C. R. Palevol 11 (2012) 563–566

Contents lists available at SciVerse ScienceDirect

Comptes Rendus Palevol

www.sciencedirect.com

General palaeontology, systematics and evolution (General phylogenetics and systematic theory)

LisBeth: New cladistics for phylogenetics and biogeography

LisBeth : un logiciel de cladistique pour la phylogénétique et la biogéographie

René Zaragüeta Bagils[a,b], Visotheary Ung[a,b], Anaïs Grand[a,*,b], Régine Vignes-Lebbe[a,b], Nathanaël Cao[a,c], Jacques Ducasse[a,b]

[a] *UMR 7207 CNRS MNHN UPMC, Centre de Recherche sur la Paléobiodiversité et les Paléoenvironnements (CR2P), 57, rue Cuvier, CP48, 75005 Paris, France*
[b] *Laboratoire Informatique et Systématique (LIS), UPMC Université Paris06, 4, place Jussieu, 75005 Paris, France*
[c] *Laboratoire Paléobotanique et Paléoécologie, UPMC Université Paris06, 4, place Jussieu, 75005 Paris, France*

ARTICLE INFO

Article history:
Received 21 May 2012
Accepted after revision 28 July 2012
Available online 1st September 2012

Presented by Philippe Taquet

Keywords:
Cladistics
Evolution
Phylogenetics
Historical biogeography

ABSTRACT

Within phylogenetics, two methods are known to implement cladistics: parsimony or maximum parsimony (MP) and three-item analysis (3ia). Despite the lack of suitable software, 3ia is occasionally used in systematic, and more regularly, in historical biogeography. Here, we present LisBeth, the first and only phylogenetic/biogeographic program freely available that uses the 3ia approach and offer some insights into its theoretical propositions. LisBeth does not rely on the conventional taxon/character matrix. Instead, characters are represented as rooted trees. LisBeth performs 3ia analyses based on maximum congruence of three-item statements and calculates the intersection tree (which differs from usual consensus). In biogeography, it applies the transparent method to handle widespread taxa and implements paralogy-free subtree analysis to remove redundant distributions. For the sake of interoperability, LisBeth may import/export characters from/to matrix in NEXUS format, allowing comparison with other cladistic programs. LisBeth also imports phylogenetic characters from Xper[2] knowledge bases.

Mots clés :
Cladistique
Évolution
Phylogénétique
Biogéographie historique

RÉSUMÉ

En cladistique, ne bénéficiant d'aucun logiciel adéquat, l'analyse à trois éléments (3ia) est moins répandue que la parcimonie (MP), bien que fréquemment utilisée en biogéographie historique. LisBeth est le premier et unique logiciel de 3ia pour la phylogénie et la biogéographie, disponible gratuitement, et développé à partir de propositions théoriques déjà publiées et originales. LisBeth ne repose pas sur la représentation matricielle conventionnelle, les caractères étant des arbres racinés dont sont issues les hypothèses à trois éléments, parmi lesquelles LisBeth recherche la congruence maximale. LisBeth reconstruit un arbre qui diffère des consensus habituels : l'arbre d'intersection. En biogéographie, la « méthode transparente » et « l'analyse des sous-arbres libres de paralogie » sont implémentées pour gérer les taxons à distributions larges et/ou redondantes. Dans un souci d'interopérabilité, LisBeth peut importer/exporter des matrices de caractères au format NEXUS, rendant ainsi comparables les différentes approches cladistiques. Un import existe aussi à partir des bases de connaissances Xper[2].

* Corresponding author.
E-mail address: grand@mnhn.fr (A. Grand).

1631-0683/$ – see front matter
doi:10.1016/j.crpv.2012.07.002

1. Introduction

Phylogenetic inference is currently performed using two main theoretical approaches, probabilistic and cladistic. Probabilistic methods search for trees that maximize the probability of attribute distribution according to a particular model of evolution, which is relevant to aligned DNA sequences analysis. Cladistic methods infer optimal trees that maximize hypotheses of homology and may be applied to both DNA sequences and morphological traits.

Within the cladistic approach, two methods exist: parsimony (MP) and three-item analysis (3ia). MP interprets homology as a series of transformations (including the transformation of absence into presence of any trait) and is the most widely used approach. 3ia was first suggested as an improvement to parsimony analysis for phylogenetic inference (Nelson and Platnick, 1991). However, in 3ia, homology is hierarchically interpreted as differentiation of particular traits from more general ones (Cao et al., 2007). Hence, both approaches deeply differ in theoretical assumptions. 3ia approach has also been shown relevant in historical biogeography (Nelson and Ladiges, 1991).

So far, no appropriate empirical comparison between MP and 3ia exists. The main reason was the lack of a 3ia-oriented software. Here, we present LisBeth (Laboratoire d'Informatique et Systématique program for Biogeography and Evolution using THree-item analysis; previous test versions were distributed as Nelson05), the first program designed to implement 3ia. LisBeth may be used for both phylogenetic and historical biogeographic analyses.

2. Methods

2.1. Character representation

3ia is the only theory of phylogenetics that does not rely on a "matrix" (i.e. table) representation of characters/taxa (taxa/areas in biogeography), and does not need outgroup taxa (or areas). LisBeth implements for the first time representation of characters as hierarchies or rooted trees of states (Cao et al., 2007) and assigns character states as synapomorphies of taxa.

Characters, i.e. rooted character state trees entered by users, are reduced to sets of three-item statements (3is hereafter). Each 3is represents the minimal hierarchical statement, i.e. given three entities (e.g. terminal taxa or areas of endemism), two of them are more closely related to each other than any is to the third. Thus, all 3is have the same minimal hierarchical structure (A (B C)). Fractional weighting (FW) is applied to 3is as described by Nelson and Ladiges (1992).

2.2. Algorithmics

Hennig (1966), the German entomologist who founded cladistic analysis, defined congruence as the criterion of optimality in cladistic analysis. To discover congruence one might use either parsimony (e.g. Farris, 1983) or compatibility analyses (e.g. Estabrook et al., 1976). It has been shown (Wilkinson, 1994a) that both criteria are computationally identical in 3ia. However, compatibility allows further theoretical implementations; hence, LisBeth uses compatibility. Relationships among taxa or biogeographic areas are found by combining the 3is deduced from all characters. So far, 3ia only benefits from exact algorithms, i.e. exhaustive and branch and bound (Hendy and Penny, 1982) searches, which are applied on 3is, associated with their FW. Thus, LisBeth first measures the sum of FW of compatible 3is. Optimal trees have the maximum amount of compatible 3is, i.e. are built from the max-clique of mutually compatible 3is that has the biggest sum of fractional weights.

3. Results and discussion

3.1. Implementation

We wanted the software to be modular; hence, LisBeth is currently a package of several programs, mainly written in C language for improved performance, including parallel computation when needed. There are command-line utilities, which can be combined to produce the desired sequence of computation. For instance, scripts can be written quickly with shell interface, while sophisticated computation can be run by confirmed programmers using modules of the popular Python language (www.python.org). Thus, LisBeth is aimed to become an open platform for computing in 3ia. LisBeth also includes a Graphical User Interface (GUI) (Fig. 1).

3.2. Character edition

Characters are edited via the GUI as rooted trees (Fig. 1A) or Venn diagrams (Fig. 1B). The classes (or nodes) of a character are first created and then filled with taxa (Fig. 1C) according to the state they present. Names of character states are shown in Fig. 1D. The list of characters appears in Fig. 1E.

3.3. Tree edition

Displayed trees may be edited manually (with parenthetical notations or Venn diagrams via the GUI), i.e. relative positions of terminals may be modified, and/or saved in eps format for printing. LisBeth automatically arranges trees in four different ways and terminal taxa may be automatically aligned.

3.4. Algorithmics

LisBeth implements exact algorithms (exhaustive and branch and bound searches, Fig. 1F).

The procedure may be separated in two steps: 1) rearrangement of characters in 3is associated with their FW; 2) search for the optimal trees. The list of optimal trees is displayed in Fig. 1G.

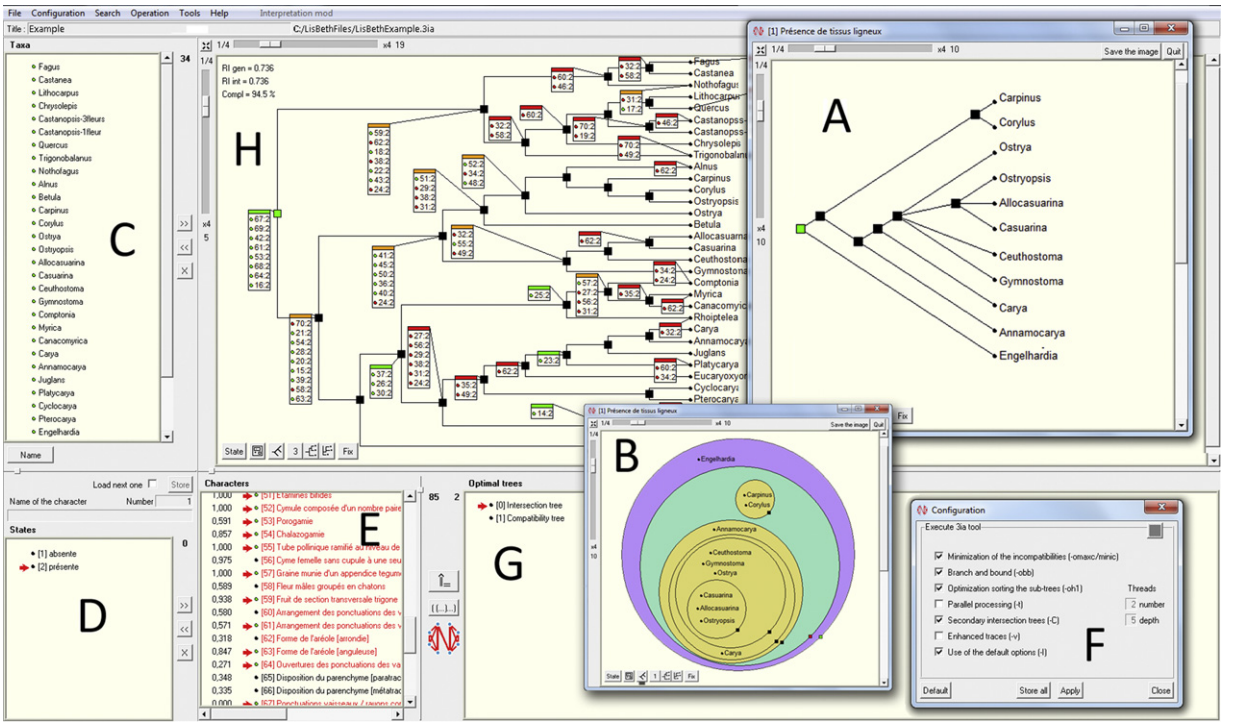

Fig. 1. LisBeth GUI. A. Character displayed as a rooted tree. B. Character displayed as a Venn diagram. C. List of taxa (or endemic areas). D. Description of characters and character-states. E. List of characters. F. Configuration window. G. List of the intersection tree and optimal trees. H. Display of characters or cladograms, retention index value, completeness index value and character interpretation.
Fig. 1. Interface de LisBeth. A. Affichage des caractères représentés sous forme d'arbres racinés. B. Affichage des caractères représentés sous forme de diagrammes de Venn. C. Liste des taxons (ou aires d'endémisme). D. Description des caractères et des états de caractères. E. Liste des caractères. F. Fenêtre de configuration. G. Liste de l'arbre d'intersection et des arbres optimaux. H. Affichage des caractères ou cladogrammes, de l'indice de rétention, de l'indice de complétude et de l'interprétation des caractères.

3.5. Intersection tree

LisBeth calculates and displays the results (i.e. most compatible trees) in a list of numbered trees from "[1]" to "[n]", and recombines the intersection tree, labelled "[0]" (Figure 1G). The intersection tree is a summary, more than a consensus, a tree built from, and only from, the 3is that are common to all the optimal trees and characters (Cao et al., 2009). It summarizes all the information that is common to the optimal trees, and only this information. The intersection tree has the desirable characteristics of consensus trees (Wilkinson, 1994b). Furthermore, it takes into account the information present in the characters that served to build the optimal trees, whereas this information is lost in consensus trees.

3.6. Character interpretation

LisBeth implements the 3ia method for character interpretation, based on the tree topology and character state distributions. For each node, synapomorphies are shown in green and homoplasies in red (Fig. 1H). Orange colour means that both synapomorphies and homoplasies are found for a given node. The character interpretation can be displayed for each character independently as well as for a set of selected characters (here, characters indicated by the red colour in Fig. 1E) or for a given node. Relative position of the lists of character states (displayed in boxes with a colourful banner) can be modified since each box is linked with the corresponding node.

3.7. Support

In 3ia the retention index (Archie, 1989; Farris, 1989) measures the proportion of 3is deduced from the primary hypotheses of homology that fit the optimal tree (Kitching et al., 1998). The retention index value of optimal trees and of the intersection tree are automatically generated and displayed, as well as the per-character values (Fig. 1H). We define a new completeness index (Compl), representing the proportion of 3is deduced from the intersection tree that are also present in the characters, and it can be displayed as well (Fig. 1H). Let A be the number of 3is common to all optimal trees found (i.e. used to rebuild the intersection tree); let B be the number of 3is deduced from the rebuilt intersection tree.

$$\text{Compl} = \frac{A}{B}$$

Compl can be seen as a measurement of the degree of support provided by primary hypotheses of homology. A low value implies that a significant proportion of 3is deduced from the intersection tree are not present in the

character set likely because few characters support the cladogram.

3.8. Biogeography

LisBeth is also an appropriate tool for cladistic biogeography (Nelson and Ladiges, 1991). Users supply a list of taxa and areas, the area(s) of endemism where each taxon lives and the cladograms of taxa that provide the evidence of area relationships. LisBeth translates cladograms of taxa to taxon-area cladograms (TACs) by replacing terminal taxa with the areas of endemism where representatives are found. This may lead to two well-known problems: redundant areas (due to taxic paralogy, i.e. taxon diversification prior to area duplication) and widespread taxa (or Multiple Areas in Single Terminals [MASTs] according to Ebach et al., 2005). LisBeth applies the Paralogy-free Subtree analysis (Nelson and Ladiges, 1996) and the Transparent method (Ebach et al., 2005), respectively, to resolve them. Redundancy of areas implies conflicting relationships; the Paralogy-free Subtree analysis combines subsets of non-conflicting 3is within TACs into subtrees. The Transparent method is the first method that supplies a separate procedure dealing with MASTs, which uses the Paralogy-free Subtree analysis rather than solving paralogy. It provides all biogeographic 3is implied by widespread taxa of the TACs.

3.9. Export/import

LisBeth offers the possibility to extend each taxonomist's daily work. Whenever a knowledge base is created, i.e. a set of structured descriptive data, it is possible to define a set of characters from an Xper^2 (Ung et al., 2010) knowledge base. For the sake of interoperability, a particular, very simple, scripting format allows choosing among descriptors, specifying hypotheses of homology to any set of descriptor-states, and accepting, discarding, or creating new dependencies among characters.

LisBeth may also export 3is matrices in NEXUS format for MP searches. It is capable of importing the results for empirical comparisons under different cladistic interpretations.

4. Conclusion

LisBeth is continually improved with new characteristics (like other platform support such as Mac and a Spanish version), implementation of a new solution for MASTs (a new method using A2 (Nelson and Platnick, 1981) is currently being implemented), and ergonomic enhancements. A new GUI intended for biogeography will be available shortly. With its accurate knowledge representation, i.e. characters as hierarchies of states or rooted dichotomous trees, LisBeth provides a powerful tool to perform efficient phylogenetic and biogeographic studies. Hence, LisBeth opens new possibilities towards empirical studies (e.g. comparative biogeography of the Pacific – Parenti and Ebach, 2009, Caribbean biogeography – Escalante et al., 2007), cladistic method comparisons and simulation-based analyses (Huelsenbeck, 1995; Laurin and Germain, 2011)

Acknowledgements

We would like to thank D.M. Williams and M.C. Ebach for fruitful discussions and tests on previous version of the software and on theoretical developments we implement now in LisBeth.

References

Archie, J.W., 1989. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. Syst. Zool. 38, 253–269.
Cao, N., Zaragüeta Bagils, R., Vignes-Lebbe, R., 2007. A hierarchical representation of the hypothesis of homology. Geodiversitas 29 (1), 5–15.
Cao, N., El Azawi, M., Zaragüeta Bagils, R., 2009. Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example. Bull. Soc. geol. France 180 (1), 13–15.
Ebach, M.C., Humphries, C.J., Newman, R.A., Williams, D.W., Walsh, S.A., 2005. Assumption 2: opaque to intuition? J. Biogeogr. 32, 781–787.
Escalante, T., Rodríguez, G., Cao, N., Ebach, M.C., Morrone, J.J., 2007. Cladistic biogeographic analysis suggests an Early Caribbean diversification in Mexico. Naturwissenschaften 94, 561–565.
Estabrook, G.F., Johnson, C.S., McMorris, F.R., 1976. A mathematical foundation for the analysis of cladistic character compatibility. Math. Biosci. 29 (1–2), 181–187.
Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in cladistics II. Columbia University Press, New York, pp. 7–36.
Farris, J.S., 1989. The retention index and the rescaled consistency index. Cladistics 5, 417–419.
Hendy, M.D., Penny, D., 1982. Branch and bound algorithms to determine minimal evolutionary trees. Math. Biosci. 59 (2), 277–290.
Hennig, W., 1966. Phylogenetic systematics. University of Illinois Press, Chicago, Illinois, 264 p.
Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44, 17–48.
Kitching, I.J., Forey, P.L., Humphries, C.J., Williams, D.W., 1998. Cladistics: the theory and practice of parsimony analysis, 11., 2nd ed. The Oxford University Press, Oxford, 228 p.
Laurin, M., Germain, D., 2011. Developmental characters in phylogenetic inference and their absolute timing information. Syst. Biol. 60, 630–644.
Nelson, G., Ladiges, P.Y., 1991. Three-area statements: standard assumptions for biogeographic analysis. Syst. Zool. 40 (4), 470–485.
Nelson, G., Ladiges, P.Y., 1992. Information content and fractional weight of three-item statements. Syst. Biol. 41 (4), 490–494.
Nelson, G., Ladiges, P.Y., 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. American Museum Novitates 3167, 1–58.
Nelson, G., Platnick, N.I., 1981. Systematics and Biogeography. In: Cladistics and Vicariance. Columbia University Press, New York, 567 p.
Nelson, G., Platnick, N.I., 1991. Three-taxon statements: amore precise use of parsimony? Cladistics 7, 351–366.
Parenti, L.R., Ebach, M.C., 2009. Comparative biogeography: discovering and classifying biogeographical patterns of a dynamic Earth. University of California Press, Berkeley, 295 p.
Ung, V., Dubus, G., Zaragüeta-Bagils, R., Vignes-Lebbe, R., 2010. Xper^2: introducing e-Taxonomy. Bioinformatics 26 (5), 703–704.
Wilkinson, M., 1994a. Three-taxon statements: when is a parsimony analysis also a clique analysis? Cladistics 10, 221–223.
Wilkinson, M., 1994b. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. Syst. Biol. 43 (3), 343–368.