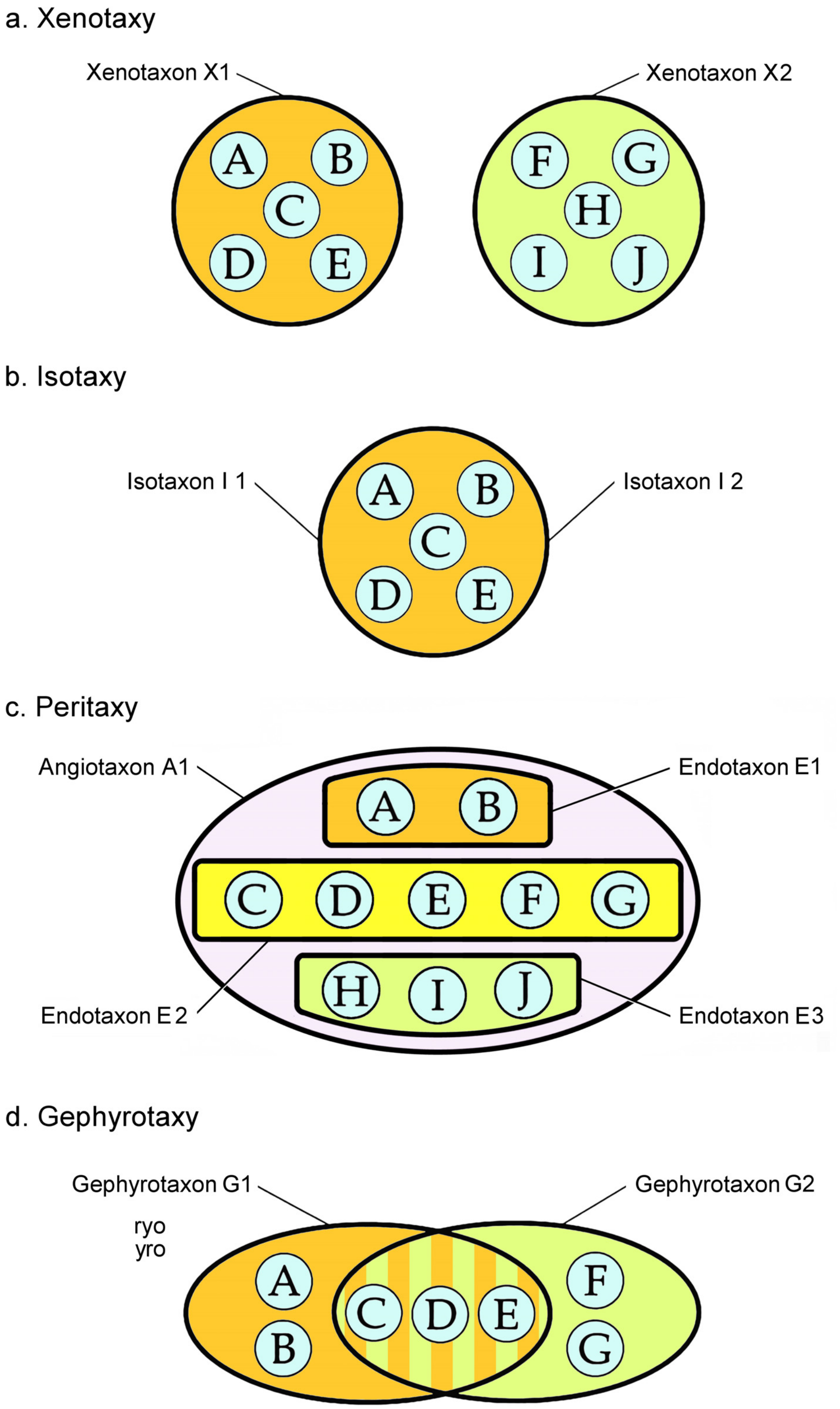

FIGURE 1. Diagrammatic illustration of the different patterns of relationships of topotaxy between taxa (A1, E1, E2, E3, G1, G2, I1, I2, X1, X2) concerning their contents or extensions. A to J are either individuals (specimens) in the case of species and subspecies, or taxa in the case of taxa of all other nominal-series. (Modified from Fig. 3 in Dubois 2005: 405).

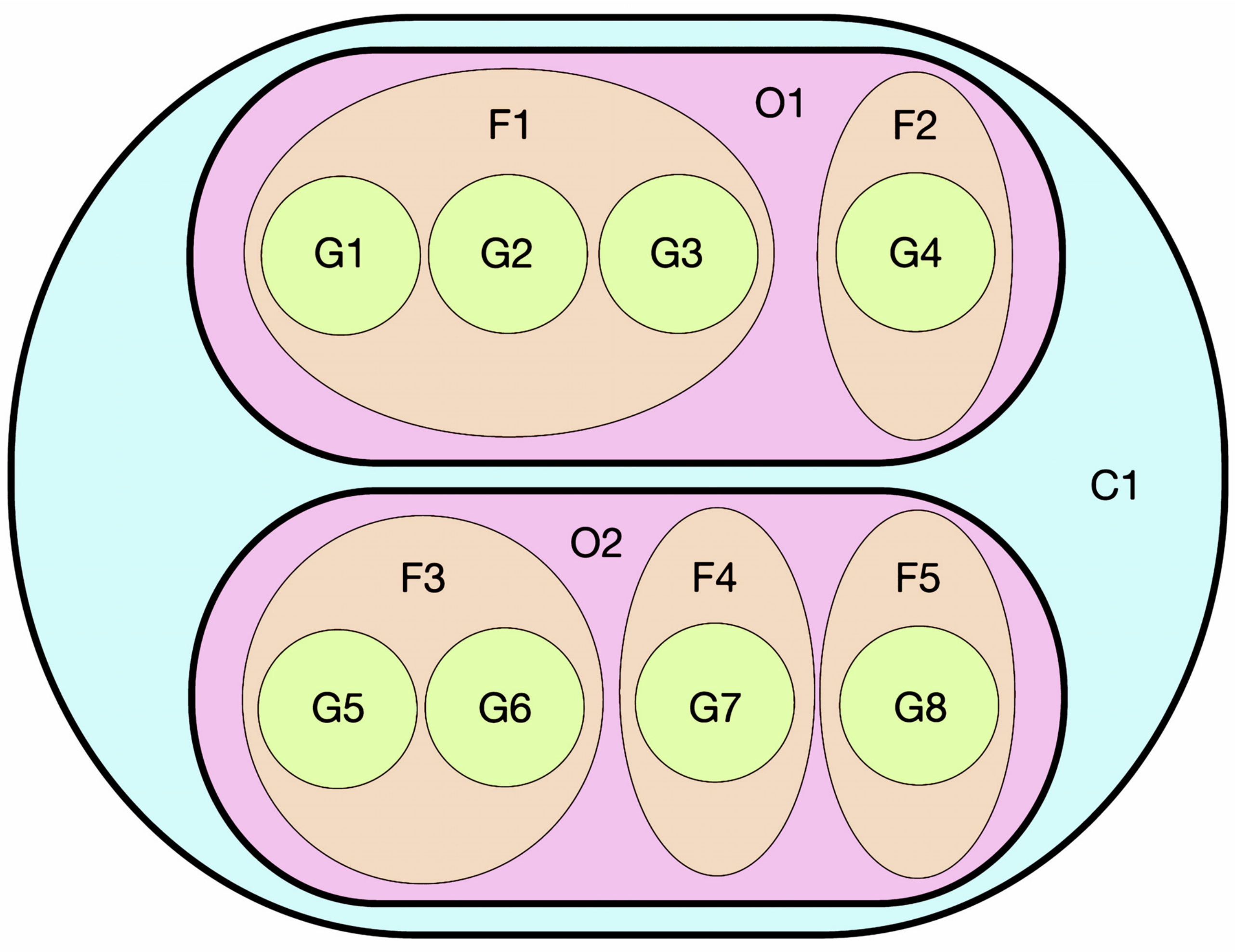

FIGURE 2. Diagrammatic illustration of the relationships of peritaxy in a taxonomic hierarchy: the class C1 is getangiotaxic to its orders O1 and O2, which are getangiotaxic to their families F1 to F5, which are getangiotaxic to their genera G1 to G8. C1 is telangiotaxic to F1 to F5 and G1 to G8, whereas O1 and O2 are telangiotaxic respectively to G1 to G4 and to G5 to G8. (Modified from Fig. 3 in Dubois 2006*a*: 201).

Acknowledgements

We thank Roger Bour (Paris) for the preparation of the figures, and Ingo Brigandt (Edmonton) and two anonymous referees for comments on a first version of this paper.

References

Anonymous [Oracle®Database SQL Reference] (2005) *9 SQL queries and subqueries. Hierarchical queries*. <http://docs.oracle.com/cd/B19306_01/server.102/b14200/queries003.htm> [Last accessed: 24 February 2012].

Celko, J. (2004) *Joe Celko's trees and hierarchies SQL for smarties*. San Francisco (Morgan Kaufmann): i–xii + 1–225.

Dubois, A. (2005) Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature*. 1. Some general questions, concepts and terms of biological nomenclature. *Zoosystema*, **27** (2): 365–426.

Dubois, A. (2006*a*) Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature*. 2. The proposed Rules and their rationale. *Zoosystema*, **28** (1): 165–258.

Dubois, A. (2006*b*) New proposals for naming lower-ranked taxa within the frame of the *International Code of Zoological Nomenclature*. *Comptes rendus Biologies*, **329** (10): 823–840.

Dubois, A. (2007) Phylogeny, taxonomy and nomenclature: the problem of taxonomic categories and of nomenclatural ranks. *Zootaxa*, **1519**: 27–68.

Dubois, A. (2008) Phylogenetic hypotheses, taxa and nomina in zoology. *In*: A. Minelli, L. Bonato & G. Fusco (ed.), *Updating the Linnaean heritage: names as tools for thinking about animals and plants*, *Zootaxa*, **1950**: 51–86.

Dubois, A. (2011) The *International Code of Zoological Nomenclature* must be drastically improved before it is too late. *Bionomina*, **2**: 1–104. <http://dx.doi.org/10.11646/bionomina.2.1.1>.

Garcia-Molina, H., Ullman, J. D. & Widom, J. (2008) *Database systems: the complete book.* Second edition. Upper Saddle River, New Jersey (Prentice Hall): i–xxvii + 1–1119.

Groff, J. R. & Weinberg, P. N. (1999) *SQL: the complete reference*. Berkeley (McGraw-Hill/Osborne): i–xxiv + 1–864.

Kamfonas, M. (1992) Recursive hierarchies: the relational taboo! *The relational Journal*, October-November 1992. <www.kamfonas.com/id3.html>.

Shipunov, A. (2011) The problem of hemihomonyms and the on-line hemihomonyms database (*HHDB*). *Bionomina*, **4**: 65–72.

Singh, P. K. (2009) *Database management system concepts*. New Delhi (VK Publication): [i–iv] + 1–188.

Su, W., Wang, J. & Lochovsky, F. (2006) Automatic hierarchical classification of structured deep web databases. *Lecture notes in computer science*, 4255: 210–221. <http://dx.doi.org/10.1007/11912873_23>.

Wittlinger, G. & Wittlinger, H. (2004) *Textbook of Dr. Vodder's manual lymph drainage*. Volume 1. *Basic course*. 7th edition. Stuttgart (Georg Thieme Verlag): 1–137.

Submitted: 16 May 2012. Accepted: 18 May 2013. Published: 28 June 2013.
Corresponding Editor: Ingo Brigandt.