



HAL
open science

Nouvelles inférences cladistiques sur la biogéographie historique du Sud-Est asiatique et de la région de l'Ouest du Pacifique. Réflexion sur une approche intégrative de la dimension temporelle en biogéographie historique

Visotheary Ung

► To cite this version:

Visotheary Ung. Nouvelles inférences cladistiques sur la biogéographie historique du Sud-Est asiatique et de la région de l'Ouest du Pacifique. Réflexion sur une approche intégrative de la dimension temporelle en biogéographie historique. Evolution [q-bio.PE]. MNHN Paris, 2013. Français. NNT : 2013MNHN0010 . tel-03936257

HAL Id: tel-03936257

<https://hal.science/tel-03936257>

Submitted on 29 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

MUSEUM NATIONAL



D'HISTOIRE NATURELLE

Ecole Doctorale Sciences de la Nature et de l'Homme – ED 227

Année 2013

N°attribué par la bibliothèque

|||||

THESE

Pour obtenir le grade de

DOCTEUR DU MUSEUM NATIONAL D'HISTOIRE NATURELLE

Spécialités : Biogéographie, Systématique, Informatique pour la Biodiversité

Présentée et soutenue publiquement par :

Visotheary UNG

Le 18 Octobre 2013

**Nouvelles inférences cladistiques sur la Biogéographie historique du Sud-Est asiatique
et de la région de l'Ouest du Pacifique.**

**Réflexion sur une approche intégrative de la dimension temporelle en
Biogéographie historique.**

Sous la direction de : **Madame VIGNES-LEBBE Régine (Pr)**, **Messieurs ZARAGÜETA i BAGILS René (Mcf)** et **WILLIAMS David (Pr)**

JURY :

Mr Philippe Grandcolas	Directeur de recherche, CNRS, Paris	Président
Mme Régine Vignes-Lebbe	Professeur, Université Pierre et Marie Curie, Paris	Directrice de Thèse
Mr René Zaragüeta i Bagils	Maître de conférences, Université Pierre et Marie Curie, Paris	Directeur de Thèse
Mr David M. Williams	Professor, Natural History Museum, London	Directeur de Thèse
Mme Lynne R. Parenti	Professor, Smithsonian National Museum of Natural History, Washington	Rapportrice
Mr David Gower	Researcher, Natural History Museum, London	Rapporteur
Mme Sandrine Meylan	Professeur, IUFM de Paris-Université la Sorbonne, Paris	Examinatrice
Mr Jorge Cubo	Professeur, Université Pierre et Marie Curie, Paris	Examineur

*« Every species has come to existence coincident both in space and time
with a pre-existing closely allied species »*

connue sous le nom de « Sarawak law », Alfred Russel Wallace, 1960

On the law which has regulated the introduction of new species



Photographie d'Alfred Russel Wallace, prise en 1862 à Singapour (*Alfred Russel Wallace, Letters and Reminiscences*, James Marchant 1916)

Remerciements

Toute thèse se doit de commencer par des remerciements, je pense même que ce doit être la partie la plus lue du document. Je ne pourrais donc me soustraire à ce difficile exercice, cependant, plutôt que de prendre le risque d'oublier quelqu'un, je vais jouer la carte de la facilité et ne mentionnerai personne en particulier, que chacun ayant partagé un tant soit peu tout ou partie de ces 4 dernières années se retrouvent dans l'expression de ma reconnaissance et de mon amitié des plus sincères.

En revanche, je me dois d'exprimer ici, mes remerciements les plus vifs aux membres de mon jury qui ont bien voulu examiner mon travail de thèse. Que Lynne Parenti, David Gower, Sandrine Meylan, Jorge Cubo et Philippe Grandcolas soient chaleureusement remerciés pour le temps qu'ils auront bien voulu me consacrer et pour leurs conseils.

Je tiens à remercier très sincèrement David Williams pour m'avoir orientée vers l'étude du Sud-Est asiatique. Lorsque nous avons discuté il y a 4 ans d'un possible sujet de thèse, il m'a proposé de m'attaquer soit à la région de l'Amérique centrale soit au Sud-Est asiatique, bien évidemment, mon appétence naturelle pour tout ce qui touche à cette région dont je suis originaire a largement orienté mon choix. Je le remercie pour sa disponibilité, sa gentillesse sans égale et son humour « so british ».

Je voudrais remercier très amicalement René Zaragüeta-Bagils d'avoir bien voulu me guider sur le long chemin sinueux de la « méthodologie », nos routes se sont croisées pour la première il y a bien longtemps et déjà, nous avons discuté de « méthodes ». Je lui suis reconnaissante de ne jamais succomber à la facilité de suivre des thématiques plus « dans le vent » et de toujours garder une intégrité scientifique sans faille. Que sa patience pour transmettre son savoir soit grandement plébiscitée !

Je remercie très chaleureusement Régine Vignes-Lebbe pour sa disponibilité et l'écoute dont elle a fait preuve tout au long de ces années et je ne parle pas que de la thèse. Je la remercie de m'avoir accueillie à bras ouverts il y a 7 ans dans son équipe. Je ne saurai oublier non plus avec quelle pédagogie elle a su m'enseigner les bases de la Systématique et l'importance de garder un esprit critique dans toute entreprise scientifique. Je voudrais aussi remercier par la même occasion Jacques Lebbe qui a eu tellement d'influence sur l'étudiante que j'étais alors et qui a su insuffler une telle énergie dans ses enseignements et fait germer la « graine de la réflexion » chez chacun de ses étudiants. « Pas de boîte noire », je l'ai bien compris, merci Jacques !

J'ai une pensée toute particulière pour Jacques Ducasse, le développeur de LisBeth sans qui tout ce travail n'aurait été possible. Je voudrais préciser ici que Jacques est chercheur bénévole affilié à notre UMR et qu'il

travaille avec nous sur son temps libre. Je le remercie très sincèrement pour sa disponibilité et son aide précieuse avec LisBeth et surtout pour la réactivité à laquelle il m'a fourni tous les petits scripts informatiques qui m'ont tant facilité la tâche lors de mes analyses.

Je remercie tous les membres de l'UMR 7207, particulièrement ceux du bâtiment de Géologie qui supportent tous les jours nos rires et éclats de voix.

J'ai aussi une pensée spéciale pour Angelina, Elyane et Suzy pour leur disponibilité, leur indispensable aide au quotidien, les organisations et suivis de missions inégalés jusqu'à présent et surtout pour leur incroyable efficacité doublée d'une si grande gentillesse.

Finalement, je ne peux résister à l'envie de citer quelques personnes qui ont eu une importance toute particulière pour moi. Je pense à certains membres de mon ancien laboratoire (UMR 7625) qui m'ont toujours soutenu et qui méritent d'avoir un petit message ici. Sandrine, Marie-Bernadette, Paola, Christelle, Elizabeth, Clotilde, Thibaud, Mathieu, Sébastien, Hélène, Claire, soyez tous remerciés pour tout ce que vous m'avez apporté !

Aussi, comment pourrais-je omettre de parler de mes acolytes de bureaux qui partagent depuis quelques années tant de choses ? Florian, Frédéric, Thomas, Nils, Antoine, Hélène, mon travail n'aurait jamais été si agréable sans votre dynamique présence (et je pèse mes mots !). Que Anaïs, Adeline et Ninon soient aussi grandement remerciées d'avoir bien voulu m'accueillir lorsque j'ai voulu momentanément changer de bureau. Passer d'un bureau de garçons à un bureau de filles présente quelques avantages certains. Merci à Brigitte pour sa bonne humeur et sa présence réconfortante !

A Peggy, que pourrais-je te dire que tu ne sais déjà ? Merci d'être là !

A mes parents et frères et sœurs, merci de si bien prendre le relais lors de mes absences, je ne vous remercierais jamais assez de votre présence.

Enfin, je ne pourrai finir sans remercier ma petite famille, mes filles, Vaddhany et Chan-Devy qui supportent avec tant de compréhension mes absences ; Pascal pour si bien « veiller au grain » lorsque, justement, je suis en déplacement et son indéfectible soutien dans tout ce que j'entreprends. Je ne pourrais jamais assez te remercier pour tout ce que tu fais pour moi !

Table des matières

Table des matières

REMERCIEMENTS	7
TABLE DES MATIERES	11
TABLE DES FIGURES	15
TABLE DES ENCADRES	17
PROLOGUE	19
INTRODUCTION	23
CHAPITRE I : CONTEXTE METHODOLOGIQUE ET IMPLEMENTATION LOGICIELLE	27
INTRODUCTION	29
1. L'ANALYSE A TROIS ELEMENTS POUR LA BIOGEOGRAPHIE AVEC LISBETH	29
1.1 PRINCIPE DE LA METHODE	29
1.2 PARTICULARITES DE LA BIOGEOGRAPHIE	30
2. LISBETH	33
3. UTILISATION DE LISBETH POUR LA BIOGEOGRAPHIE	39
3.1 SAISIE DES CLADOGRAMMES DE TAXONS VIA L'INTERFACE GRAPHIQUE	39
3.2 SUBSTITUTION DES TAXONS TERMINAUX PAR LES AIRES D'ENDEMISME ET APPLICATION DE LA « PARALOGY-FREE SUBTREE ANALYSIS » ET DE LA METHODE TRANSPARENTE	39
4. AMELIORATIONS DU MODULE DE BIOGEOGRAPHIE	43
4.1 INTERFACE UTILISATEUR	43
4.1.1 Création du fichier .txt pour 3area (module réalisant la paralogy-free sub-tree analysis et la méthode transparente)	43
4.1.2 Saisie de la distribution des taxons	44
4.1.3 Lancement de 3area	44
4.1.4 Lancement de 3ia (sous-programme réalisant à proprement parlé l'analyse : décomposition en 3is, analyse de congruence et reconstruction de l'arbre d'intersection)	45
5. CREATION D'UNE « BOITE A OUTILS »	45
6. UNITES (OBJETS) D'ANALYSE EN BIOGEOGRAPHIE : VERS UNE REPRESENTATION SEMANTIQUE DES CARACTERES BIOGEOGRAPHIQUES	47
6.1 « ON THE ONTOLOGY OF AREAS OF ENDEMISM »	49
6.1.1 Contexte	49

6.1.2	Résumé	49
6.1.3	Conclusions	49
6.2	« HENNIG'S REVOLUTION HAS ARRIVED »	56
6.2.1	Contexte de l'article	56
6.2.2	Conclusions	56
6.3	« TO ADD A LITTLE SEMANTICS CAN HELP FOR CONSERVATION POLICIES »	58
6.3.1	Contexte de l'article	58
6.3.2	Résumé	58
6.3.3	Conclusions	58
	CONCLUSIONS ET PERSPECTIVES	71
1	CONCLUSIONS	71
2	PERSPECTIVES	71
 <u>CHAPITRE II: NOUVELLES INFERENCE CLADISTIQUES SUR LA BIOGEOGRAPHIE HISTORIQUE DU SUD-EST ASIATIQUE ET DE LA REGION DE L'OUEST DU PACIFIQUE</u>		73
 <u>INTRODUCTION</u>		75
 1. HISTOIRE GEOLOGIQUE DE L'ASIE DU SUD-EST OU COMMENT REASSEMBLER UN « PUZZLE » GEANT ?		78
1.1.	SEPARATION DU GONDWANA (-160 MA)	78
1.2	OUVERTURE DE LA MER DE TASMANIE (-95 MA)	82
1.3	COLLISION ET SUBDUCTION (-50 MA)	82
1.4	FRACTURATION DE L'ARC DU PACIFIQUE OUEST OU ARC MELANESIEN (-40 MA)	82
1.5	REORGANISATION DES LIMITES DES PLAQUES AU CENOZOÏQUE (VERS -25 MA)	83
1.6	MODIFICATIONS DES MOUVEMENTS ET DES LIMITES DES PLAQUES AU NEOGENE (VERS -5 MA)	85
2. DISTRIBUTION DES TAXONS DANS LA REGION : LIGNES TRACEES		86
3. CLIMAT ET CHANGEMENTS DU NIVEAU DE LA MER AU QUATERNAIRE		88
4. « COMPARATIVE BIOGEOGRAPHY OF SOUTHEAST ASIA AND THE WEST PACIFIC REGION »		90
4.1	CONTEXTE DE L'ARTICLE	90
4.2	RESUME	91
4.3	CONCLUSIONS	91
CONCLUSION ET PERSPECTIVES		125
1.	CONCLUSIONS	125
2.	PERSPECTIVES	125
2.1	Sélection de taxons selon leur âge	125
2.2.	Wallacea : une zone tampon majeure	126
2.3.	Sulawesi : une île modèle pour tester la vicariance	127
2.4.	Tester l'existence de « Pandora »	129

2.5. Tester la classification de Silder et Silder 1938-1939	129
2.6 Nouvelles pistes pour établir des règles de délimitations d'aires de conservation	131

CHAPITRE III : REFLEXION SUR UNE APPROCHE INTEGRATIVE DE LA DIMENSION TEMPORELLE EN BIOGEOGRAPHIE HISTORIQUE **135**

INTRODUCTION **137**

1. ETAT DE L'ART **137**

1.1 « THE TIME CONTROL REFINEMENT » 137

1.2 « TEMPORAL CONGRUENCE AND CLADISTIC ANALYSIS BIOGEOGRAPHY AND CO-SPECIATION » 139

1.3 LE DECOUPAGE EN « TRANCHES TEMPORELLES » OU « TIME-SLICING » 141

1.3.1 « The Chronobiogeographic paradigm » 141

1.3.2 « The Temporally Partitioned Component Analysis » 144

1.3.4 « The Test of Time » 145

1.3.5 « The Cladistic Biochronologic Analysis » 148

2. SYNCHRONIE, DIACHRONIE : LE TEMPS, UNE DIMENSION OUBLIEE ? **152**

2.1. POURQUOI LA BIOGEOGRAPHIE HISTORIQUE EST-ELLE UN PROBLEME CLADISTIQUE ? 152

2.2. SYNCHRONIE : ANALYSE DANS UNE MEME TRANCHE TEMPORELLE 153

2.3 DIACHRONIE : ANALYSE D'UNE OU PLUSIEURS PERIODES TEMPORELLES SIMULTANEMENT 154

3. VERS UNE BIOGEOGRAPHIE HISTORIQUE INTEGRATIVE **154**

4. « BRIDGING THE TEMPORAL GAP IN COMPARATIVE BIOGEOGRAPHY » **156**

4.1 CONTEXTE DE L'ARTICLE 156

4.2 RESUME 156

4.3 CONCLUSIONS 157

CONCLUSIONS ET PERSPECTIVES **163**

1. CONCLUSIONS 163

2. PERSPECTIVES 164

CONCLUSION GENERALE **165**

CONCLUSION GENERALE **167**

1. LES APPORTS DE LA 3IA A LA BIOGEOGRAPHIE **168**

2. PROFESSION BIOGEOGRAPHE ? **169**

3. VERS UNE BIOGEOGRAPHIE INTEGRATIVE **170**

EPILOGUE **173**

REFERENCES BIBLIOGRAPHIQUES **179**

ANNEXES	191
ANNEXE I : TOOLBOX DE LISBETH	193
• Décomposition des caractères en 3is (« Character treatment »)	193
• Traçabilité des caractères (« Trace char history »)	193
• Facilités d'import/export	193
• Import d'arbres (« Tree import »)	193
• Import de matrices (« Character import »)	194
• Export de caractères (« Character export »)	194
• Extraction de différentes parties indépendantes contenues dans le fichier .3iz	194
• Arbre d'intersection (« Intersection tree »)	194
• Calcul des RI (« RI Computation »)	194
ANNEXE II : TEST DE LA CLASSIFICATION DE SHILDER ET SHILDER	195
ANNEXE III : BIOTIC EVOLUTION AND ENVIRONMENTAL CHANGE IN SOUTHEAST ASIA.	203
ANNEXE IV : THREE-ITEM ANALYSIS AND GADIFORMES (TELEOSTEI, PARACANTHOPTERYGII): HOW TO CONDUCT A CLADISTIC ANALYSIS USING LISBETH.	211
ANNEXE V : COMPARATIVE BIOGEOGRAPHY AND GEOGRAPHIC SPECIATION OF NORTH AMERICAN HIGHLAND FISHES FROM THE MESA CENTRAL TO THE CANADIAN SHIELD	249
ANNEXE VI : COMMENTS ON THE INTERNATIONAL CODE OF AREA NOMENCLATURE (ICAN)	291
ANNEXE VII : XPER²: INTRODUCING E-TAXONOMY	295
ANNEXE VIII : IKEY+ : A NEW SINGLE-ACCESS KEY GENERATION WEB SERVICE.	299
ANNEXE IX : MKEY+: INTERACTIVE IDENTIFICATION WEBSERVICE	305
RÉSUMÉ	317
ABSTRACT	317

Table des figures

Figure 1 : Schéma présentant les étapes d'une analyse 3ia	30
Figure 2 : Cladogramme du taxon <i>Gallicolumba</i> utilisé dans l'analyse de Biogéographie comparative du Chapitre II (interface de LisBeth)	31
Figure 3 : Cladogramme du taxon <i>Gallicolumba</i> dans lequel les taxons terminaux ont été substitués par les aires d'endémisme où sont présents les taxons terminaux (interface de LisBeth)	32
Figure 4 : Cladogramme du taxon <i>Gallicolumba</i> via l'interface entière de LisBeth	32
Figure 5 : Cladogramme du taxon <i>Gallicolumba</i> sur lequel nous avons superposé la distribution des différentes espèces et indiqué un exemple de taxon à large répartition ainsi qu'une aire répétée	33
Figure 6 : Interface de saisie de LisBeth	39
Figure 7 : Lancement du sous-module 3area pour la Biogéographie	40
Figure 8 : Saisie des cladogrammes de taxons pour la création d'un fichier texte	43
Figure 9 : Proposition d'interface pour la saisie de la distribution des taxons	44
Figure 10 : Proposition d'ajout d'options	44
Figure 11 : Menu de lancement de l'analyse via l'interface de LisBeth	44
Figure 12 : Carte de la région d'étude (modifiée d'après OpenStreetMap)	77
Figure 13 : Reconstructions paléogéographiques à -120 Ma	79
Figure 14 : Carte de Bornéo (actuel)	81
Figure 15 : Reconstruction géographique	84
Figure 16 : Répartition des blocs continentaux, des fragments de terranes et des principaux points de suture de l'Asie du Sud-Est	85
Figure 17 : Lignes biogéographiques tracées dans la région (d'après van Oosterze 1997)	87
Figure 18 : Reconstruction de la région de Banda à 4 Ma	128
Figure 19 : Sous-régions biogéographiques Pandora et Indo-Malaysienne représentées sur une reconstruction globale du monde à -30 Ma	129
Figure 20 : Carte mondiale des régions présentant des « points-chauds » de biodiversité	132
Figure 21 : Modèle hypothétique présentant comment une histoire biogéographique complexe pour une aire test pourrait donner un pattern irrésolu	138
Figure 22 : « Réconciliation » entre des patterns incohérents d'hôtes et de parasites (A) pour les rendre congruents (B)	140
Figure 23 : Exemples de phylogénies et de leur adéquation temporelle	141
Figure 24 : Séquence de fragmentation hypothétique de l'aire XYZ (t1 et t2) suivie d'une dispersion (t3)	142
Figure 25 : Représentation d'une partition temporelle pour un cladogramme de taxons	145

Figure 26 : « The test of Time »	146
Figure 27 : Relation entre le temps et les nœuds d'un cladogramme enraciné	149
Figure 28 : Phylogénie calibrée d'un clade hypothétique à 6 taxons dont les distributions couvrent 3 paléobiotas	151
Figure 29 : Equivalence entre la Systématique et la Biogéographie historique (Zaragüeta Bagils, comm. pers.)	153
Figure 30 : Représentation hiérarchique des intervalles temporels (durées) (modifié d'après Zaragüeta Bagils et al. 2004)	155
Figure 31 : Deux exemples de cladogrammes non pectinés pour lesquels l'intégration d'un nouveau taxon X dans l'échantillon n'ajoute pas d'ambiguïté concernant le temps (modifié d'après Zaragüeta Bagils et al. 2004)	156

Table des encadrés

Encadré 1 : Biogéographie évolutive (d'après Parenti et Ebach 2009)	24
Encadré 2 : Biogéographie systématique (d'après Parenti et Ebach 2009)	24
Encadré 3 : Wallace et le Sud-Est asiatique	25
Encadré 4 : Squelette du format de fichier .txt pour la Biogéographie, exemple du taxon «Gallicolumba »	41
Encadré 5 : Exemple de fichier de sortie du sous-module 3area	42
Encadré 6 : Aires d'endémisme/ aires biogéographiques	48
Encadré 7 : « Hennig's revolution has arrived »	57
Encadré 8 : Carte du Gondwana au Crétacé	78
Encadré 9 : Carte du Sud-Est asiatique	80
Encadré 10 : La ligne de Wallace	86
Encadré 11 : Représentation simplifiée de la formation d'une chaîne volcanique divisant un biota au milieu de l'océan (modifié d'après Parenti et Ebach 2013)	90
Encadré 12 : Analyse préliminaire testant la classification de Silder et Silder	130
Encadré 13 : Algorithme du « paradigme chronobiogéographique »	144
Encadré 14 : « <i>Nothofagus</i> and Pacific biogeography » (Linder et Crisp 1995)	147
Encadré 15 : « Phylogenetic patterns in Northern hemisphere plant biogeography » (Donoghue et al. 2001)	148
Encadré 16 : « Brooks Parsimony Analysis » (Brooks 1981, Kluge 1988, Wiley 1988 a, b)	152
Encadré 17 : Procédure proposée pour une analyse biogéographique intégrative	163

Prologue

De manière non conventionnelle, je souhaite donner ici mon avis sur mes choix et orientations scientifiques, tant pour les expérimentations que pour l'écriture du présent manuscrit.

Il est normalement de bon ton que le thésard écrive à la première personne et fasse apparaître très clairement quels ont été ses apports scientifiques et ses perspectives. Cependant, par confort personnel, j'ai pris la décision d'écrire à la première personne du pluriel ou à la troisième personne. Je m'excuse par avance auprès de mes lecteurs si cette convenance personnelle trouble leur lecture que je souhaite pourtant agréable.

Néanmoins, je voudrais préciser maintenant que les contenus des chapitres I, II et III ont été réalisés à mon initiative propre suite aux premières discussions avec mes différents directeurs.

En vous souhaitant une bonne lecture !

Introduction

Introduction

La Biogéographie est de ces disciplines qui intriguent. Son étymologie même en fait une science pluridisciplinaire et qui se veut aujourd'hui intégrative. Cette branche de la géographie s'attache à étudier la composante géographique des taxons biologiques, leur répartition, leur distribution sur le globe terrestre.

Pourquoi les taxons sont-ils distribués de la sorte ? Pourquoi n'y-a-t-il pas de marsupiaux en Asie ni de tigres en Australie ? Pourquoi existe-t-il des taxons endémiques ?

On peut aussi se demander comment les taxons sont arrivés à une telle distribution. Par quels mécanismes ou processus ?

Enfin, la troisième question et non des moindres, concerne l'échelle temporelle, quand les événements se sont-ils produits ?

La dure tâche de la Biogéographie et donc du biogéographe qui la met en application est de répondre à ces questions qui semblent à première vue indépendantes : ***pourquoi, comment, quand ?***

Malgré tout, ces questions peuvent trouver des réponses faisant appel à des événements liés.

Cette thèse s'articule autour de deux questions, l'une empirique, l'autre fondamentale. La première s'intéresse à la Biogéographie comparative de l'Asie du Sud-Est et de la région à l'Ouest du Pacifique. La deuxième concerne l'intégration de la dimension temporelle en Biogéographie comparative. Nous allons successivement traiter ces deux points dans le présent document.

Avant d'entrer dans le vif du sujet qui nous préoccupe il nous faut faire, ici, un commentaire sur les différents courants qu'englobe la discipline.

Ebach et Parenti (2009), dans leur livre « Comparative biogeography » en donne une classification bien précise. Ils y distinguent la Biogéographie évolutive (Encadré 1) de la Biogéographie systématique (Encadré 2). La Biogéographie évolutive concerne plutôt des thématiques d'ordre écologique et va rendre compte des distributions taxiques sur la base de l'interprétation des processus à l'origine de ces distributions, elle comprend également toutes les méthodes basées sur des analyses matricielles; alors que la Biogéographie systématique comprend des méthodes de recherche de patterns généraux, entre autres et se divise en Biogéographie descriptive et Biogéographie cladistique.

En dépit de l'intérêt de chacune de ces approches, il nous est impossible de les détailler toutes dans le présent manuscrit et nous renvoyons donc le lecteur à chacune des références ci-après pour plus d'explications. Nous souhaitons juste préciser que notre discours se place clairement dans une approche de Biogéographie systématique cladistique (voir chapitre I).

Encadré 1 Biogéographie évolutive (d'après Parenti et Ebach 2009)

- Méthodes échographiques
 - Modélisation de niches écologiques (Peterson et al. 1999)
- Méthodes phylogénétiques
 - Biogéographie Phylogénétique (Hennig 1966, Brundin 1966)
 - Phylogéographie (sensu Avise 2000)
 - Analyses phylogénétiques pour comparer des arbres (Wojcicki and Brooks 2005)
- Méthodes basées sur des matrices
 - Parsimony analysis of endemism (PAE) (Rosen 1985, 1988)
 - Brooks parsimony analysis (BPA) (Wiley 1987)
 - Ancestral area analysis (AAA) (Bremer 1992, 1995)
 - Dispersal and Vicariance Analysis (DIVA) (Ronquist 1997)

Encadré 2 Biogéographie systématique (d'après Parenti et Ebach 2009)

- Biogéographie descriptive
 - Distribution maps (sensu Zimmerman 1777)
 - Biogeographic maps (sensu Lamarck et Candolle 1805)
 - Panbiogeography (Croizat 1964, Craw et al. 1999)
- Méthodes cladistiques
 - Component analysis (Nelson and Platnick 1981, Nelson 1984)
 - Assumptions 1 and 2 (Nelson and Platnick 1981)
 - Paralogy-free Subtree analysis (Nelson and Ladiges 1996)

Ce point de méthodologie étant fait, nous aimerions revenir sur les problématiques que nous allons aborder dans ce document. Ainsi que nous l'avons dit, la principale interrogation de la Biogéographie concerne l'étude de la distribution des taxons sur la Terre. Deux régions du monde sont tout particulièrement adaptées à ce genre de questionnement : l'Amérique centrale et les caraïbes et l'Asie du Sud-Est. Nous avons choisi de nous pencher sur l'Asie du Sud-Est et avons étendu la région d'étude à tout l'Ouest du Pacifique. Ce choix est d'autant plus justifié que c'est le lieu même de la découverte de la « Théorie de l'Évolution par sélection naturelle » énoncé par Alfred Russell Wallace (Encadré 3). Nous tâcherons, dans les chapitres I et II, de répondre aux questions « **pourquoi** et

comment » appliquées à la région du Sud-Est asiatique et de l'Ouest du Pacifique. Enfin, nous tenterons de répondre à la question « **quand** » dans le chapitre III.

Encadré 3 Wallace et le Sud-Est asiatique

Alfred Russel Wallace est né en 1823 proche du village de Usk sur la frontière galloise, d'une famille plutôt modeste. Ce qui l'amena, d'ailleurs, à arrêter son enseignement académique assez tôt, à l'âge de 14 ans. Il est connu pour être un autodidacte, contrairement à son non moins prestigieux contemporain: Charles Darwin qui, quant à lui, n'eut jamais à travailler pour vivre.

Wallace commença sa réelle carrière de naturaliste par un premier voyage dans le bassin Amazonien en 1848. Il écrivit plusieurs livres sur la région dont le premier "Palm trees of the Amazon" (Wallace, 1853a) a été beaucoup critiqué. Il décida ensuite de retourner sous les tropiques mais cette fois-ci dans une tout autre région du globe : le Sud-Est asiatique ! Il fit de Singapour son camp de base pendant les 8 années qu'il a passé à explorer la région et les nombreuses îles environnantes. En totalité, il a collecté 125 660 spécimens de collection (van Wyhe 2012), dont plus de 80 000 scarabées. Il a proposé une ligne de séparation, connue sous le nom de « ligne de Wallace » (voir Encadré 10), qui distingue les faunes typiquement Australiennes des faunes Asiatiques. Wallace est souvent considéré comme le « père » de la Biogéographie moderne de part l'ensemble de ses travaux de recherche (van Oosterzee 1997). Il est étonnamment bien moins célèbre que Charles Darwin et pourtant il a élaboré sa théorie de l'Évolution sur le terrain, là-bas, dans le Sud-Est asiatique alors que Darwin a formalisé ses conclusions une fois revenu à Londres (van Wyhe 2012). Cette région du monde est incontestablement le lieu de la découverte de la théorie de l'Évolution par Wallace à la différence des îles Galápagos, rendues célèbres par Darwin mais qui finalement, ne furent pas le lieu de cette découverte scientifique.

NB : Nous voudrions informer le lecteur que nous avons, autant que possible, utilisé des termes français pour l'ensemble de notre texte, cependant, certains termes, comme « pattern » par exemple trouve tout leur sens en anglais et perdrait beaucoup trop de sémantique en les traduisant. Que nos lecteurs puristes nous pardonnent cet abus de style qui est toutefois nécessaire pour rendre notre propos intelligible et sans ambiguïté aucune.

Nous aimerions ajouter que nous emploierons, indifféremment, tout au long du document les termes suivants : Biogéographie historique, Biogéographie cladistique et Biogéographie comparative. Conscient que ces différents termes ne sont pas des synonymes mais présentent des spécificités trop fines pour être étayées ici, nous allons, sans doute de manière un peu abusive, les utiliser comme des équivalents.

Chapitre I : Contexte méthodologique et implémentation logicielle

*« The biogeographic literature is cluttered with a range of perspectives and methods around this modern
threefold parallelism of space, time and form and whether biogeography determines general patterns of
individual species histories. »*

Christopher J. Humphries, 2004 : 225.

La Biogéographie cladistique, historique, recherche les relations de parenté entre aires d'endémisme, tout comme la Systématique recherche des relations de parenté entre taxons, à savoir une classification naturelle des taxons ; il en découle que la Biogéographie recherche quant à elle une classification naturelle des aires d'endémisme. Dans ce premier chapitre, nous allons (1) présenter le principe de la méthode d'analyse à trois éléments, (2) présenter le logiciel LisBeth qui implémente cette méthode et enfin (3) discuter des unités ou objets d'analyse en Biogéographie : les aires biogéographiques et plus particulièrement les aires d'endémisme.

1. L'analyse à trois éléments pour la Biogéographie avec LisBeth

1.1 Principe de la méthode

La Biogéographie historique est la discipline qui s'occupe d'établir les relations de parenté entre aires biogéographiques, comprises comme des ensembles d'aires de distribution de taxons. L'analyse cladistique a été appliquée à la Biogéographie historique depuis les travaux de Nelson G. J. et Platnick N. I. (1981), Nelson, G. J. et Ladiges P. Y. (1991a, 1992b). Ces auteurs ont montré l'équivalence entre les théories cladistiques en Systématique et en Biogéographie (voir aussi chapitre III figure 9 pour plus de détails). En Systématique, les relations hiérarchiques de taxons, c'est-à-dire d'emboîtement les uns dans les autres, sont obtenues par analyse de leurs caractères. En Biogéographie, les relations hiérarchiques entre aires biogéographiques s'obtiennent par analyse des cladogrammes de taxons après remplacement des taxons terminaux par les aires correspondantes. Dans les études phylogénétiques, la solution finale est l'arbre le plus congruent entre tous les arbres reconstruits. Similairement, concernant l'analyse biogéographique, le modèle le plus congruent avec tous les arbres nous informe sur le processus évolutif des aires biogéographiques étudiées.

Finalement, ce que l'on teste, n'est ni plus ni moins, que la congruence des aires homologues. Morrone en 2009 précisait que : « *Biogeographic homology is the basic concept of evolutionary biogeography (Morrone 2004a). In its more general form, homology means equivalence of parts and constitutes a sorting procedure used to establish meaningful comparisons within a hierarchical system (de Pinna 1991, Nelson 1994, Rieppel 2004, Williams 2004). Biogeographic homology allows one to identify biotic components, namely, the set of spatiotemporally integrated taxa that coexist in given areas.* » (Morrone 2009 :12).

Parmi toutes les méthodes existantes (Encadré 2) nous avons choisi d'appliquer l'analyse à 3 éléments (3ia), une méthode cladistique de reconstruction phylogénétique qui est également parfaitement adaptée pour des

analyses de Biogéographie (Nelson et Ladiges 1991a, b). En 3ia, la représentation des caractères ne se fait pas sous forme matricielle mais directement comme des hiérarchies ou des arbres racinés d'états de caractère (Cao et al. 2007). Chaque caractère est ensuite décomposé en une suite élémentaire d'assertions à 3 éléments (3is pour *Three-item statements*), chacun représentant la hiérarchie minimale, c'est-à-dire que pour 3 entités (des taxons terminaux pour la Systématique ou des aires d'endémisme pour la Biogéographie), deux d'entre eux sont plus proche l'un de l'autre qu'ils ne le sont du 3^{ème}. La structure se présente donc toujours de la sorte : (C, (A, B)) (Figure 1). La méthode de pondération fractionnaire énoncée par Nelson et Ladiges (1992) leur est ensuite appliquée.

L'analyse de congruence recherchant l'arbre d'intersection (Cao et al 2007) se base sur le principe de la compatibilité (Estabrook et al. 1976), les relations de parenté entre taxons terminaux ou aires d'endémisme sont obtenues en combinant les 3is déduits de tous les caractères auxquels sont associés leur pondérations fractionnaires. Les arbres optimaux (c'est-à-dire les plus compatibles) sont ceux qui présentent le nombre maximal de 3is compatibles (avec la somme de pondération fractionnaire maximale) (Cao 2008).

Nous renvoyons le lecteur à la littérature existante (citée précédemment) pour plus de détails sur la méthode.

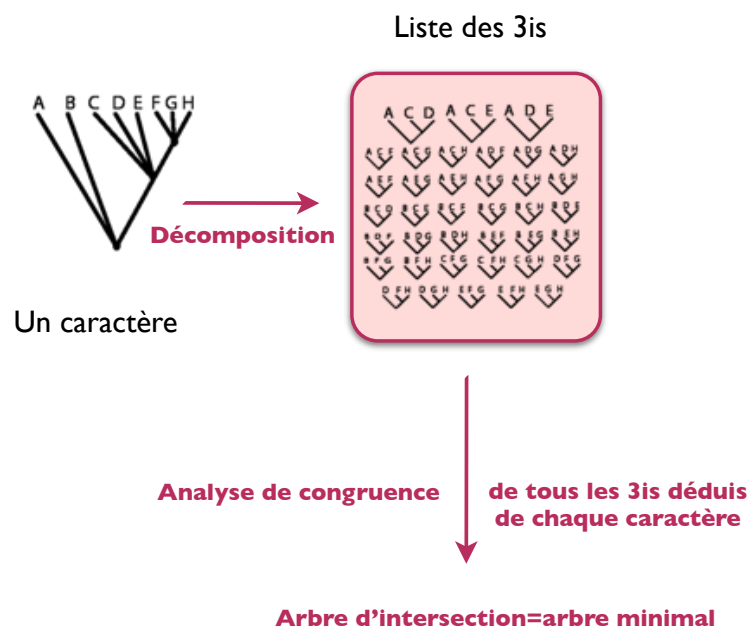


Figure 1 Schéma présentant les étapes d'une analyse 3ia

1.2 Particularités de la Biogéographie

Lorsque la méthode est appliquée à la Biogéographie, les taxons terminaux sont remplacés (Figure 2, Figure 4) par les aires dans lesquels nous trouvons ces taxons (et nous obtenons ainsi une structure hybride que l'on appelle

Taxon-Area cladogram (Figure 3) (ou TAC), ce qui peut conduire à deux problèmes bien connus que sont : la redondance d'aires (due à de la paralogie taxique, c'est-à-dire, une diversification taxique antérieure à la duplication des aires) (Figure 3) et la présence de taxons à large répartition (ou Multiple Areas in Single Terminals, MAST) (Figure 5). Nous avons implémenté dans LisBeth (voir paragraphe suivant) la méthode de « Paralogy-free subtree analysis » (Nelson et Ladiges 1996) et la méthode transparente (Ebach et al. 2005) pour respectivement trouver une solution à la redondance d'aires et aux taxons à large répartition.

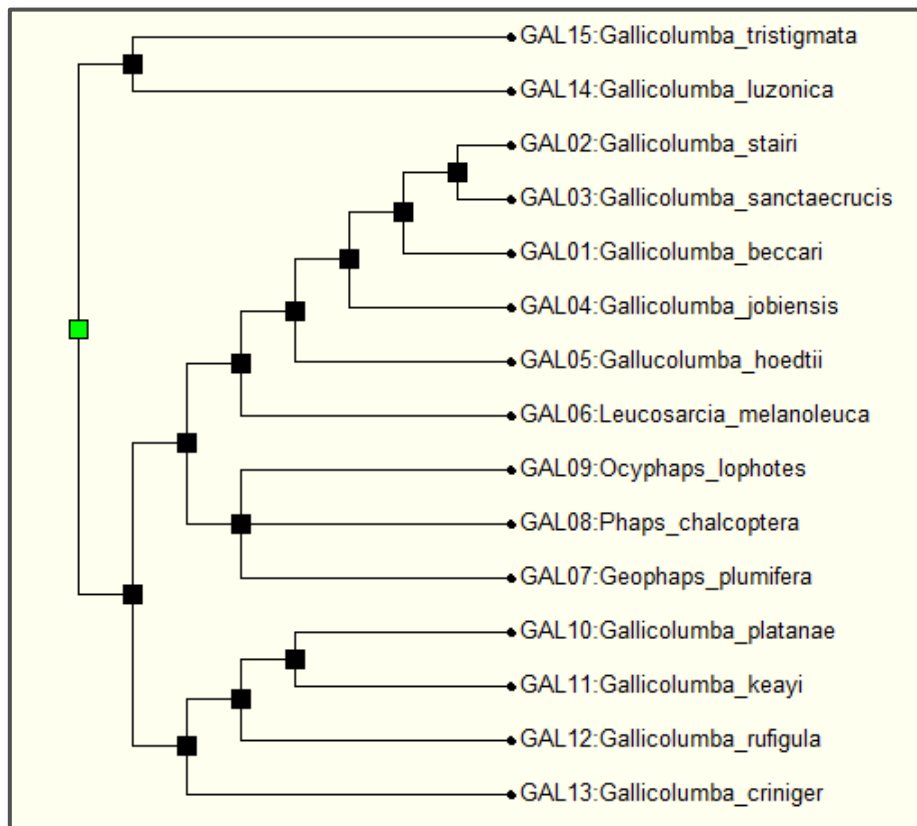


Figure 2 Cladogramme du taxon *Gallicolumba* utilisé dans l'analyse de Biogéographie comparative du Chapitre II (interface de LisBeth)

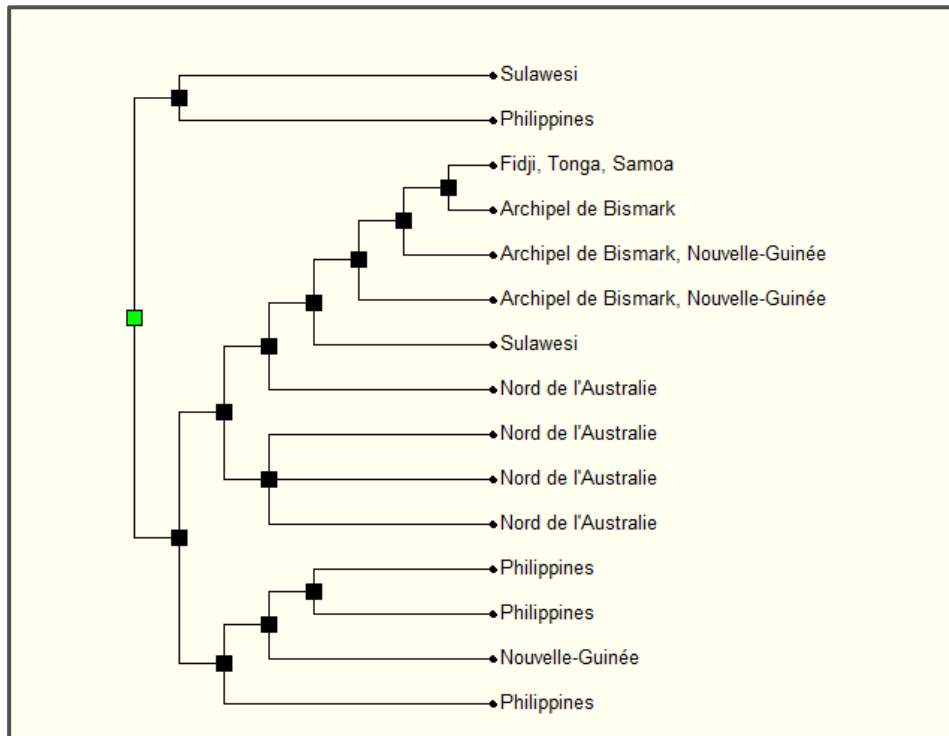


Figure 3 Le même cladogramme dans lequel les taxons terminaux ont été substitués par les aires d'endémisme où sont présents les taxons terminaux (interface de LisBeth)

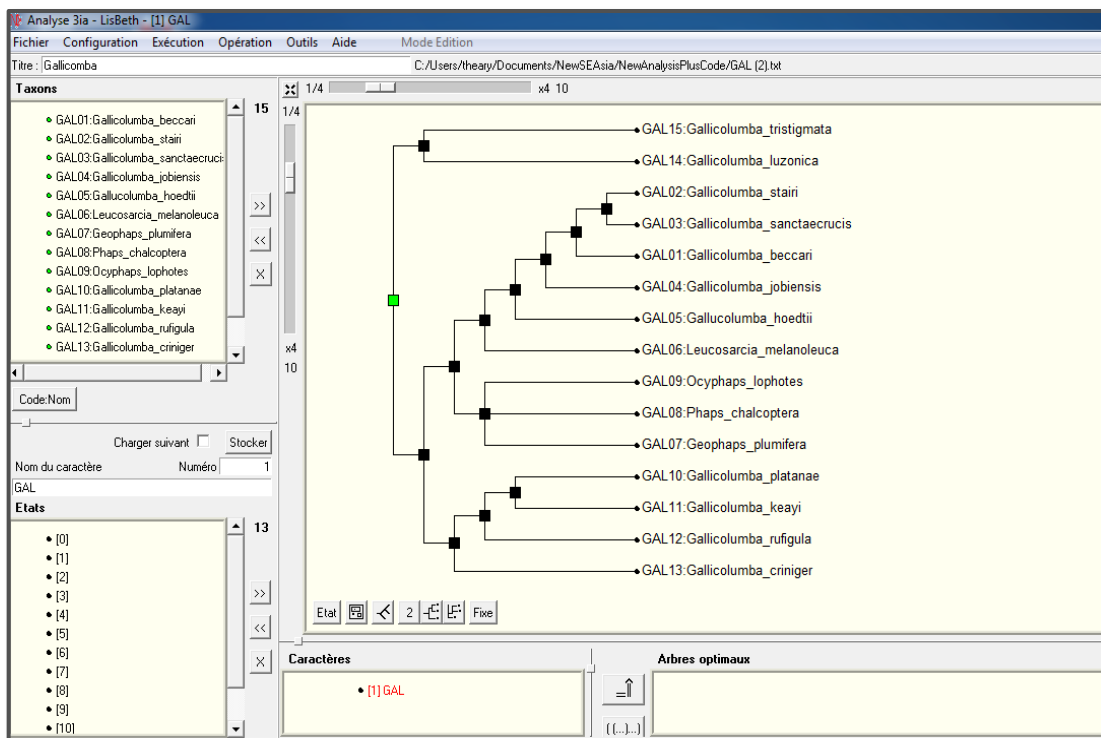


Figure 4 Le même cladogramme via l'interface entière de LisBeth

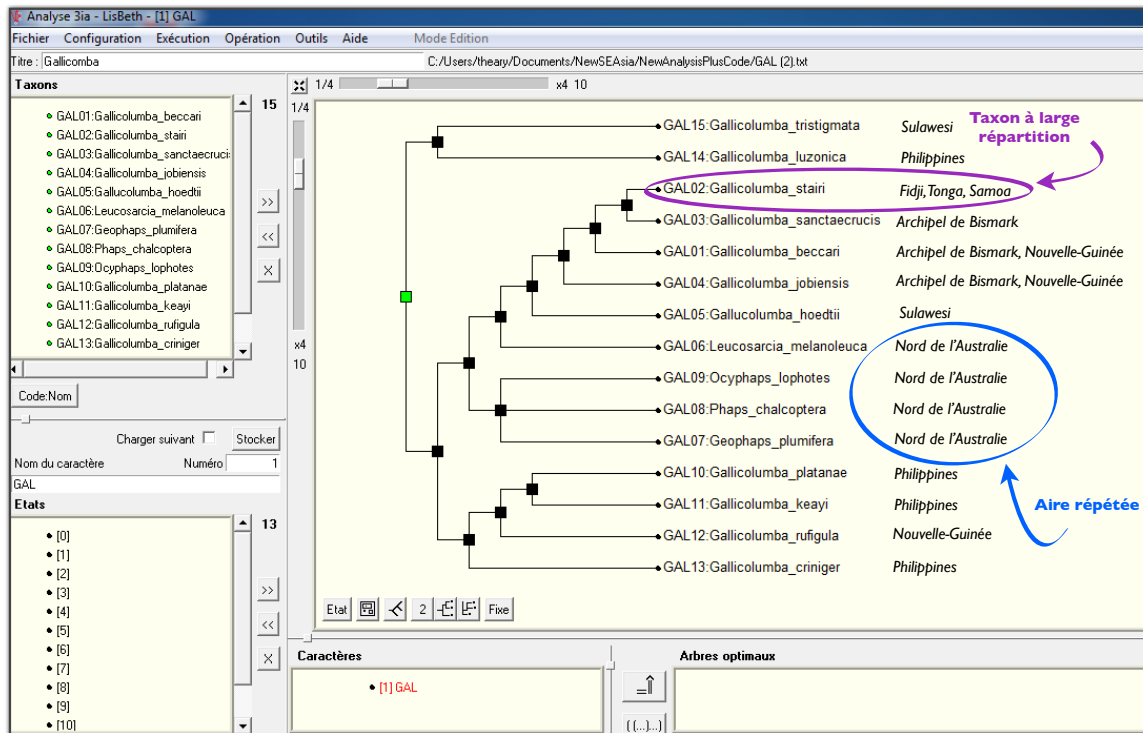


Figure 5 Le même cladogramme de taxon sur lequel nous avons superposé la distribution des différentes espèces et indiqué un exemple de taxon à large répartition ainsi qu'un aire d'aire répétée

2. LisBeth

Les développements autour d'un programme implémentant la 3ia ont commencé en 2005 avec les travaux de thèse de Nathanaël Cao et surtout grâce à l'implication de Jacques Ducasse, chercheur bénévole au laboratoire, sans qui le logiciel n'aurait pas vu le jour. Le premier programme : Nelson05 a vite été remplacé par LisBeth en 2011-2012 avec les évolutions logicielles et l'orientation Biogéographique (Lisbeth = Laboratoire Informatique et Systématique program for Biogeography and Evolution using THree-item analysis).

L'article présentant cette nouvelle interface est parue en 2012 : « Zaragüeta-Bagils R, Ung V., Grand A., Vignes-Lebbe R., Cao N., Ducasse J, 2012 : LisBeth : New cladistics for phylogenetics and biogeography. C.R Palevol 11(8):563-566. »

Quelques références utilisant LisBeth pour la Systématique :

1. Bourdon E. (2006) L'avifaune du Paléogène des phosphates du Maroc et du Togo. Diversité, systématique et apports à la connaissance de la diversification des oiseaux modernes (Neornithes). Thèse de doctorat (MNHN, Paris France) ;
2. Cao N. (2008) Analyse à trois éléments et anatomie du bois des Fagales Engl. Thèse de doctorat (MNHN, Paris, France) ;
3. Corvez A. (2012) L'origine de la mégaphylle chez les Monilophytes. Thèse de doctorat (MNHN, Paris, France) ;
4. Grand A., Corvez A, Duque Velez LM, Laurin M (2013) Phylogenetic inference using discrete characters: performance of ordered and unordered parsimony and of three-item statements. *Biol J Linnean Soc* (sous presse) ;
5. Grand A., Zaragüeta-Bagils R., Duque-Velez L. M. & Ung V. (2013) Three-item analysis and Gadiformes (Teleostei, Paracanthopterygii): how to conduct a cladistic analysis using LisBeth. *Zootaxa* (révisé, soumis).

Pour la Biogéographie:

6. Escalante, T.; Rodríguez, G.; Cao, N.; Ebach, M.C. & Morrone, J.J. (2007). Cladistic biogeographic analysis suggests an early Caribbean diversification in Mexico. *Naturwissenschaften*, Vol. 94, pp. 561—565, ISSN 0378-1844 ;
7. Cao N., Bourdon E., El Azawi M., & Zaragüeta-Bagils R. (2009) Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example. *Bulletin de la Société Géologique de France*, 180 (1) : 13-15 ;
8. Parenti L. et Ebach M. C. (2009) *Comparative Biogeography Discovering and Classifying Biogeographical Patterns of a Dynamic Earth*. Berkeley (CA): University of California Press. p. 312 ;
9. Hoagstrom C. W., Ung V., Taylor K. (2013) Miocene Rivers and Tertiary Taxon Cycles Clarify the Comparative Biogeography of North American Highland Fishes. *Journal of Biogeography* (accepté);
10. Ung V. (2013) Bridging the temporal gap in Biogeography. *Journal of Biogeography* (en préparation) ;
11. Ung V. Zaragüeta-Bagils R. & Williams D. M. (2013) Comparative biogeography of Southeast Asia and the West Pacific region. *Journal of Biogeography* (en révision).

Concernant les outils de métrologie que nous avons mis en place (depuis juillet 2012), nous comptons 623 téléchargements de l'application ainsi que 128 inscrits sur notre liste de diffusion. (lisbeth@listes.upmc.fr). L'article a aussi un bon rang en terme de téléchargement, c'est le 29ème article le plus téléchargé en 2012 parmi tous les articles publiés par la revue *CR Palevol* (toutes années confondues depuis 2002) (Laurin 2013, communication personnelle, rapport Elsevier pour le comité éditorial des *CR Palevol*).



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

Mots clés :

Cladistique

Évolution

Phylogénétique

Biogéographie historique

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 Xper² knowledge bases.

© 2012 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

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 Xper².

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

* Corresponding author.

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

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 (BC)). 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.

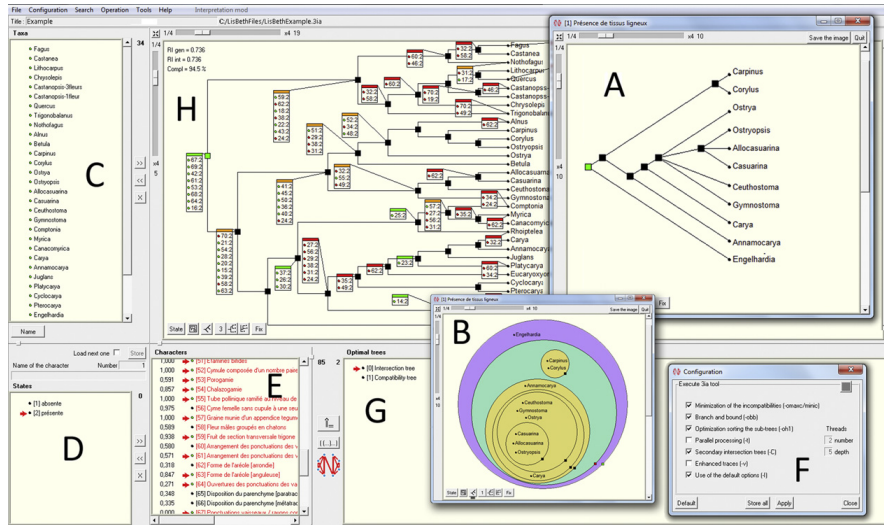


Fig. 1. LisBeth GUI. A. Character displayed as a rooted tree. B. Character displayed as a Venn diagram. C. List of taxa (or endemic areas). D. Description of characters and character-states. E. List of characters. F. Configuration window. G. List of the intersection tree and optimal trees. H. Display of characters or cladograms, of the 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 recombinates 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.

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

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 Xper² (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, J.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. Xper²: 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.

3. Utilisation de LisBeth pour la Biogéographie

3.1 Saisie des cladogrammes de taxons via l'interface graphique

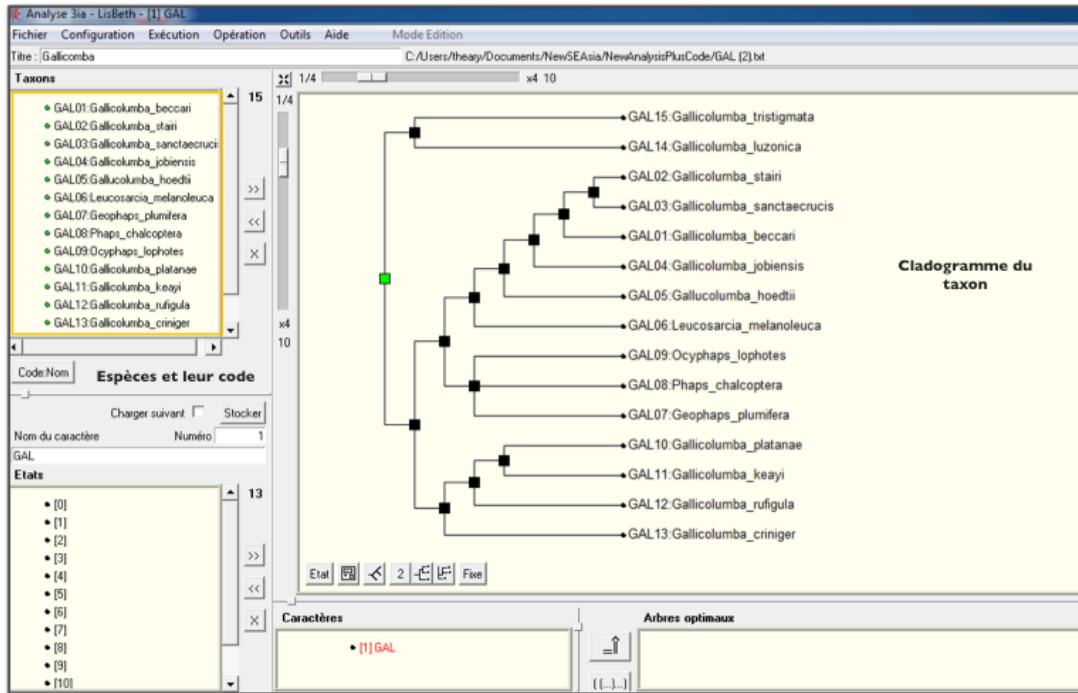


Figure 6 Interface de saisie de LisBeth

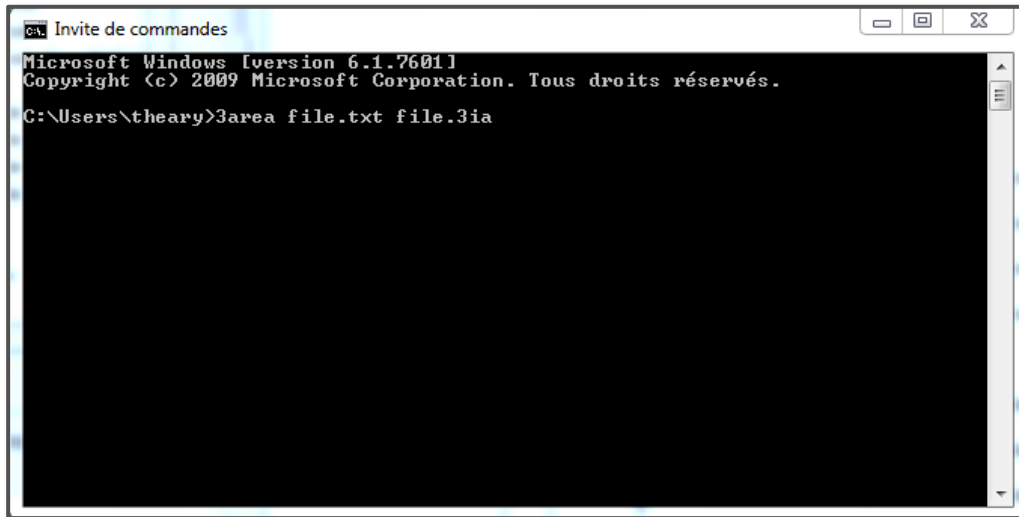
LisBeth peut être utilisée comme un éditeur de cladogrammes, chaque taxon est saisi l'un à la suite de l'autre et compose ainsi un fichier global contenant tous les taxons de l'analyse.

3.2 Substitution des taxons terminaux par les aires d'endémisme et application de la « Paralogy-free subtree analysis » et de la méthode transparente

Cette étape est automatiquement faite par LisBeth via le sous-module *3area* mais ne dispose pas encore d'interface graphique pour le faire. Il est donc nécessaire de lancer une invite de commande et de saisir la commande suivante (Figure 7):

```
>3area file.txt file.3ia
```

Dans les encadrés 4 et 5 figurent un exemple de fichier .txt et de fichier .3ia.



```
CA Invite de commandes
Microsoft Windows [version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Tous droits réservés.
C:\Users\theary>3area file.txt file.3ia
```

Figure 7 Lancement du sous-module 3area pour la Biogéographie

Encadré 4 Squelette du format de fichier .txt pour la Biogéographie, exemple du taxon *Gallicolumba*

```
Gallicolumba
;
Dimensions
  ntax=15 nchar=1
;
Charlist
  [1] GAL
;

Taxa

GAL01 = Gallicolumba_beccari
GAL02 = Gallicolumba_stairi
GAL03 = Gallicolumba_sanctaecrucis
GAL04 = Gallicolumba_jobiensis
GAL05 = Gallucolumba_hoedtii
GAL06 = Leucosarcia_melanoleuca
GAL07 = Geophaps_plumifera
GAL08 = Phaps_chalcoptera
GAL09 = Ocyphaps_lophotes
GAL10 = Gallicolumba_platanae
GAL11 = Gallicolumba_keyi
GAL12 = Gallicolumba_rufigula
GAL13 = Gallicolumba_criniger
GAL14 = Gallicolumba_luzonica
GAL15 = Gallicolumba_tristigmata
;

Areas

Bismark_Solo:      GAL01 GAL03 GAL04
North_Aust:        GAL06 GAL07 GAL08 GAL09
NewGuinea:         GAL01 GAL04 GAL12
Sulawesi:          GAL05 GAL15
Philippines:       GAL10 GAL11 GAL13 GAL14
Fiji:              GAL02
Tonga:             GAL02
Samoa:             GAL02
;

Characters
[1]
((GAL15,GAL14),((((((GAL02,GAL03),GAL01),GAL04),GAL05),GAL06),(GAL09,GAL08,GAL07)),((GAL10,GAL11),GAL12),GAL13)) ;
```

Encadré 5 Exemple de fichier de sortie du sous-module 3area

```
# Fichier genere par ttsbiog
# a partir du fichier <C:\Users\theory\Documents\NewSEAsia\NewAnalysisPlusCode\GAL (2).txt>
# 0 h. 00 m. 00 s.

Gallicomba
;

Dimensions
ntax=8 nchar=15
;

Taxa
AA = Bismark_Solo
AB = North_Aust
AC = NewGuinea
AD = Sulawesi
AE = Philippines
AF = Fiji
AG = Tonga
AH = Samoa
;

Characters
#Source[1]
((((GAL01,(GAL02,GAL03)),GAL04),GAL05),GAL06),(GAL07,GAL08,GAL09)),((GAL10,GAL11),GAL12),GAL13)),(GAL14,GAL15))
# Renorm : [1.Exp0001] ((((((AA,AF),AA),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[1] (((((AA,AF),AD),AB),(AC,AE))
[2] (AB,(AC,AE))
# Renorm : [1.Exp0002] (((((((AA,AF),AA),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[3] (((((AA,AF),AC),AD),AB)
# Renorm : [1.Exp0003] (((((((AA,AF),AC),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[4] (((((AA,AF),AC),AD),AB)
# Renorm : [1.Exp0004] (((((((AA,AF),AC),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[5] (((((AA,AF),AC),AD),AB)
# Renorm : [1.Exp0005] (((((((AA,AG),AA),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[6] (((((AA,AG),AD),AB),(AC,AE))
[7] (AB,(AC,AE))
# Renorm : [1.Exp0006] (((((((AA,AG),AA),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[8] (((((AA,AG),AC),AD),AB)
# Renorm : [1.Exp0007] (((((((AA,AG),AC),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[9] (((((AA,AG),AC),AD),AB)
# Renorm : [1.Exp0008] (((((((AA,AG),AC),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[10] (((((AA,AG),AC),AD),AB)
# Renorm : [1.Exp0009] (((((((AA,AH),AA),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[11] (((((AA,AH),AD),AB),(AC,AE))
[12] (AB,(AC,AE))
# Renorm : [1.Exp0010] (((((((AA,AH),AA),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[13] (((((AA,AH),AC),AD),AB)
# Renorm : [1.Exp0011] (((((((AA,AH),AC),AA),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[14] (((((AA,AH),AC),AD),AB)
# Renorm : [1.Exp0012] (((((((AA,AH),AC),AC),AD),AB),(AB,AB,AB)),((AC,(AE,AE)),AE)),(AD,AE))
[15] (((((AA,AH),AC),AD),AB)
;
```

Les TACs (sous-arbres hybrides) « libres de paralogie et de MASTs » (15 dans l'exemple de l'Encadré 5) constituent les caractères de l'analyse biogéographique en tant que telle (voir Figure 29 du chapitre III).

Une fois cette étape réalisée, la suite de l'analyse peut se faire directement via l'interface de LisBeth (nous renvoyons le lecteur au manuel d'utilisation de LisBeth disponible à cette adresse pour plus de détails : <http://www.infosyslab.fr/lis/?q=fr/ressources/logiciels/lisbeth/documentation>).

4. Améliorations du module de Biogéographie

4.1 Interface utilisateur

La première concerne l'interface utilisateur spécifique à la Biogéographie. LisBeth ne dispose pas à l'heure actuelle d'interface utilisateur ce qui rend la prise en main assez fastidieuse. Nous avons proposé les interfaces suivantes en vue d'améliorer l'ergonomie de notre outil.

4.1.1 Création du fichier .txt pour 3area (module réalisant la paralogy-free sub-tree analysis et la méthode transparente)

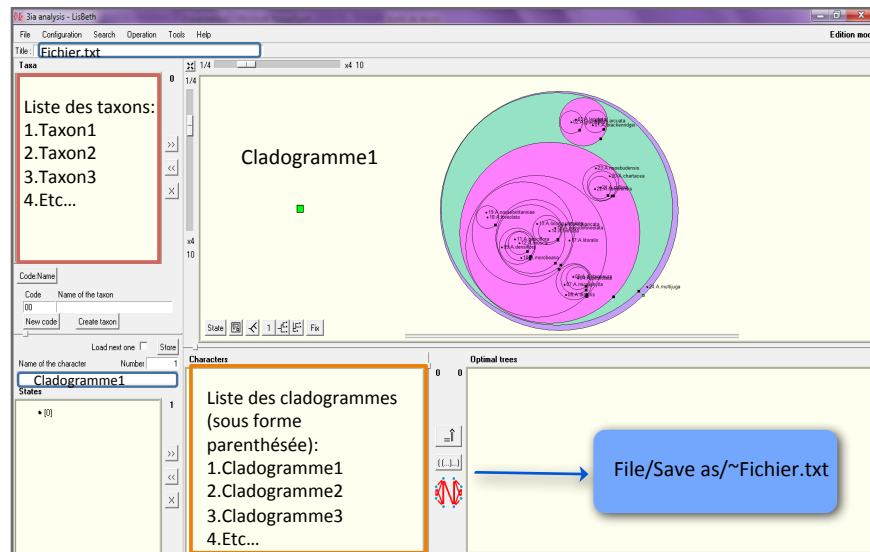


Figure 8 Saisie des cladogrammes de taxons pour la création du fichier texte

4.1.2 Saisie de la distribution des taxons

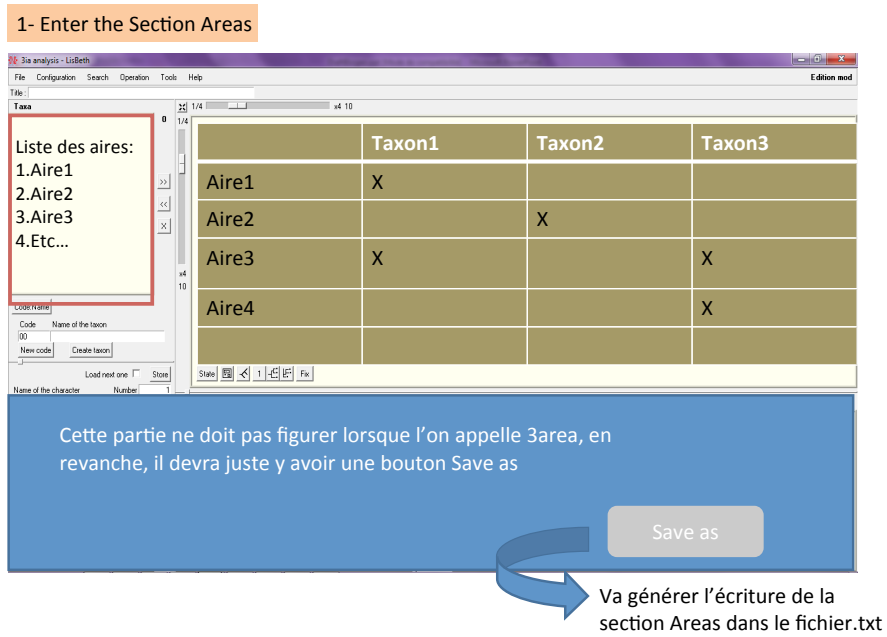


Figure 9 Proposition d'interface pour la saisie de la distribution des taxons

4.1.3 Lancement de 3area

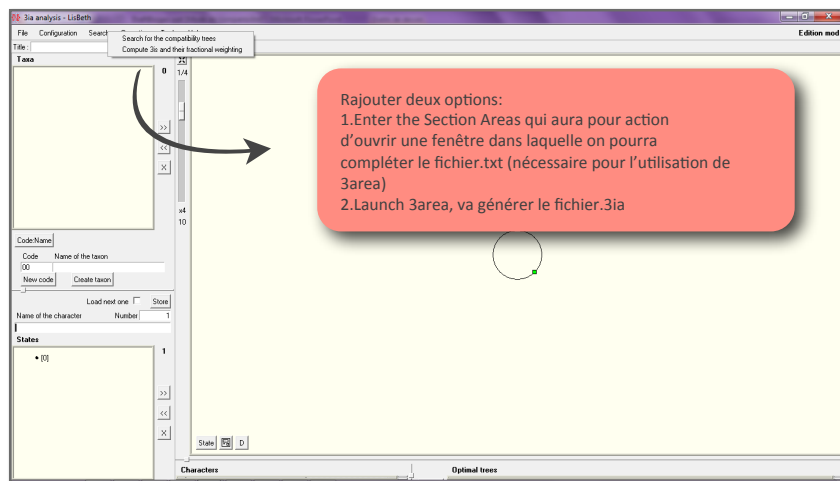


Figure 10 Proposition d'ajout d'options

4.1.4 Lancement de 3ia (sous-programme réalisant à proprement parlé l'analyse : décomposition en 3is, analyse de congruence et reconstruction de l'arbre d'intersection)

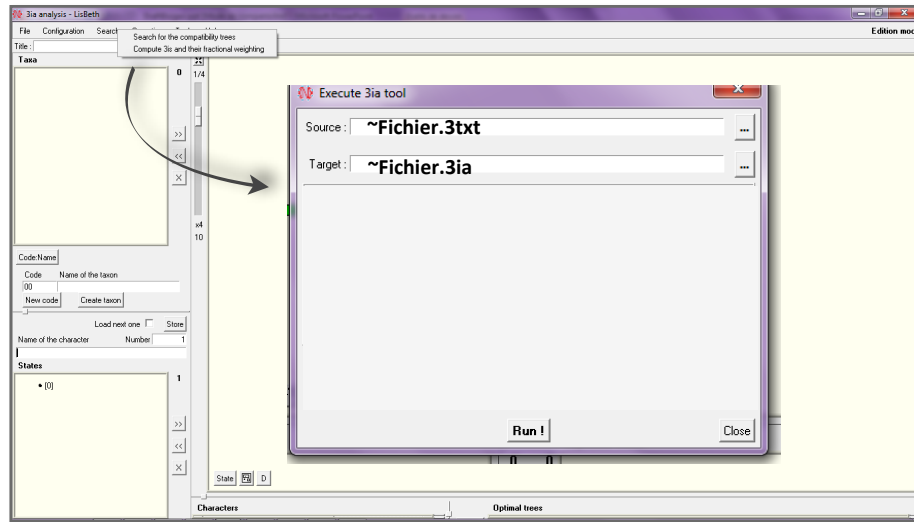


Figure 11 Menu de lancement de l'analyse via l'interface de LisBeth

5. Création d'une « boîte à outils »

Le lecteur est prié de se référer à l'Annexe I du présent manuscrit pour voir le détail du contenu de cette nouvelle fonctionnalité dans LisBeth. Nous avons choisi de fournir aux utilisateurs de LisBeth tout un ensemble de programmes indépendants afin de proposer une granularité des plus fines si nécessaire. Entre autre, la principale amélioration concerne la traçabilité des caractères. Pour la Systématique, cette traçabilité revient au « mode Interprétation » déjà disponible dans LisBeth et replace les synapomorphies aux nœuds (Zaragüeta Bagils et al 2012). Pour La Biogéographie, cette fonctionnalité permet de garder la trace des taxons issus des cladogrammes sources ayant servis à reconstruire l'arbre d'intersection c'est à dire l'aréogramme. Si nous continuons notre parallèle entre les deux disciplines, nous pouvons dire que pour la Biogéographie, les « synapomorphies » sont les ensembles de taxons que l'on retrouve aux nœuds internes de l'aréogramme (voir chapitre III Figure 29) or jusqu'à présent, il n'en a jamais été question. Aucune étude biogéographique n'a jamais évoqué ces « synapomorphies » biogéographiques. Grâce à cette fonctionnalité, il est possible, dorénavant de garder la trace de ces taxons, et par la même, de connaître quels sont les « taxons synapomorphiques » ayant

servis à reconstruire l'arbre. Même si nous ne pouvons pas parler ici de « synapomorphie vraie » *sensu* Brundin (Brundin 1966 dans Williams 2004), il n'en reste pas moins que cette fonctionnalité est une première étape des plus avant-gardiste car elle offre enfin un argument analytique pour la discipline. Les relations de parenté entre aires d'endémisme sont caractérisées par une combinaison de taxons présents, les « taxons synapomorphiques ».

6. Unités (objets) d'analyse en Biogéographie : vers une représentation sémantique des caractères biogéographiques

S'intéresser à la Biogéographie historique implique inévitablement de s'intéresser, à un moment donné, aux unités (ou objets ou encore « briques ») de base auxquelles on applique des méthodes. Dans notre cas, il s'agit d'aires d'endémisme (qui, en définitive, ne sont que des cas particuliers d'aires biogéographiques, voir Encadré 6).

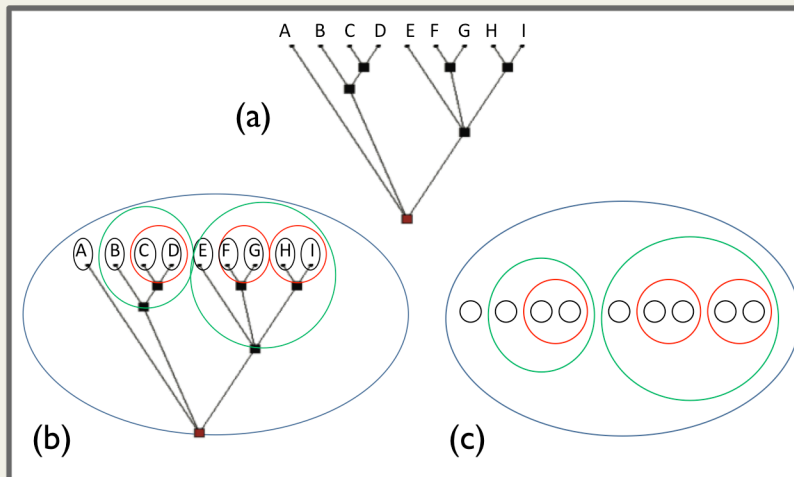
La réflexion autour de ce point précis nous a amené à la préparation de 3 articles, le premier (1) est une réponse à un article publié en 2011 décrivant une « ontologie pour les aires d'endémisme », le second (2) se voulait être une réponse à un article plus récemment publié (Avril 2013) mais finalement nous avons décidé d'étayer un peu plus notre propos, et seulement mis en ligne sur le forum de « Nature » une première réponse très succincte, et le troisième (3) décrit de manière succincte mais précise pourquoi la Biogéographie comparative peut s'ancrer dans le paradigme de la conception sémantique.

Encadré 6 Aires d'endémisme/ aire biogéographiques

Plusieurs définitions co-existent et finalement, chacun s'autorise à utiliser celle qui lui convient le mieux. Parmi les plus couramment utilisées, on peut mentionner :

- « an area is the congruent distributional limits of two or more species » Platnick 1991 ;
- « an area occupied by two taxa, with overlapping area identified as a separate area » Axelieus 1991 ;
- « an area is a geographic region comprising the distributions of two or more monophyletic taxa that exhibit a phylogenetic and distributional congruence and having their respective relatives occurring in other such-defined regions » Harold and Mooi 1994 ;
- « an area is recognized for the coincident distribution of two or more organisms » Humphries et Parenti 1999.

Dans notre cadre méthodologique de représentation hiérarchique des caractères (ici des aires biogéographiques), les aires d'endémisme sont des cas particuliers d'aires biogéographiques. Sur le schéma suivant, en (a) est représenté un cladogramme d'aires (*id est* un aréagramme), en (b) le même aréagramme avec la représentation en diagramme de Venn superposé et en (c) que la représentation sous forme de diagramme de Venn. Dans cet exemple, A, B, C, D, E, F, G, H et I sont des aires d'endémisme et les nœuds internes des aires biogéographiques.



6.1 « On the ontology of areas of endemism »

6.1.1 Contexte

La publication d'un article en 2011 : « Ontology of areas of endemism » par Crother et Murray a suscité un vif intérêt dans notre équipe et nous avons collaboré avec deux étudiants en philosophie des sciences sur l'écriture d'une réponse à leur publication.

Avec cet article, pour la première fois, il était question de la nature ontologique des aires d'endémisme. Il existe beaucoup de références autour de la définition/délimitation/reconnaissance des aires d'endémisme (Nelson et Platnick 1981, Axelius 1991, Platnick 1991, Harold et Mooi 1994, Hovenkamp 1997, Humphries et Parenti 1999 parmi d'autres). Mais jusqu'alors, rien n'avait été énoncé autour de leur nature propre. Cet article est essentiel car il pose pour la première fois la question qui sous-tend toute la théorie de la Biogéographie.

6.1.2 Résumé

Le concept d'aire d'endémisme est crucial pour les études biogéographiques. Un certain nombre d'auteurs ont déjà tenté de le clarifier d'un point de vue théorique et méthodologique, par exemple Axelius (1991) ou Harold et Moi (1994). Pour la première fois, Crother et Murray (2011, 2013) porte le problème dans le domaine de l'« ontologie ». Ils proposent une « ontologie » pour les aires d'endémisme conformément à l'ontologie des espèces exposés par Guiselin (1974) et Hull (1978). Une compréhension adéquate du concept d'aire d'endémisme est indéniablement primordiale pour la Biogéographie, nous pensons, cependant, que l'approche de Crother et Murray n'est pas suffisamment satisfaisante. Nous allons discuter un certain nombre de problèmes de leur argumentation et nous préciserons notre point de vue.

6.1.3 Conclusions

Cet article de Crother et Murray est essentiel parce qu'ils pointent du doigt un problème conceptuel complexe : la signification des concepts biogéographiques et des aires d'endémisme (leur unité de base). Néanmoins, nous pensons que leur conclusion est prématurée et que leur argument manque de contenu philosophique. Dans ce court commentaire, nous avons montré que (1) pour pouvoir comprendre des concepts scientifiques, nous devons résister à la tentation de proposer des ontologies pour chacun d'eux mais plutôt d'essayer de comprendre ce qu'ils veulent dire et comment ils sont utilisés selon les contextes, et que (2) dans le contexte de la Biogéographie historique, les aires sont caractérisées par intension avec la distribution géographique des taxons et donc, comprises comme des classes (au sens mathématique).

On the ontology of areas of endemism

Sophie Pécaud¹, Visotheary Ung², Stéphane Prin³ and René Zaragüeta Bagils²

¹ Institut d'histoire et de philosophie des sciences et des techniques, IHPST UMR 8590 CNRS/Paris 1/ENS, Université Paris-1–Panthéon Sorbonne, 13, rue du Four, 75006 Paris, France

² UPMC Univ Paris 06, CR2P - UMR 7207 CNRS, MNHN, Univ Paris 06, 4 Place Jussieu, Tour 46-56, 5ème étage, F-75005, Paris, France

³ Origine, structure et évolution de la biodiversité, OSEB UMR 7205 CNRS MNHN, 45 rue Buffon, 75005 Paris, France

Summary

The concept of areas of endemism is crucial for biogeographical studies. A number of authors have already faced the task of clarifying it from both a theoretical and methodological perspective, *e.g.* Axelius (1991) or Harold and Mooi (1994). For the first time, Crother and Murray (2011, 2013) presented problem in the context of “ontology:” They aim at presenting an “ontology” of areas of endemism aligned with Ghiselin’s (1974) and Hull’s (1978) ontology of species. An adequate understanding of the concept of “area of endemism” is undeniably of the most importance for biogeography; yet we do not think that Crother and Murray’s solution is satisfactory. Here we discuss a number of problems raised by their argument and articulate our own point of view.

Keywords

Area of endemism, physical geographical area, biogeographical area, individual, class, classificatory structure, historical biogeography

Introduction

In their article “Ontology of areas of endemism” Crother and Murray (2011) argue that areas of endemism should not be considered as classes, but rather as individuals *sensu* Ghiselin (1974) and Hull (1978). In this commentary, we discuss some of the conceptual confusions we suspect this thesis to be based on (i.e. the species as individuals concept, the argument with respect to particulars and instances, definition by intension or ostension and finally the boundaries problem), and we maintain that biogeographical areas, including areas of endemism, are best understood as classes, i.e. as parts of classificatory structures *sensu* Lebbe (1998: 422 *sq.*).

Are areas of endemism individuals because species are individuals?

In the domain of systematics, species-as-individuals (SAI) have been proposed to resolve the problem of essentialism, which says that “a taxon can be defined by describing traits that all and only its members possess”

(Sober 2000: 182). Since evolutionary phenomena like secondary losses prevent a definition of species (as well as more inclusive taxa) in terms of essential, i.e. necessary and sufficient properties (e.g. “chiridian limbs” cannot be a necessary and sufficient property of the taxon *Tetrapoda* since some tetrapods are legless), Ghiselin (1974), soon followed by Hull (1978), proposed that species should be defined. Crother and Murray (2011) (CM hereafter) carried this line of reasoning into historical biogeography. Since “there is no list of specific characteristics that will always define areas of endemism” (CM 2011: 1010), they proposed that those areas should also be defined as historical individuals.

Ghiselin’s thesis is well known: species are unique occurrences; once disappeared, they cannot reappear. Ruse (1995), amongst others, noted that this thesis is mostly based on an intuition: if I met someone who looked exactly like Napoleon, I would not conclude that this person actually *is* Napoleon (Ruse, 1995: 183). In the SAI thesis, this intuition is transposed to species: although the cloned *Tyrannosaurus rex* of *Jurassic Park* looks exactly like the *T. rex* of the Late Cretaceous, I would not conclude it is a true *T. rex*.

The use of intuition is the main persuasive force of the argument as well as its greatest logical weakness: who knows if, in between the jaws of the cloned tyrannosaurus of *Jurassic Park*, my intuition would not be different? Ruse, at least, has no hesitation:

“As I disappeared down the throat of a recreated tyrannosaurus, I know where my intuition would be” (Ruse, 1995:183).

Particulars and instance: are biogeographical and physical geographical areas the same?

In their attempt to clarify the ontological status of areas of endemism, CM state that “areas themselves,” and not only areas of endemism, are individuals in that they are “wholes” made up of “parts,” and not “classes” which can be “instantiated.” To support this idea, they simply combine a set of example: Jamaica, Lake Tangyanika or Lower Central America are such “wholes” made up of “parts;” they cannot be instantiated and therefore are not classes. This argument relies on the conflation between the physical geographical and the biogeographical concepts of area; it actually relies on the *reduction* of the biogeographical concept of area to the physical geographical one. However, if the meaning of the concept of area of endemism is to be found, it is not in physical geography, but in biogeography itself. Areas of endemism are *not* physical geographical areas, they are *not* composed of “physical parts” made up of soil and stones; they are *geographical distributions of taxa*—since what provides epistemic access to relationships among biogeographical areas are relationships among taxa. A particular organism (e.g. a *C. duranii*) at a particular place (upper Dordogne and Loire drainage) may instantiate a biogeographical area (Central France) but it cannot be understood as being a part of it.

This confusion highlights the necessity of criticizing CM’s very project of proposing an *ontology* of areas of endemism. One of the classical definitions of ontology is the study of the nature or essence of things, i.e. the study of their essential properties as opposed to their accidental ones. In seeking for the ultimate material substrate of areas of endemism in physical geography, CM search for the “essence” of such areas, their “essential” properties. This essentialist view of the problem drives them to reduce the different concepts of area to the

physical geographical one (understood as the more “essential”), and it prevents them from clarifying the very meaning of the biogeographical concept itself.

As a consequence, we must abandon the ontological or essentialist perspective: in order to understand the proper meaning of the biogeographical concept of area, we must not try to root it in the most “essential” concept of area we can find, but rather to understand what it means and how it is used in the context of biogeography. In other words, we must replace methodological essentialism by methodological nominalism *sensu* Popper (Popper, 1945: 34; Gayon, 2011: 130-132): defining what areas of endemism ultimately *are* belongs to metaphysics, whereas deciding to name “areas of endemism” a particular set of geographical distributions of taxa is a scientific decision which is justifiable (and therefore testable) within a given theoretical framework.

Definition: ontology and identification, characterization and description

CM claim that it is not possible to define areas of endemism by intension because “there is no list of specific characteristics that will always define areas of endemism” (CM, 2011: 1010), therefore, they propose that areas of endemism can only be defined by pointing out their endemic taxa (or endemic combinations of taxa), i.e. by ostension.

This assertion relies on confusion between *definition* and *identification*: from the possibility of *identifying* an area of endemism by pointing out its endemic taxa, CM are lead to conclude the necessity of *defining* areas of endemism by ostension. Saying that areas of endemism can be *identified* by ostension, i.e. by pointing out their endemic taxa is true and fruitful for biogeographical practice, i.e. description, naming, identification, and inventory of biogeographical areas. In this, we agree with CM when they say that “every [endemic] species above one simply adds to the diagnosis of the area of endemism” (CM, 2011: 1010). Nevertheless, it does not mean that areas of endemism should be ostensively *defined*.

From a logical point of view, identifying an object is “associating [it] to one or several classes of a classificatory structure” (Lebbe, 1998: 424). Usually, an object is associated to a class by checking that it meets certain requirements. Identifying an object therefore requires that a specific classificatory structure, i.e. a set of classes described in terms of properties, is *already* available; in other words, identification by ostension requires that intensional definitions (or simply descriptions) have already been made. What allows the identification of biogeographical areas by ostention, i.e., by the process of pointing out their endemic taxa, is the fact that “classificatory structures” (areagrams) made up of “classes” (biogeographical areas) and “properties” (geographical distributions of taxa) already exists.

Boundaries: essentialism and ontology, definition and classification

CM derive an ontological fact— areas of endemism “are clearly not classes” (CM 2011: 1012)—from an epistemological one—since “areas of endemism can evolve” (CM 2011: 1012), they “are not intensionally defined: there is no list of specific characteristics that will always define areas of endemism” (CM 2011: 1010). This statement is mistaken. While we agree that biogeographical areas—including areas of endemism—evolve,

and therefore cannot be defined by essence, i.e. necessary and sufficient properties, we do not understand why areas of endemism are not classes. Biogeographers do not try to *define* areas, but rather *search evidence* for their existence; the evidence they use are geographical distributions of taxa (or of sets of taxa). To make it clearer, let us here draw a comparison with systematics: Sober argued that, in systematics, traits do not *define* taxa, but are *evidence* for the existence of taxa. “The fact that sparrows and robins have wings, but snakes do not, is evidence that sparrows and robins belong to a monophyletic group that does not include snakes” (Sober, 2000: 182). If we carry this analysis into biogeography: geographical distributions of taxa are evidence for the existence of biogeographical areas just as traits are the evidence for the existence of taxa. For instance, the presence of *Cottus duranii* sculpins in upper Dordogne, Lot and Loire drainages, but not in Lake Geneva, is evidence for the existence of a biogeographical area that includes upper Dordogne, Lot and Loire drainages (let us call it “Central France”), but excludes Lake Geneva.

Since biogeographers do not try to define areas, they do not need to find properties which are actually *capable* of defining them. The relationship between classes and their properties can be biconditional, or simply conditional. In the first case, properties are thought as necessary and sufficient (“if and only if”). “If x has chirochordian limbs, then x belongs to the taxon *Tetrapoda*,” and “only if x has chirochordian limbs, then x belongs to the taxon *Tetrapoda*”. Or, “if a *C. duranii* is found at place y , then y belongs to Central France,” and “only if a *C. duranii* is found at place y , then y belongs to Central France”. Both of those examples are patently wrong: some members of *Tetrapoda* lack chirochordian limbs; some places in upper Dordogne, Lot and Loire drainages do not host *C. duranii*. But, as we just said, biogeographers do not need to find such properties; since they are looking for *evidences* for the existence of areas, they only need to find properties which are simply *sufficient*. Indeed, the relationship between properties and their classes does not have to be biconditional, it can also simply be conditional (“if... then”). “If specimen x has chirochordian limbs, then specimen x belongs to the taxon *Tetrapoda*”, “if a *C. duranii* is found at place y , then place y belongs to Central France”. Chirochordian limbs nonetheless remain an evidence for the existence of *Tetrapoda* even if some tetrapods are legless; and *C. duranii* remain an evidence for the existence of Central France as a biogeographical area even if some places in upper Dordogne, Lot and Loire drainages do not host any.

CM’s argument then fails to account for those two facts: (1) biogeography’s aim is not to *define* areas, but to *search evidence* for their existence (epistemological fact); (2) therefore it does not need to find *necessary and sufficient* properties to describe its objects (biogeographical areas), but only *sufficient* one (logical fact).

Conclusion

Crother and Murray’s paper is prominent in pointing out a complex conceptual problem: the meaning of biogeographical concepts. However, we think that they their conclusion is premature and that their arguments lack philosophical content. In this short commentary, we have shown that (i) in order to understand scientific concepts, we must resist to the temptation of proposing their “ontology,” but rather try to understand what they means and how they are used in the scientific context under investigation; and that (ii) in the context of historical

biogeography, areas are characterized by intension with geographical distributions of taxa, and therefore understood as classes. This idea was retained in the International Code of Area Nomenclature (Ebach *et al.*, 2008), even though the initial proposition had many inconsistencies (Zaragüeta Bagils *et al.*, 2009).

Biosketches

Sophie Pécaud is a Ph.D. student in philosophy of biology at the University of Paris 1 Panthéon-Sorbonne. Her research focuses on the relationships between phylogenetic systematics and evolutionary biology.

Visotheary Ung is an engineer at CNRS and PhD student at the MNHN Paris. Her research focuses on historical biogeography of Southeast Asia and integrated temporal framework in cladistic biogeography through diachronic events analysis.

Stéphane Prin is a Ph.D. student in phylogenetics and systematics at the MNHN Paris. His research focuses on theoretical, conceptual, and formal aspects of phylogenetics, mainly under its cladistic approach.

René Zaragüeta Bagils is an associate professor at UPMC Univ Paris 06. My research concerns the theory of cladistic phylogenetics and biogeography and its philosophical implications. I am particularly interested in the approach known as "tree-item analysis".

Author contributions: all four authors conceived the ideas and Sophie Pécaud led the writing.

Bibliography

Axelius, B. (1991). Areas of distribution and areas of endemism. *Cladistics*, **7**, 197-199.

Crother, B. I. and Murray, C. M. (2011). Ontology of areas of endemism, *Journal of Biogeography*, **38**, 1009-1015.

Ebach, M. C., Morrone, J. J., Parenti, L. R and Vilorio, Á. L. (2008). International Code of Area Nomenclature. *Journal of Biogeography*, **35**, 1153-1157.

Gayon, J. (2011). De Popper à la biologie de l'évolution : la question de l'essentialisme. *Philonsorbonne*, **6**, 127-136.

Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Zoology*, **23**, 536-544.

Harold, A. and Mooi, R. D. (1994). Areas of endemism: definition and recognition criteria. *Systematic Biology*, **43**, 2, 261-266.

Hull, D. L. (1978). A matter of individuality. *Philosophy of Science*, **45**, 335-360.

Lebbe, J. (1998). Représentations par objets et classifications biologiques. *Langages et modèles à objets. État des recherches et perspectives* (ed. by R. Ducounau, J. Euzenat, G. Masini and A. Napoli), pp. 433-459. INRIA, Rocquencourt.

Popper, K. R. (1945). *The Open Society and its Enemies*, Vol. 1: *The Spell of Plato*. Routledge, London.

Ruse, M. (1995). The species problem. *Concepts, Theories and Rationality in the Biological Science* (ed. by G. Walters and J. G. Lennon), pp. 171-193. Universitätsverlag Konstanz and University of Pittsburgh Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Translated by G. E. M. Anscombe and D. Paul, Blackwell Publishing, Oxford.

Zaragüeta Bagils, R., Bourdon, E., Ung, V., Vignes-Lebbe, R. and Malécot, V. (2009). On the International Code of Nomenclature (ICAN). *Journal of Biogeography*, **36**, 1617-1619.

6.2 « Hennig's revolution has arrived »

Systematics and Biodiversity (en préparation)

6.2.1 Contexte de l'article

Cet article est toujours en préparation cependant nous avons mis en ligne (<http://www.nature.com/nature/report/index.html?comment=58849&doi=10.1038/496295a>) sur le forum de la revue « Nature » une courte correspondance (Encadré 7) sur les conseils de l'éditeur de Nature (Dr Gee).

Nous voulions rebondir sur un article récemment publié (2013) dans Nature par Wheeler, Assis et Rieppel (Wheeler et al. 2013) dans lequel ils soulignent l'importance des données descriptives morpho-anatomiques pour la taxonomie *a contrario* de l'ère actuelle du « tout moléculaire ».

6.2.2 Conclusions

Nous appuyons leur propos en y ajoutant une dimension sémantique pour laquelle nous proposons un cadre analytique, formel qui sert de trame à la représentation des caractères utilisés en Systématique. De plus, nous rappelons que le contexte actuel Européen et International s'oriente vers une standardisation des données y compris pour les données descriptives.

Ce court article s'intègre parfaitement avec notre proposition d'un cadre formel pour la Biogéographie. Le parallèle ayant été déjà démontré entre la Systématique et la Biogéographie, notre modèle se veut clairement trans-disciplinaire, c'est ce que nous étayons dans l'article suivant.

Encadré 7 Hennig's revolution has arrived

In 2008, Wheeler (*Systematic entomology* 33, 2-7; 2008) pointed out that Willi Hennig "re-evaluated taxonomy, as *phylogenetic systematics*, to its rightful place as a rigorous, free-standing and central field of the biological sciences". With Assis and Rieppel, he enabled the message to be "hammered home" claiming that « the way Willi Hennig discovered evolutionary relationships should not be forgotten » (*Nature* 496, 295-296; 2013). They also defended an unusual position in the current context of an « all molecular » world that morphological characters are still significant and essential for evolutionary studies. They are absolutely right: both types of characters should be integrated in systematic studies aimed at reconstructing monophyletic groups, i.e. no type of characters should prevail over another. Wheeler (2008) concluded that a "Cyber-Enabled Taxonomy" would add a new step to Hennig's unfinished revolution.

During the last decade, various initiatives emerged with this aim of providing collaborative and enhanced platforms for this new delocalized e-taxonomy. Among them, EDIT (the European Distributed Institute of Taxonomy, <http://www.e-taxonomy.eu/>) and ViBRANT (Virtual Biodiversity Research and Access Network to Taxonomy, <http://vbrant.eu/>) paved the road and brought new infrastructures to speed taxonomic work, with standardized exchange format for descriptive data (SDD, Standardized Descriptive Data) following TDWG (Taxonomic Database Working Group, <http://www.tdwg.org/>) recommendations, and webservices-oriented platforms offering an easy access and sophisticated tools to assist taxonomists in their daily work.

Here we suggest that, in addition with cybertaxonomy, the focal point of the next revolution will be the semantic representation of characters. Systematics has to be formalized in a semantic modelistic framework allowing explicit conceptualization of characters such as promoted by the "semantic conception" of empirical theories. This philosophical approach to scientific theories, developed in the 60's, claims that a theory can be formalized as a model, a set of objects and relationships satisfying a set of axioms, i.e. constraints. The model allows the interpretation of these axioms mathematically. For instance, the evolutionary tree may be considered a model of nested classes accounting for relationships among monophyletic taxa. Taxa are formal objects related by hierarchical relationships and constrained by the monophyly condition. The model-theoretic view is fruitful, especially in bridging taxonomic characters with the Semantic Web. The development of the Resource Descriptive Framework (RDF) and Web Ontology Language (OWL), i.e. the standards of the Semantic Web for the development of ontologies, relies on theoretical models and meets the taxonomic character representation. A "Semantic Web"-based taxonomy is the outlook for the future.

Wheeler, Q.D., editor 2008. The new taxonomy. CRC Press Inc. New York, USA.

Wheeler Q.D. 2008. Undisciplined thinking: morphology and Hennig's unfinished revolution. *Systematic entomology* 33, 2-7.

Both authors have equally participated in the writing of this comment.

CR2P UMR 7207 CNRS/MNHN/UPMC, 57 rue Cuvier CP 48 75005 Paris

6.3 « To add a little semantics can help for conservation policies »

6.3.1 Contexte de l'article

Faisant suite à l'article que nous avons écrit concernant une « ontologie des aires d'endémisme », il nous est apparu opportun de poursuivre la réflexion sur les objets (*id est* les OTUs) de la Biogéographie que sont les aires d'endémisme. Le présent article a été présenté sous une forme d'une présentation orale au congrès de systématique de Vienne « Biosystematics 2013 ».

6.3.2 Résumé

Les priorités concernant la conservation biologique restent d'un intérêt majeur et une question cruciale dans un contexte de perte de biodiversité mondiale. La sélection et la délimitation des aires protégées ou à protéger sont toujours faites empiriquement, ce qui est un processus difficilement réalisable sans modèle. Jusqu'alors, aucun modèle mathématique n'a été suggéré pour définir de manière non-ambigüe une aire protégée. Il est plus que jamais nécessaire d'ajouter un peu de sémantique dans la façon de considérer ces dernières. Nous proposons une ontologie pour les aires biogéographiques afin de disposer de règles pour la conservation basées sur un modèle sémantique.

6.3.3 Conclusions

Dans un contexte mondial où la gestion de la biodiversité est un problème majeur, définir un modèle conceptuel pluridisciplinaire de représentation des connaissances peut s'avérer être un atout supplémentaire. Nous proposons un modèle ancré dans le paradigme de la Conception Sémantique qui s'applique à la Systématique mais également à la Biogéographie comparative et qui offre un outil pour la détermination des aires protégées. Cet article est toujours en préparation.

To add a little semantics can help for conservation policies

Ung V¹. and Grand A¹.

Both authors have equally participated in the writing of this article.

¹Centre de Recherche en Paléobiodiversité et Paléoenvironnements

UMR 7207: CNRS MNHN UPMC

57 rue Cuvier CP 48

75005 Paris

France

ABSTRACT

Conservation priorities are of the major interest and remain a crucial issue in a context of global biodiversity loss. The selection and delimitation of protected areas are human-dependant and empirical. Yet, no mathematical model has been suggested to unambiguously define a protected area. It is more than ever necessary to add semantics in the way areas are viewed. Here we would like to propose an ontology for areas of endemism implying a modelistic framework for conservation policies.

KEY-WORDS

Biodiversity hotspots, Conservation, Endemism, General pattern, Mathematics, Ontology, Semantics.

Endemism is the ecological state of being unique to a defined geographic location. This notion is crucial nowadays, in the current context of global biodiversity loss (Myers et al. 2000; Sachs et al. 2009; Butchard et al. 2010) and climate change (IPCC 2007; Dawson et al. 2011). While some researchers claim that we are facing a major biodiversity crisis, the main question for policy-makers remains: how to delimit protected areas? Biodiversity hotspots are recognized all over the globe (Myers et al. 2000; Seligmann et al. 2007; Woodruff 2010; Bellwood et al. 2012). Orme et al. in 2005 stressed the fact that “global hotspots of species richness (Figure 1) are not necessarily congruent with endemism or threat” (Orme et al. 2005). Thus, shouldn't it be worthwhile to find out a rational and standard way of delimiting endangered areas? Even the main international organizations have their own definition of a protected area. The IUCN definition is “*an area of land and/or sea especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means*”. Whereas for the Convention on Biological Diversity, a protected area is “*a geographically defined area which is designated or regulated and managed to achieve specific conservation objectives*”. *In fine*, despite various attempts (e.g. Kerr 1997; Dudley and Stolton 2008, Stolton and Dudley 2009, Gillepsie 2009 among others), a real standardization is missing.

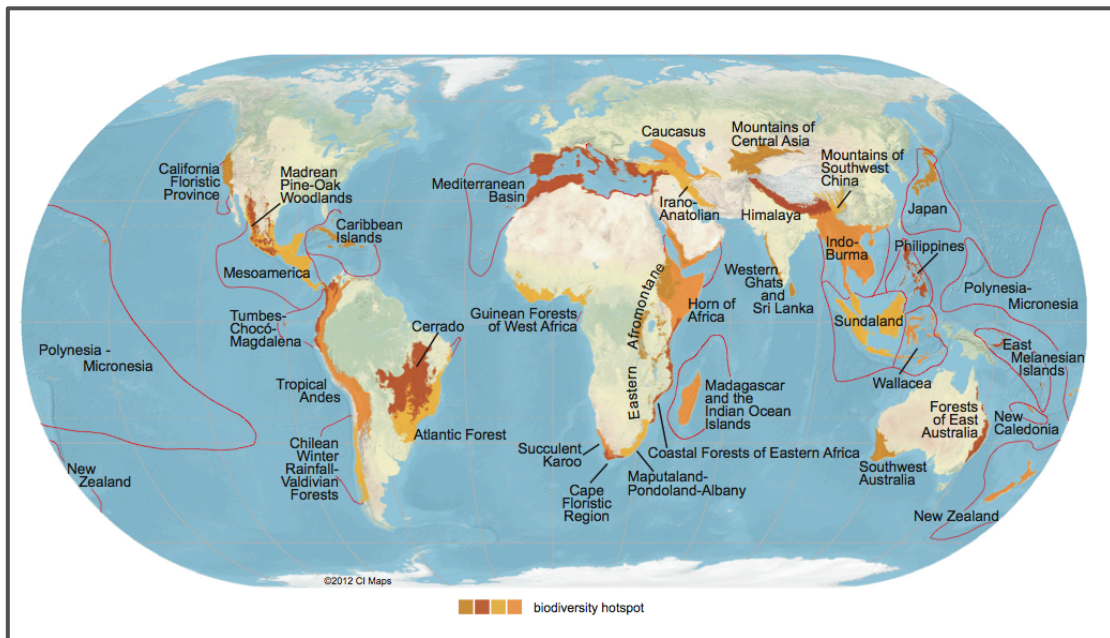


Figure 1 Conservation « hotspots » edited by the Conservation International organization (2011)

In 1994, Anderson already stated that:

«There are three major difficulties encountered by those dealing with the phenomenon of endemism: a semantic problem, the absence of a clear conceptual framework, and an analytical problem.» **Sydney Anderson, 1994: 451.**

We propose a conceptual framework that aims to help solving the analytical problem (i.e. defining areas that should be targets for conservation efforts) and give an ontology for areas of endemism that will address the semantic question.

Our work has its foundations in comparative biogeography that aims to discover and classify biogeographical patterns thanks to a congruence analysis. We suggest that comparative biogeography could be usefully formalized in a semantic modelistic framework allowing explicit conceptualization of biogeographic knowledge such as promoted by a semantic conception of empirical theories. Semantic interpretation of scientific theories developed in the 1960's in philosophy of science. Patrick Suppes, Bas van Fraassen, Frederic Suppe and Ronald Giere, among others, after Tarski's model-theoretic approach, were among the major promoters of the approach. Under a semantic framework, a scientific theory is formalized as a model. A model is a set of objects and relationships satisfying a set of axioms, i.e. constraints. The model allows the interpretation of these axioms mathematically. In biology, Elisabeth Lloyd, John Beatty, and Paul Thompson presented semantic models of the population genetics and of the Mendelian genetics (Beatty 1987; Lloyd 1986, 1987, 1988; Thompson 1983, 1986,

1987). The fruitfulness of their approach has recently been taken up by Gildenhuis (2013). Thompson (1986) viewed the evolutionary theory as a family of interacting theories. He stated that “the semantic conception of theories provides a framework within which a formalization of evolutionary theory understood as a family of interacting theories can be given [...]”. Among these evolutionary theories, we focus on phylogenetics and historical biogeography. While Thompson showed that the theory of heredity interacts with the theory of natural selection, we emphasize that comparative biogeography interacts with systematic phylogenetics. Moreover, historical biogeography depends, i.e. inherits the structure of its model from phylogenetics. Thus, a semantic model accounting for comparative biogeography is actually domain-independent. As a consequence, such a model also accounts for phylogenetics.

Here, we highlight that traditional phylogenetic trees and areagrams, which are widely used in phylogenetics and biogeography respectively, can be described under the semantic framework. Within systematic phylogenetics, the evolutionary tree is a model interpretable as nested groups accounting for relationships among monophyletic taxa. Taxa are related by hierarchical relationships and constrained by the monophyly condition. The objects under study are Operational Taxonomic Units (OTUs), i.e. terminal taxa, that have to be defined prior to the modelling. Taxa are represented by formal objects and the monophyly constraint is represented by a particular formal relationship of partial order. Within comparative biogeography, the objects studied are areas of endemism deduced from taxa distributions. They have to be defined prior to the semantic modelling. By way of example, figure 2 illustrates an empirical case study based on the distributions of taxa through Southeast Asia and the West Pacific region (Turner et al. 2001) where the authors determined 23 areas of endemism to perform a biogeographic analysis of the region. The semantic model we describe here is a nested model for which an OTU or an area of endemism is only a particular case (i.e. a terminal) of a taxon or of a biogeographic area respectively. As an example, figure 2 illustrates an empirical case study based on distributions of taxa through Southeast Asia and the West Pacific region (Turner et al. 2001) where the authors determined 22 areas of endemism to perform a biogeographical analysis of the region.

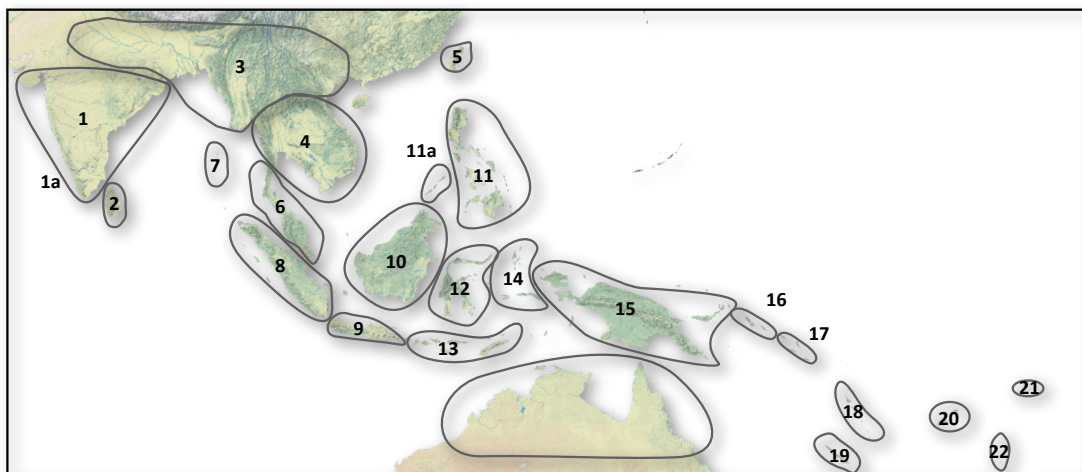


Figure 2 The areas of endemism employed. (1) India, (1a) Kerala, (2) Sri Lanka, (3) Himalayas, (4) Southeast Asia, (5) Taiwan, (6) Malay Peninsula, (7) Andaman and Nicobar Islands, (8) Sumatra, (9) Java, (10) Borneo, (11) Philippines, (11a) Palawan, (12) Sulawesi, (13) Lesser Sunda Islands, (14) Moluccas, (15) New Guinea, (16) Solomon Islands, (17) Santa Cruz Group, (18) Vanuatu, (19) New Caledonia, (20) Fiji, (21) Samoa, (22) Tonga, (23) Australia. (modified after Turner et al. 2001)

In this semantic context, our contribution goes further and consists in a suggestion about the ontology of species and areas of endemism; this issue is essential to grasp the meaning of the term “biodiversity” (taxic and biogeographic) more accurately. Therefore, this ontological consideration impacts the way the methods of discovery of biodiversity and the conservation of biodiversity are implemented.

In philosophy, ontology is the study of that which is, i.e. the existence of objects and concepts outside of our minds. In the field of artificial intelligence, ontology can be defined as “a logical theory which gives an explicit, partial account of a conceptualization”, i.e. “an intensional semantic structure which encodes the explicit rules constraining the structure of a piece of reality” (Guarino and Giaretta 1995). This definition is indeed rooted in a semantic framework that makes explicit the object and its relationships (i.e. the structure) and the constraints that a model accounts for. Since the philosophical interpretation in logico-mathematic terms perfectly meets the requirements of Computer Science developments, it is not surprising that a semantic model is an ontological model. Semantics allows the integration of the term ontology across philosophy and artificial intelligence.

Recently, Crother and Murray (2011, 2013) proposed their own ontology (in its raw philosophical meaning) for areas of endemism that they called the “areas of endemism as individuals theory” based on the “species as individuals theory” *sensu* Ghiselin (1974) and Hull (1978). They stated that:

“The critical first step for the argument of areas of endemism as individuals is the recognition of species as individuals.” Crother and Murray 2011:1010.

They support 5 conditions to establish their ontology: (1) an area must be a particular thing; (2) there cannot be any instance of it; (3) an individual is defined ostensively and not intensionally; (4) individuals have boundaries in time and space, and (5) individuals experience cohesiveness in their response to change.

However, it can be shown that the individual thesis lacks the monophyly constraint. As an individual is limited in time (it is born and then it dies), it cannot contain all its descendants. For instance, we won’t say that a great grandfather is still alive (while he died years ago) because we spend time with his great grandchildren. Thus, areas viewed as individuals do not contain all their descendants (e.g. areas that were born after a vicariance event) and are not monophyletic. As a consequence, a theory that accounts for objects viewed as individuals cannot be assimilated in a model. This theory is instead formalized as an unrestricted structure. We won’t discuss the 5 points proposed by Crother and Murray (2011, 2013) that describe such a structure since here we

focus on models. Frigg (2002) demonstrated that structures are not enough to draw fruitful conclusions. It is not relevant to refer to the ontology outside of a semantic framework since the ontology cannot be grasped by a structure (but only by a semantic model). For this reason, we propose that Crother and Murray's ontology, as opposed to its name, is not an accurate ontology of areas of endemism. We conclude that the monophyly of biogeographic areas is an ontological constraint.

As a consequence, our contribution firstly impacts the way we refer to "biodiversity". Biodiversity is usually measured as a count of species. The term "species" is highly ambiguous. We have emphasized that the ontology of species cannot be grasped in the framework of the "species as individuals" theory. We also consider that the "species as individuals" theory is not accurately implemented by phylogenetic trees, i.e. the objects of the "species as individuals" theory cannot be represented as members of phylogenetic trees. Instead, phylogenetic trees are semantic models that account for the ontology of monophyletic taxa and OTUs. This view is compatible with the "phylogenetic species" theory in which species are regarded as phylogenetic units (Cracraft 1987; Velasco 2008). We join all the authors that insisted for the integration of the phylogenetic dimension in the various measures of biodiversity (e.g. Faith 2002; Carvalho et al 2013). This conclusion is also true for the measures of the biogeographic diversity. Counting areas of endemism is not enough; we have to consider their biogeographic history too through an areagram accounting for the ontology of biogeographic areas in a general sense.

Our contribution also impacts the way the methods that grasp the phylogenetic and biogeographic patterns are implemented. Various cladistic tools aim at reconstructing phylogenetic trees and areagrams (e.g. PAUP*, Swofford 2001; TNT, Goloboff et al 2008; DIVA, Ronquist 1996). However, they are not faithful to the semantic hierarchical model over the course of an entire analysis. For instance, they represent phylogenetic characters with another classificatory structure than a hierarchy, namely a partition, i.e. characters are represented as matrix columns. In biogeography, taxon-area cladograms, that stand for biogeographic characters, are also represented as matrices (e.g. with the Brooks Parsimony Analysis, Wiley 1987, Brooks et al 2001). Faithfulness to a hierarchical model throughout the steps that relate the knowledge level (i.e. the conceptual level) to the empirical level (i.e. observations, measures) is important for internal consistency. Philosophers that formalized the semantic conception of theories stressed the importance of implementing isomorphic models (e.g. Van Fraassen 1994; Frigg 2002; French 2003). We claim that the hierarchical model should be implemented at the character level, and since this model accounts for the interdisciplinarity of theories, the implementation should be the same for phylogenetics and biogeography. Authors have already stressed the parallel between systematics and biogeography (e.g. Nelson and Platnick 1981; Nelson and Ladiges 1991a, 1991b, 1992). Thus, our team has implemented the model at every level of an analysis in the software LisBeth (Zaragüeta-Bagils et al. 2012), aimed at systematics and biogeographic analysis. The objects of the model are implemented as classes. The hierarchical relationship among classes leads to the creation of subclasses. OTUs and areas of endemism are

instances of the classes. Empirical biogeographic analysis using LisBeth have led to the reconstruction of general area cladograms (Cao et al. 2009; Escalante et al. 2007; Parenti and Ebach 2009) and systematic examples are also available (Bourdon 2006; Cao 2008; Corvez 2012; Grand et al. 2013). Figure 3 shows how the general pattern can be retrieved thanks to LisBeth and the delimitation of subsequent biogeographical subregions by the authors (Parenti and Ebach 2009).

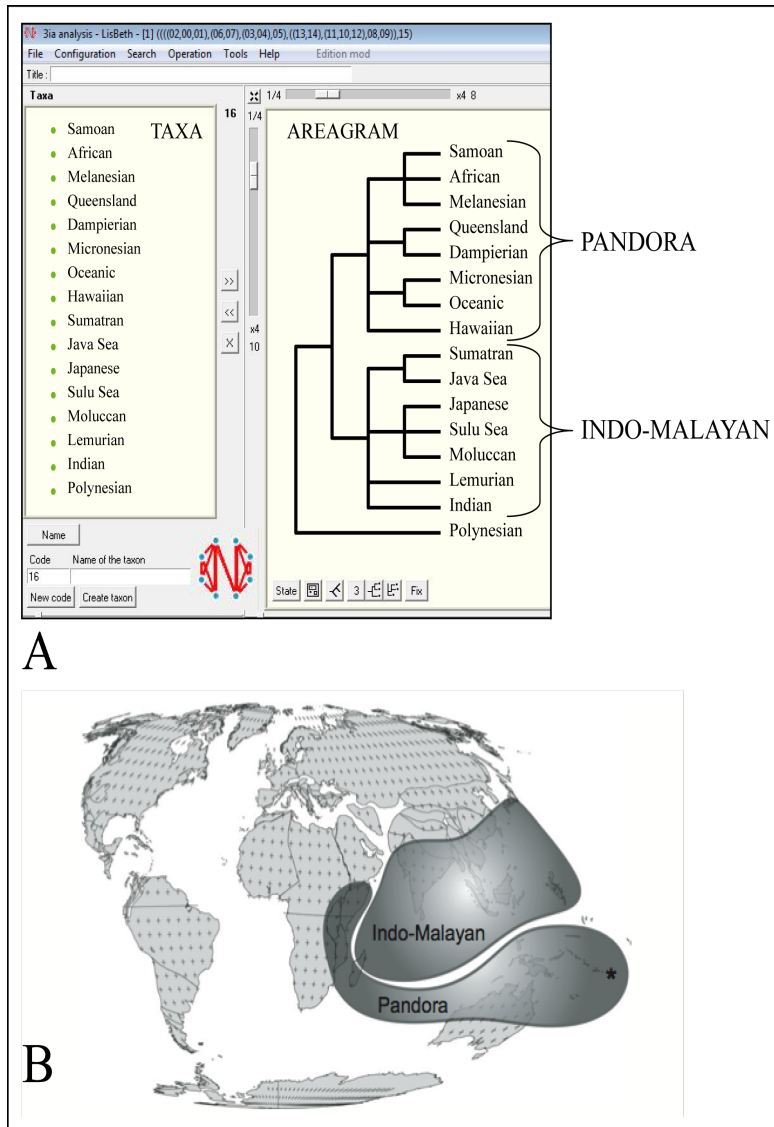


Figure 3 (A) Areagram generated with LisBeth based on the analysis of 16 areas of endemism, (B) the subregions defined

The model of embedded classes gives an ontology for areas of endemism and as such can be useful in the process of delimiting protected areas, i.e. areas of conservation. The associate methodology accounts for a process-free analysis and gives new clues coming from a mathematical representation of knowledge and a calculation procedure. Protected areas would be the ones with the highest number of species retrieved analytically and not only enumerated.

Nelson and Platnick wrote (N&P 1981): "*knowledge of general characters is useful for another reason: it is the basis for expectations about order in nature.*" They gave the example of the axial skeleton character in lampreys, sharks and perches. Considering this character, sharks and perches are grouped based on the presence of a vertebral column. They concluded that this pattern "will prove true also for all other general characters that might subsequently be discovered". Thus, we would expect that a general pattern of areas would prove true not only for the distributions of taxa used to reconstruct the general pattern but also for other taxa that might subsequently be discovered. That means that delimitating protected areas on the basis of particular taxa distributions could contribute to protect other unexpected taxa. Thus, due to its phylogenetic dimension, the conservation of taxic and biogeographic diversity has a serendipitous element in it.

The phylogenetic and semantic dimension of conservation provides information on "what" we want to protect; but the question of "how" we protect remains. While "what" we want to protect are taxa and biogeographic areas, the "how" side of the question implies nomenclatural norms and identification tools. As Agapow et al. (2004) stated, "the species problem is emblematic of conservation and the gulfs between science policy and practice". We agree that a conceptual contribution should not be promoted to the detriment of pragmatic actions and conservation decisions. With this requirement in mind, we emphasize that (taxic and biogeographic) conservation is embedded in a taxonomic framework and as highlighted by Wheeler (2007), characters are central within taxonomy. The taxonomic characters allow for formalizing the knowledge we have about taxa or areas and can be used to produce descriptions, identification keys as well as phylogenies. Thus, the phylogenetic ontology ("what") and the pragmatic identification ("how") are integrated by the means of characters. We showed that it is worth implementing the character level as a semantic hierarchical model. We can now extend this claim to the descriptive area of biology: the descriptive modelling should account for hierarchical relationships among characters. That is, the hierarchy among characters allows representing inapplicability conditions. Our team developed the software Xper² (Ung et al. 2010) that is interoperable with LisBeth (Zaragüeta-Bagils et al. 2012). The descriptive knowledge management tool Xper² is designed for editing descriptions and creating interactive identification keys. We consider that conservation actions could gain from using such interoperable tools.

In a Philosophical framework, ontology is the study of that which is (i.e. the existence of things outside of our minds). With an Artificial Intelligence meaning, ontology can be defined as "*a logical theory which gives an explicit, partial account of a conceptualization*", i.e. "*an intensional semantic structure which encodes the explicit rules constraining the structure of a piece of reality*" (Guarino and Giarretta 1995). This definition is rooted in a semantic framework that makes explicit the object and relationships (i.e. the structure) and the constraints that

a model accounts for. Since the philosophical interpretation in logico-mathematic terms perfectly meets the requirements of Computer Science developments, it is not surprising that a semantic model is an ontological model. Semantics allows the integration of the term ontology across Philosophy and Artificial Intelligence.

We also assess that, since we face a deluge of descriptive data, the way of designing ontological models (i.e. engineering ontologies of artificial intelligence) would be of great help to mine the data. We are currently working at bridging the models implemented in LisBeth and Xper² with engineering ontologies (e.g. we refer to the Comparative Data Analysis Ontology, Prosdocimi et al 2009, and the OBO Relation Ontology, Smith et al 2005).

Biology (as a whole) is evolving towards being fully integrative and should benefit from mathematical (i.e. theoretical) models (Murray 2000). In that sense, it has paved the road for new conservation policies.

References

- Agapow, P-M., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C. and Purvis, A. (2004) The Impact of Species Concept on Biodiversity Studies. *The Quarterly Review of Biology*, 79 (2), 161-179.
- Andersen, Sydney. Area and Endemism, *The Quarterly Review of Biology*, December 1994, Vol. 69, No. 4.
- Beatty, J. (1987) On Behalf of the Semantic View. *Biology and Philosophy*, 2, 17-23.
- Bellwood, D.R., Renema, W. and Rosen B.R. (2012) Biodiversity hotspots, evolution and coral reef biogeography: a review. Chapter 9, 216-245. In: Gower, DJ, Johnson, KG, Richardson, JE, Rosen, BR, Rüber, L & Williams, ST. (eds) 2012 *Biotic evolution and environmental change in Southeast Asia*. Cambridge University Press, U.K.
- Bourdon, E. (2006) *L'avifaune du Paléogène des phosphates du Maroc et du Togo. Diversité, systématique et apports à la connaissance de la diversification des oiseaux modernes (Neornithes)*. D Phil. Thesis, Muséum National d'Histoire Naturelle, Paris France.
- Brooks, D. R., van Veller, M. G. P. and McLennan, D. A. (2001) How to do BPA, really. *Journal of Biogeography*, 28, 345-358.
- Butchart S.T.H.M., Walpole M., Collen B., van Strien A., Scharlemann J.P.W., Almond R.E.A., Baillie J.E.M., Bomhard B., Brown C., Bruno J., Carpenter K.E., Carr G.E., Chanson J., Chenery A.M., Csirke J., Davidson N.C., Dentener F., Foster M., Galli A., Galloway J.N., Genovesi P., Gregory R.D., Hockings M., Kapos V., Lamarque J.F., Leverington F., Loh J., McGeoch M.A., McRae L., Minasyan A., Hernández Morcillo M., Oldfield T.E.E., Pauly D., Quader S., Revenga C., Sauer J.R., Skolnik B., Spear D., Stanwell-Smith D., Stuart S.M., Symes A.,

- Tierney M., Tyrrell T.D., Vié J.C., Watson R. (2010) Global biodiversity: indicators of recent declines. *Science* 328, 1164-1168.
- Cao N. (2008) Analyse à trois éléments et anatomie du bois des Fagales Engl. Thèse de doctorat (MNHN, Paris, France).
- Cao, .N., Bourdon, E., El Azawi, M., Zaragüeta-Bagils, R. (2009) Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example. *Bulletin de la Société Géologique de France*, 180, 13-15.
- Carvalho, M. R., Ebach, M. C., Williams, D. M., Nihei, S. S., Trefaut Rodrigues M., Grant, T., Silveira L. F., Zaher, H., Gill, A. C., Schelly, R. C., Sparks, J. S., Bockmann, F. A., Séret, B., Ho, H-C., Grande, L., Rieppel, O., Dubois, A., Ohler, A., Faivovich, J., Assis, L. C. S., Wheeler, Q. D., Goldstein, P. Z., de Almeida, E. A. B., Valdecasas, A. G. and Nelson, G. (2013) Does counting species count as taxonomy ? On misrepresenting systematics, yet again. *Cladistics*, 1-8.
- Corvez, A. (2012) L'origine de la Mégaphylle chez les Monilophytes. D Phil Thesis, Muséum National d'Histoire Naturelle, Paris France.
- Cracraft, J. (1987) Species Concepts and the Ontology of Evolution. *Biology and Philosophy*, 2, 329-346.
- Dawson T.P., Jackson S. T., House J. I., Prentice I. C. and Mace G. M. (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, 332, 53-58.
- Dudley N. and Stolton S. (eds) (2008). *Defining protected areas: an international conference in Almeria, Spain*. Gland, Switzerland: IUCN. 220 pp
- 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.
- Faith, D. P. (2002) Quantifying Biodiversity: a Phylogenetic Perspective. *Conservation Biology*, 16 (1), 249-252.
- French, S. R. D. (2003) A Model-Theoretic Account of Representation (Or, I Don't Know Much About Art...But I Know It Involves Isomorphism). *Philosophy of Science*, 70 (5), 1472-1483.
- Frigg, R. (2002) Models and representation: Why structures are not enough. *Measurement in physics and economics discussion paper series DP MEAS 25/02*, London School of Economics, London.
- Gildenhuis, P. (2013) Classical population genetics and the semantic approach to scientific theories. *Synthese*, 190, 273-291.
- Gillespie A. (2009) Defining Internationally Protected areas. *Journal of International Wildlife Law and Policy*, 12, 229-247.
- Goloboff, P. A., Farris, J. S. and Nixon, K. C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, 24(5), 774-786.
- Grand, A., Corvez, A., Duque Velez, L. M. and Laurin, M. (2013) Phylogenetic inference using discrete characters: performance of ordered and unordered parsimony and of three-item statements. *Biological Journal of the Linnean Society*, (in press).

- Guarino, N. and Giarretta, P. (1995) Ontologies and knowledge bases: towards terminological clarifications. 25-32. In: Mars, N. (ed) 1995 Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. IOS Press, Amsterdam.
- Ghiselin, M. T. (1974). A radical solution to the species problem. *Systematic Zoology*, 23, 536-544.
- Hull, D. L. (1978). A matter of individuality. *Philosophy of Science*, 45, 335-360.
- Kerr J. T. (1997) Species richness, endemism, and the choice of areas for conservation. *Conservation Biology*, 11 (5), 1094-1100.
- Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change 2007 : Synthesis report*. Geneva : IPCC.
- Lloyd, E. A. (1986) Thinking about models in evolutionary theory. *Philosophica*, 37, 87-100.
- Lloyd, E. A. (1987) Confirmation of Ecological and Evolutionary Models. *Biology and Philosophy*, 2, 277-293.
- Lloyd, E. A. (1988) *The structure and confirmation of evolutionary theory*. Princeton University Press, New Jersey, UK.
- Murray, J. D. (2000) Pattern formation in integrative biology – a marriage of theory and experiment. *Comptes Rendus de l'Académie des Sciences de Paris, Life Sciences*, 323, 5-14.
- Myers N., Mittermeier R. A., Mittermeier C. G., Da Fonseca G. A. B. and Kent J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.
- Nelson, G. J. and Platnick, N. L. (1981) *Systematics and Biogeography: Cladistics and Vicariance*, Columbia Univ Press, New York.
- Nelson, G. J. and Ladiges, P. Y. (1991a) Standard assumptions for biogeographic analysis, *Australian Systematic Botany*, 4, 41-58.
- Nelson, G. J. and Ladiges, P. Y. (1991b) Three-area statements: Standard assumptions for biogeographic analysis, *Systematic Zoology*, 40, 470-485.
- Nelson, G. J. and Ladiges, P. Y. (1992) Information Content and Fractional Weight of Three-Item Statements, *Systematic Biology*, 41, 490-494.
- Orme et al (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature*, 436, 1016-1019.
- Orbach M, Karrer L. 2010. *Marine Managed Areas: What, Why, and Where*. Science and Knowledge Division, Conservation International, Arlington, Virginia, USA.
- Parenti, L. R., Ebach, M. C. (2009) *Comparative Biogeography: Discovering and Classifying Biogeographical Patterns of a Dynamic Earth*, University of California Press, Berkeley.
- Prosdocimi, F., Chisham ,B., Pontelli, E., Thompson, J. D. and Stoltzfus, A. (2009) Initial Implementation of a Comparative Data Analysis Ontology. *Evolutionary Bioinformatics*, 5, 47-66.
- Ronquist, F. (1996) DIVA, version 1.0. Computer program for MacOS and Win32. Available from www.systbot.uu.se/personel/f_ronquist.html.

- Sachs J.D., Baillie J. E. M., Sutherland W. M., Armsworth P. R., Ash N., Beddington J., Blackburn T. M., Collen B., Gardiner B., Gaston K.J., Godfray H. C. J., Green R. E., Harvey P. H., House B., Knapp S., Kümpel N. F., Macdonald D. W., Mace G. M., Mallet J., Matthews A., May R. M., Petchey O., Purvis A., Roe D., Safi K., Turner K,
- Walpole M., Watson R. and Jones K. E. (2009) Biodiversity conservation and the millennium development goals. *Science*, 325, 1502-1503.
- Seligmann P., Mittermeir R. A., Da Fonseca A. B., Gascon C., Crone N., Da Silva J. M. C., Famolare L., Bensted-Smith R., Rajaobelina L., and Beehler B. (2007) Centers for Biodiversity Conservation. CBC Ressources. Conservation international.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. and Rosse, C. (2005) Relations in biomedical ontologies. *Genome Biology* 6, R46.
- Stolton S. and Dudley N. (eds) (2010). Arguments for protected areas : Multiple benefits for conservation and use. Earthscan Ltd, Dunstan House, 14a St Cross Street, London EC1N 8XA, UK.
- Swofford, D. L. (2001) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.
- Thompson, P. (1983) The Structure of Evolutionary Theory: A Semantic Approach. *Studies in History and Philosophy of Science*, 14, 215-229.
- Thompson, P. (1986) The interaction of theories and the semantic conception of evolutionary theory. *Philosophica*, 37, 73-86.
- Thompson, P. (1987) A Defense of the Semantic Conception of Evolutionary Theory. *Biology and Philosophy*, 2, 26-32.
- Turner H., Hovenkamp P. and van Welzen P. C. (2001) Biogeography of Southeast Asia and the West pacific. *Journal of Biogeography*, 28, 217-230.
- Ung, V., Dubus, G., Zaragüeta-Bagils, R. and Vignes-Lebbe, R. (2010) Xper2: introducing e-taxonomy. *Bioinformatics*, 26, 703-704.
- Van Fraassen, B. (1994) Interpretation of Science: Science as Interpretation. 169-187. In: Hilgevoord (ed) 1994 *Physics and Our View of the World*. Cambridge University Press, Cambridge.
- Velasco, J. D. (2008) Species concepts should not conflict with evolutionary history, but often do. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 39, 407-414.
- Whiley E.O. (1987) Methods in vicariance biogeography. In *Systematics and Evolution: a matter of diversity*, P. Hovenkamp (ed.). Utrecht: Institute of Systematic Botany, Utrecht University, pp.283-306.
- Wheeler, Q. D. (2007) Invertebrate systematics or spineless taxonomy? *Zootaxa*, 1668, 11-18.
- Woodruff, David S. (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and future of the remaining refugial-phase biodiversity. *Biodivers Conserv*, 19, 919-941.

Zaragüeta Bagils, R., Ung, V., Grand, A., Vignes-Lebbe, R., Cao, N. and Ducasse, J. (2012) LisBeth: New cladistics for phylogenetics and biogeography. *Comptes Rendus Palevol*, 11, 563-566.

Conclusions et perspectives

1 Conclusions

Il est primordial que le lecteur perçoive ici toute l'importance du contexte théorique et méthodologique dans lequel nous évoluons depuis une soixantaine de pages. Il était indispensable pour la bonne compréhension de la suite d'expliquer ce cadre formel et conceptuel.

Ainsi, nous avons explicité (succinctement) les principes de la méthode d'analyse à 3 éléments, mais nous avons aussi abordé quelles sont les particularités inhérentes à la Biogéographie (MASTs, taxons à large répartition) et quelles sont les solutions implémentées pour les solutionner. Nous avons présenté les différentes étapes nécessaires pour réaliser une analyse biogéographique avec LisBeth.

Enfin, nous nous sommes penchés sur la problématique des unités d'analyse de la Biogéographie historique que sont les aires d'endémismes et avons initié une réflexion autour des modèles sémantiques que sont les ontologies. Cette dernière approche, certes, encore sous la forme d'un « canevas » présente l'avantage de poser une trame formelle pour définir ce que sont les aires d'endémisme et plus généralement les aires biogéographiques.

2 Perspectives

Les dernières propositions de modifications ergonomiques étant déjà en phase d'implémentation dans LisBeth, nous nous focaliserons dans un futur proche sur des améliorations d'ordre algorithmique, notamment sur la recherche d'un algorithme de reconstruction heuristique qui pour le moment n'est pas implémenté dans LisBeth.

Le cadre méthodologique étant ainsi posé, nous pouvons désormais aller plus avant dans notre analyse empirique : de l'Asie du Sud-Est (chapitre II) et des réflexions d'ordre méthodologique concernant l'intégration de la dimension temporelle qui nous permettra d'approcher d'un véritable modèle de Biogéographie intégrative (chapitre III).

A ce moment précis de notre discours, nous informons notre lecteur qu'il peut indépendamment lire les chapitres II et III qui suivent et dans n'importe quel ordre. Rappelons aussi que ce premier chapitre reste volontairement assez succinct pour ne pas alourdir le propos global de cette thèse.

Chapitre II: Nouvelles inférences cladistiques sur la Biogéographie historique du Sud-Est asiatique et de la région de l'Ouest du Pacifique

«The phenomena of distribution in the Malay Archipelago, to which I have here called attention, teach us that, however narrow may be the strait separating an island from its continent, it is still an impassable barrier against the passage of any considerable number and variety of land animals ; and that in all cases in which such islands possess a tolerably rich and varied fauna of species mostly identical, or closely allied with those of the adjacent country, we are forced to the conclusion that a geologically recent disruption has taken place. »

Alfred Russel Wallace, *On the zoological geography of the Malay Archipelago*, 1859.

Introduction

L'histoire géologique complexe du Sud Est asiatique et de la région de l'Ouest du Pacifique en fait une région de choix pour les biogéographes.

Nous nous sommes intéressés à cette région située au cœur de l'archipel Indo-Australien en raison de sa complexité tant biologique que géologique. Elle se décompose en deux régions géographiques : l'une, continentale, est constituée par la Birmanie (Myanmar), le Cambodge, le Laos, la Thaïlande, le Vietnam et la Malaisie (ou pour être plus précis, la péninsule malaise) et l'autre, maritime, se compose de Brunei, de l'Est du Timor, de l'Indonésie, des Philippines et de Singapour (Michaux, 2010)¹. Nous avons étendu la région d'étude plus à l'Ouest pour y intégrer l'Inde et à l'Est, tout le Pacifique Ouest (partant de la Nouvelle-Guinée jusqu'aux îles du Pacifique) (Figure 12). Placée à l'intersection de deux plaques géologiques (asiatique et australienne), cette région possède une activité sismique et volcanique forte, traduite par une histoire géologique des plus intéressantes. Les différentes îles constituant les ensembles maritimes ont émergés à des âges distincts (Hall 2002, 2011, 2012) ce qui fait apparaître toute la problématique de la Biogéographie, à savoir la distribution des organismes vivants (plantes et animaux) dans cette région.

Voici un court extrait d'un article de Robert Hall de 2012 dans lequel il propose une reconstruction géologique de la région, plus avancée par rapport aux précédentes, et dit :

« The plate tectonic development of this region provides the most important framework for interpreting its biogeography, but the explanatory power of this hypothesis is sometimes exaggerated. Plate tectonics has been the engine, but the consequence has been more than merely movements of pieces across the Earth's surface. Plates, and possibly microplates, have carried their own biotas, but the land mass of Australia has crosses climatic zones, mountains have risen and disappeared, deep waterways have closed and opened, and islands have provided connections. We are still far from a complete understanding of the geology of this region, still less the links between geology, paleogeography, ocean-atmosphere circulation and climate which have influenced biological change, biogeographical patterns and biodiversity. It seems more and more likely that the geological history of Wallacea is important not only for understanding the impact on life in this equatorial region, but also for the planet because of the probable consequences for global climate. » **Hall 2012, p 33-34.**

En une dizaine de lignes, Hall a condensé et saisi toute la portée de la Biogéographie historique. Cette région du monde est considérée depuis longtemps comme un centre de biodiversité tant pour les taxons terrestres que marins (Michaux 2010, Woodruff 2010). Notamment, sa richesse floristique la rend particulièrement intéressante. En 2001, déjà, Frodin et collaborateurs le remarquent et énoncent que « le Sud-est asiatique est un

¹ Nous aimerions remarquer que même si nous délimitons la région d'étude avec des dénominations de frontières d'ordre politique, nous sommes conscients qu'il aurait fallu rechercher les frontières naturelles correspondant à ces même limites territoriales. A notre corps défendant, nous reprenons aussi la terminologie existante pour la région publiée par d'autres auteurs.

remarquable centre de diversité florale. La région malaise qui s'étend de la péninsule Malaise à la Nouvelle-Guinée, contient à elle-seule une estimation de 42000 espèces de plantes vasculaires ».

La complexité des mouvements géotectoniques fait de cette région un formidable laboratoire pour notre propos, à savoir l'exploration des relations de parenté entre aires d'endémisme, depuis que nous savons que les changements qui ont eu lieu à des positions relatives des nombreuses îles (et de certaines parts) suggèrent un grand nombre d'évènements de spéciation par vicariance. Compte-tenu de cette richesse biologique, il est primordial d'analyser et de comprendre les processus à l'origine d'une telle diversité, non seulement pour des besoins liés à la Systématique (recensement, identification, classification, monitoring etc.) mais aussi pour la connaissance de l'Évolution en général et en particulier de la Biogéographie (le lecteur est prié de se référer au chapitre I pour des explications complémentaires concernant la discipline elle-même et ses méthodes). Nous proposons une nouvelle approche, cladistique, pour l'étude de cette région située autour de la célèbre "Ligne de Wallace", en nous plaçant dans le paradigme de la vicariance où la spéciation apparaît suite à la mise en place d'une barrière géographique (par opposition au modèle dispersaliste à partir d'un centre d'origine). Les résultats attendus sont une meilleure connaissance de l'histoire évolutive de la région ainsi que des traits d'histoire de vie des taxons qui la peuplent. Néanmoins, ce n'est pas pour autant que la possibilité de dispersion est niée. Simplement, dans le contexte cladistique, la cause la plus probable de congruence entre les distributions de taxons non apparentés est la vicariance ; les dispersions aléatoires pouvant expliquer toute distribution, en cas de congruence elle est peu vraisemblable d'un simple point de vue probabiliste (au sens du calcul de fréquences).

Dans ce chapitre, nous allons, tout d'abord (1) faire un point sur les connaissances géologiques de la région, examiner les différentes propositions de reconstructions paléogéographiques, (2) exposer notre démarche analytique pour la détermination des relations de parenté entre les différentes aires d'endémisme que nous avons définies et enfin (3) explorer d'autres possibilités de recherche pour cette région si diversifiée et pourtant si méconnue.

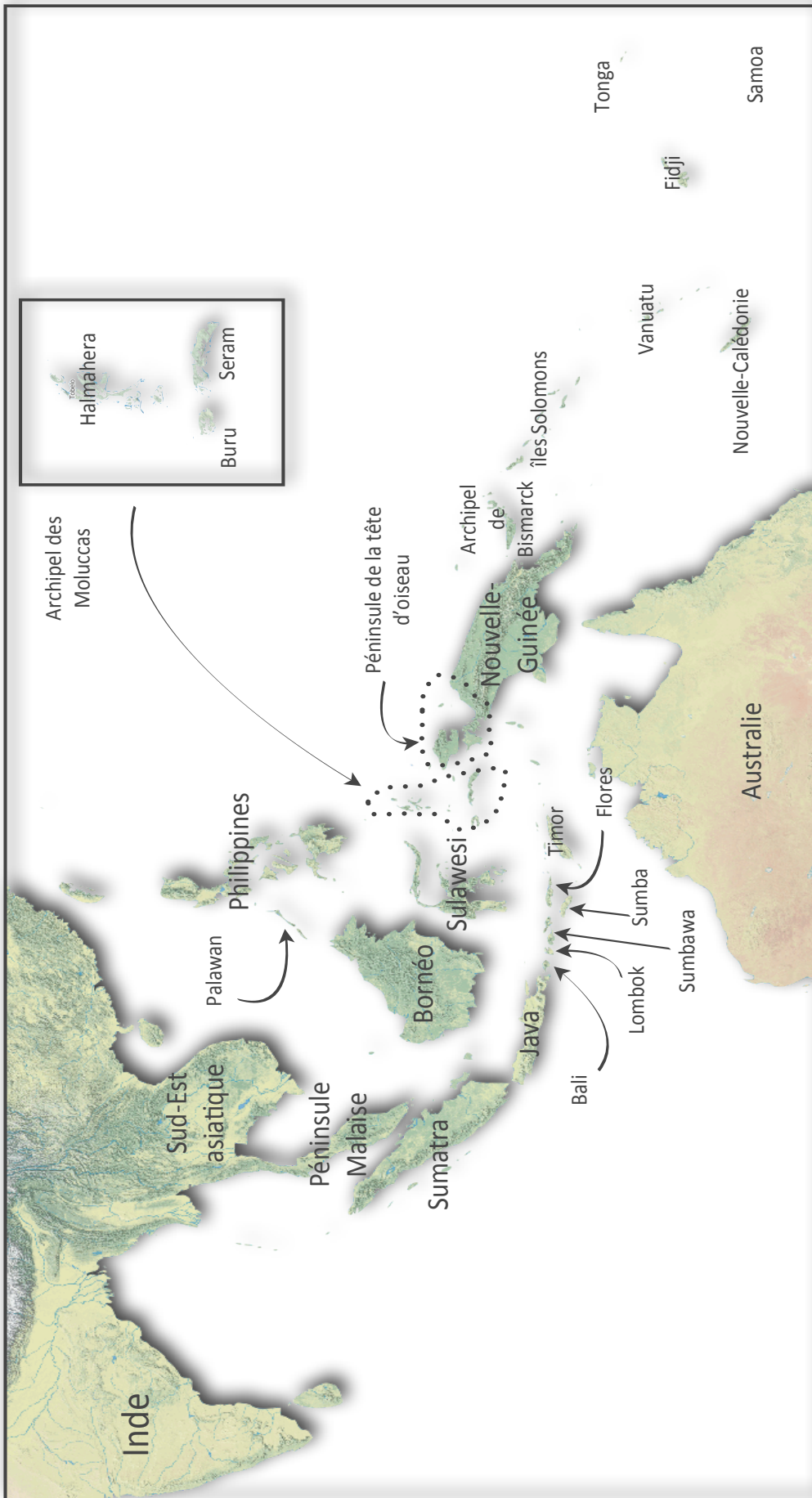


Figure 12 Carte de la région d'étude (modifiée d'après OpenStreetMap)

1. Histoire géologique de l'Asie du Sud-Est où comment réassembler un « puzzle » géant ?

Comme nous l'avons déjà précisé, l'Asie du Sud-Est et la région de l'Ouest du Pacifique sont très fragmentés en raison d'une forte activité géologique, à la fois volcanique et sismique. Étant entendu qu'il n'est pas possible ici de détailler toutes les étapes de la mise en place géologique de ces régions nous avons choisi de ne faire référence ici, qu'à 5 événements (ou périodes majeures) qui ont grandement contribué à façonner la région d'étude.

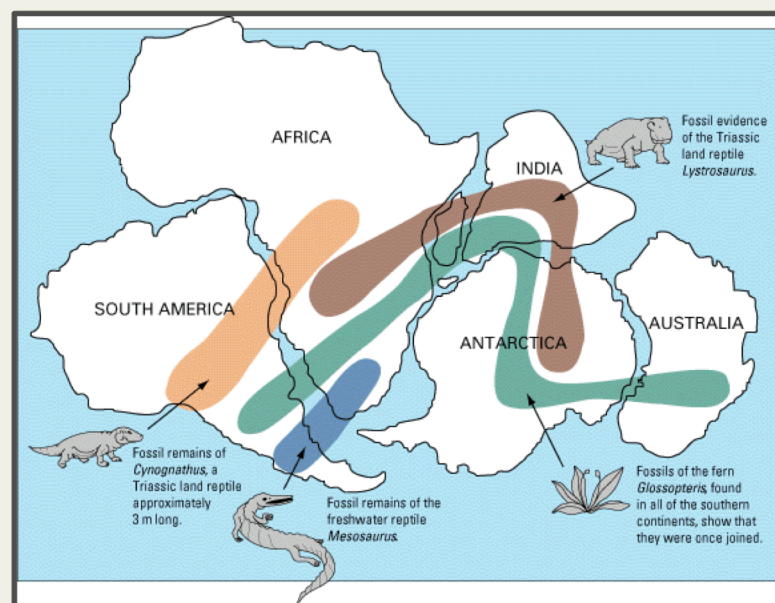
1.1. Séparation du Gondwana (-160 Ma)

Le premier événement à avoir façonné la région est l'éclatement du Gondwana (Encadré 8, Figure 13), qui a conduit à la séparation de l'Inde et du continent formé par l'Australie et l'Antarctique de l'Afrique (de Boer 1995 ; Veevers 2000 ; Hall 2002, 2011, 2012 ; Metcalfe 1988, 2002, 2011).

Encadré 8 Carte du Gondwana au Crétacé

Le Gondwana est un supercontinent formé à la toute fin du Protérozoïque (-600 millions d'années) et qui commence à se fracturer au Jurassique (-160 millions d'années).

Sur la carte ci-dessous sont représentées certaines distributions florales et fauniques fossiles montrant leur large répartition sur tout le Gondwana.



(Mc Carthy 2009)

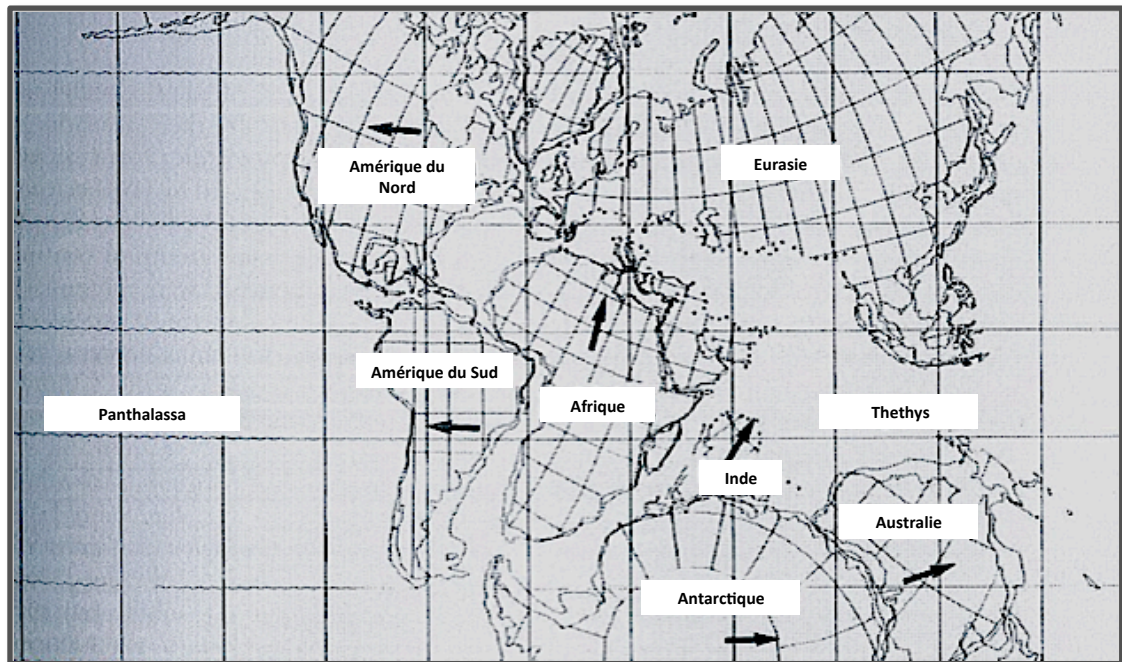


Figure 13 Reconstruction paléogéographique à -120 Ma montrant l'éclatement du Gondwana. Les flèches noires indiquent la direction des mouvements des principaux fragments (modifié d'après de Boer 1995c)

Ce supercontinent a commencé à dériver vers -180 Ma. L'Afrique et l'Amérique du Sud sont les premiers à s'individualiser tandis que l'Inde, l'Australie et l'Antarctique se déplace vers l'Est, tout en constituant un seul bloc (de Boer 1995c et références internes). Peu après, si ce n'est presque simultanément, l'Inde se sépare du bloc Australie-Antarctique et se déplace vers le nord.

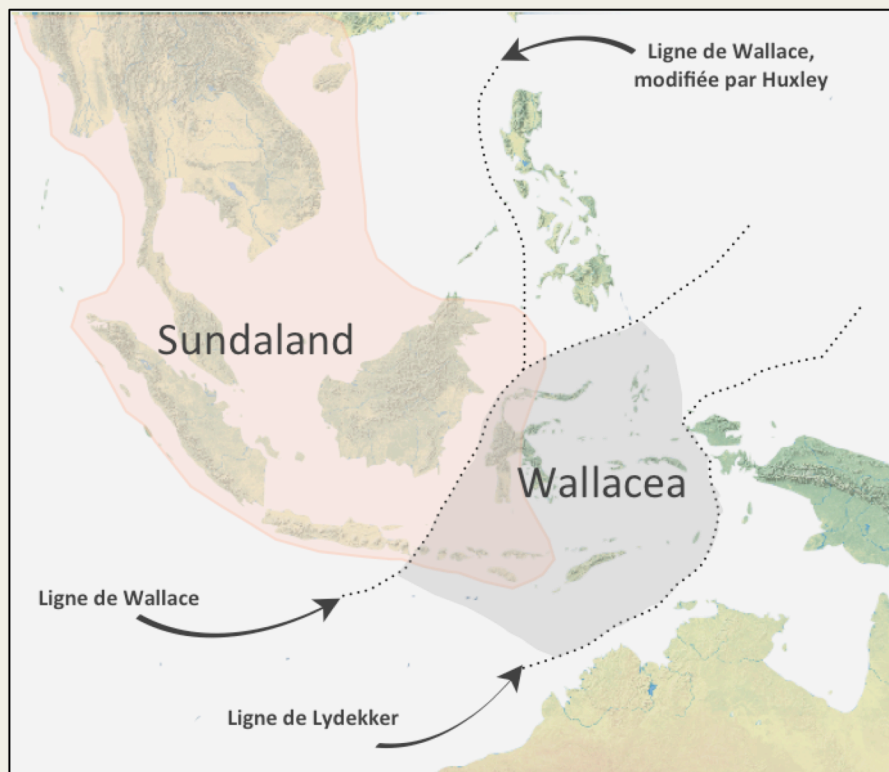
L'appartenance Gondwanienne de tous les composants continentaux de l'Asie du Sud-Est (*id est* le cœur du Sundaland, voir Encadré 9) est largement accepté (Metcalf 2011 ; Hall 2009, 2011, 2012 ; Hall et al. 2011).

Le Sundaland a été assemblé à partir de blocs continentaux qui se sont séparés du Gondwana au Paléozoïque et qui se sont amalgamés aux blocs asiatiques au Triassique (Metcalf 2011 ; Hall 2009, 2011, 2012 ; Hall et al. 2011). D'ailleurs, le bloc Sud de la Chine-Indochine-Est de la Malaise, qui est interprété comme ayant fait partie de la marge Indo-Australienne, s'est séparé du Gondwana par l'ouverture de la Paléo-Téthys au tout début du Dévonien (Metcalf 2002, 2011). Le bloc de l'ouest de Sumatra est aussi interprété comme ayant fait parti de cet ensemble de terranes ²(Metcalf 2006, 2011). La Figure 13 montre les diverses origines des blocs continentaux constituant l'Asie du Sud-Est et la région Wallacéenne.

² En géologie, un terrane est un fragment de croûte terrestre qui s'est détaché d'une plaque tectonique et agrégé sur une autre plaque (Dictionnaire de géologie, Foucault et Raoult, 2009)

Encadré 9 Carte du Sud-Est asiatique montrant le Sundaland ainsi que quelques exemples de lignes tracées dans la région et la Wallacea.

Le Sundaland une masse continentale incluant l'Asie du sud-est continentale, la Péninsule malaise (Provinces thaïlandaises, Malaisie péninsulaire, Singapour) ainsi que Bornéo, Java, Sumatra et Bali. Ces îles ont constamment été séparées ou reliées de l'Asie continentale à travers la montée ou l'abaissement du niveau de la mer imputée aux glaciations du Pléistocène. Raison pour laquelle la faune et - dans une moindre mesure - la flore sont très similaires à ce que l'on trouve en Asie du sud-est continentale : tigres, rhinocéros, primates, etc. Plus au sud, la ligne de Wallace sépare un biota asiatique et indomalais d'un biota australasien, formant une zone tampon, biologiquement mixte, appelée *Wallacea*.



Malgré des reconstructions géologiques/géographiques de plus en plus fines, il n'en reste pas moins certaines incertitudes concernant l'assemblage du Sundaland. Notamment en ce qui concerne Bornéo qui contient les plus larges blocs et dont les âges sont antérieurs au Mésozoïque. De plus, il est souvent admis que le sud-ouest de Bornéo était attaché au Sundaland avant le Crétacé. Les régions environnantes (Sarawak, Abah, Est de

Kalimantan et de l'Est de Java au Sud de Sumatra) sont interprétés comme des assemblages suivant des subductions au Crétacé et au Tertiaire (Figure 14). Aussi, Hall (2012) considère que certains fragments du Sundaland (constituant le bloc ouest de la Birmanie) auraient pour origine le nord-ouest de l'Australie et auraient dérivé au Jurassique. Au milieu du Crétacé, une collision a eu lieu entre un ou plusieurs fragment(s) asiatique(s) avec l'est du Sundaland. Une origine asiatique est soutenue pour le Nord-Ouest de Kalimantan (Williams et al. 1988) et Sarawak (Hutchison 2005). Des fragments provenant encore du nord-ouest de l'Australie se seraient accrochés au bloc est-malaisien du début du Crétacé jusqu'au milieu du Crétacé (Hall et al. 2009a) dont le sud-ouest de Bornéo constitue la plus importante partie.



Figure 14 Carte de Bornéo (actuel) avec le découpage des provinces
(<http://www.borneo.com.au/>)

Le bloc est de Java-ouest de Sulawesi est interprété comme le « bloc Argo » (Veevers 1988, Powell et al. 1988) et se serait séparé du nord-ouest de l'Australie vers -155 Ma. Ce bloc serait d'ailleurs constitué de nombreux petits blocs plutôt qu'un bloc d'une seule pièce, ajouté au Sundaland vers -90 Ma.

1.2 Ouverture de la mer de Tasmanie (-95 Ma)

Ce deuxième événement majeur a eu pour conséquence la séparation du bloc Australie-Antarctique, avec l'Australie changeant de direction de déplacement (de l'est vers le nord) (Figure 13) (de Boer 1995 ; Metcalfe 1988, 2002).

La dérive de fragments provenant de l'Australie a déterminé la forme et le caractère de la lisière australienne (Hall, 2012), ce qui a eu une influence majeure sur les développements de la collision Australie-Sundaland au Néogène.

1.3 Collision et subduction (-50 Ma)

Le troisième fait marquant est la collision de l'Inde avec le continent asiatique vers -50 Ma (Hall 2002, 2011, 2012 ; Metcalfe 1988, 2002).

À la même période, la plaque Australienne continua à avancer obliquement sur la plaque du Pacifique ce qui força le sol de la mer Thetys (mer qui séparait initialement les plaques Australienne et Pacifique) à subduire sous la plaque du Pacifique (de Boer et Duffels 1996 et références internes). L'activité volcanique accompagnant cette subduction a donné naissance à un arc volcanique sur les bords sud et ouest de la plaque Pacifique : l'arc du Pacifique Ouest (aussi connu comme l'arc Indo-Mélanésien (de Boer 1995) ou encore l'arc mélanésien externe par Duffels 1986, Holloway 1979 et Hall 2002).

Depuis l'Eocène, la distance marine entre les îles de Bornéo et de Sulawesi était au moins aussi grande que maintenant (Hall 2012).

1.4 Fracturation de l'arc du Pacifique Ouest ou arc mélanésien (-40 Ma)

Vers -42 Ma, la plaque des Philippines s'individualise de la plaque du Pacifique. Une collision entre la plaque du Pacifique et le continent asiatique provoqua une fracture de l'arc du Pacifique Ouest. La partie nord de cet arc changea de direction et provoqua une collision entre la partie centrale des Philippines et la partie ouest des Philippines, tandis que l'autre partie de cet arc (*id est* le Sepik arc, à l'ouest (de Boer et Duffels 1996)) continua sa course vers l'ouest jusqu'à atteindre l'Australie. Cette collision provoqua des fracturations qui donneront naissance aux îles actuelles suivantes : Vanuatu, Salomons et Fidji (Hall 2002).

1.5 Réorganisation des limites des plaques au Cénozoïque (vers -25 Ma)

Cette période a vu les changements les plus importants au niveau des limites des plaques. Cette massive réorganisation est sans doute le reflet d'évènements autres que ceux localisés dans cette région comme la collision entre le bloc Philippines Est-Halmahera-Arc de Caroline du Sud avec les marges du nord de la Nouvelle-Guinée ce qui changea la nature des limites sud de la plaque de la mer des Philippines (Hall 2002) (Figure 15, Figure 16). Une autre cause possible serait le début d'une collision entre le plateau de Ontong Java et de l'arc mélanésien. Enfin, on peut aussi envisager que le premier contact entre la croûte continentale Australienne, autour de la région de l'actuelle péninsule de la tête d'oiseau (partie nord et antérieure de la Nouvelle-Guinée) avec les marges du Sud-Est asiatique (dans la région de l'actuel Sulawesi) soit une des raisons pour ce réarrangement de plaques (Hall 2002).

Au tout début du Miocène, il y a eu collision entre l'Australie et le Sundaland (-25 Ma), et plus tard une collision au nord de Bornéo avec l'extension continentale du sud de la Chine (Hutchinson et al. 2000, Hall and Wilson 2000). Ces collisions ont donné lieu à l'émergence de montagnes à Sulawesi, l'arc de Banda (ensemble des îles constituant le groupe des « Lesser Sunda islands», incluant Timor, Flores, Komodo pour les plus connues) et Bornéo.

Ces collisions ont pu permettre selon un certain degré, des échanges floristiques entre les flores Australasiennes et du Sud-est asiatique mais ces dispersions n'ont eu que peu d'effets sur la composition générale du Sundaland, à la différence de l'invasion suivant la collision de l'Inde avec la plaque asiatique (Hall 2012).

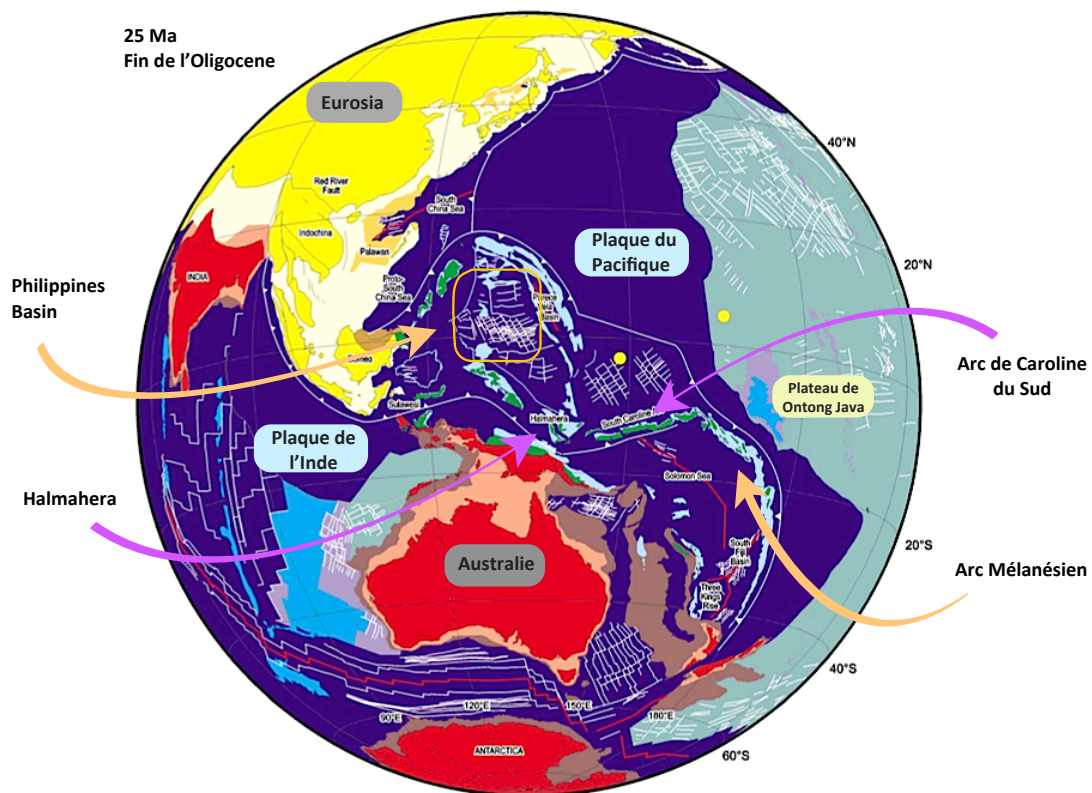


Figure 15 Reconstruction géographique. Le bloc Philippines Est-Halmahera-Arc de Caroline du Sud commence à se heurter à la plaque Australienne sur la marge nord de la Nouvelle Guinée et la plateau de Ontong-Java entre en collision avec l'arc Mélanésien (modifié d'après Hall 2002, avec autorisation de reproduction de l'auteur).

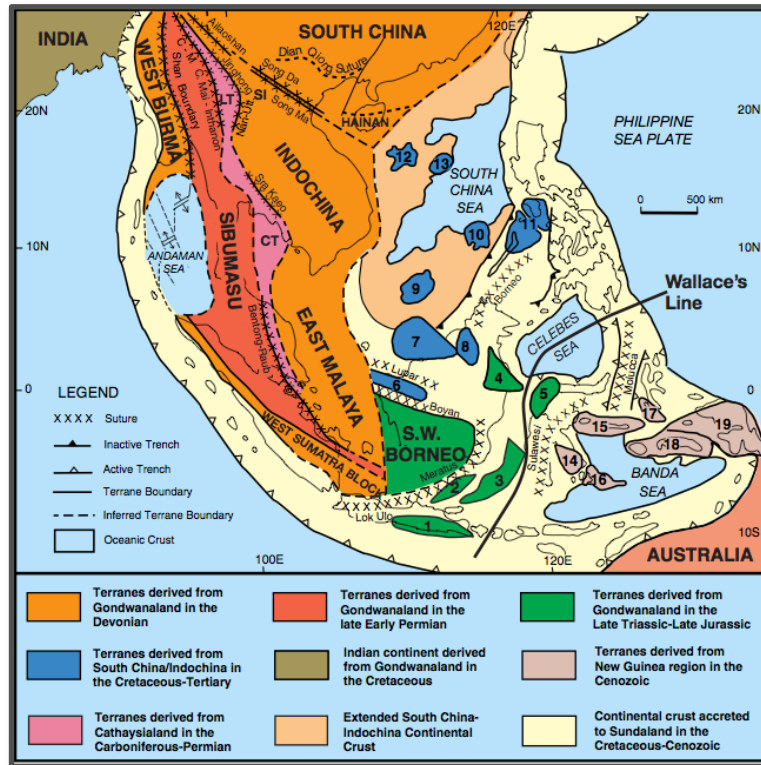


Figure 16 Répartition des blocs continentaux, des fragments et des terranes, et les principaux points de suture de l'Asie du Sud-Est. Numérotation des blocs microcontinentaux: 1, Java-Est, 2, Bawean, 3, Paternoster, 4, Mangkalahat, 5, Sulawesi Ouest, 6, 7, Semitau; Luconia, 8, Kelabit-Longbowan, 9, sol îles Spratly-dangereuses; 10, Banque Reed, 11, Nord de Palawan, 12, Iles Paracel, 13, Macclesfield Bank, 14, East Sulawesi, 15, Bangai-Sula, 16, Buton, 17, Obi-Bacan, 18, Buru-Seram, 19, West Irian Jaya . LT, Lintang terrane, CT, Chanthaburi terrane, CM, Changning-Menglian Suture; C.-Mai - Inthanon, Chiang Mai - Inthanon Suture; Nan-UTT, Nan-Uttaradit suture (Metcalfé 2011 avec autorisation de reproduction de l'auteur).

1.6 Modifications des mouvements et des limites des plaques au Néogène (vers -5 Ma)

D'importants changements de direction et de vitesse dans les mouvements des plaques ainsi que leurs délimitations ont eu lieu à cette période et perdurent encore aujourd'hui. Suite à la fracturation de l'arc mélanésien, Fiji se trouve maintenant bien isolée, les principaux terranes constituant la Nouvelle-Guinée se trouvent à une position très proche de leur position actuelle (Hall 2002). Hall (2002) suggère que c'est à cette période que l'élévation actuelle du nord de la Nouvelle-Guinée a débuté, ce qui a eu un impact certain sur les

modèles climatiques. En effet, cela a eu pour conséquences un accroissement de l'aridité dans le sud de la Nouvelle-Guinée ainsi que des précipitations dans l'ensemble de l'île. Ce rapide changement climatique aurait fourni un environnement varié qui a certainement contribué à développer l'incroyable diversité florale et animale actuelle de la Nouvelle-Guinée (Hall 2002).

2. Distribution des taxons dans la région : lignes tracées

Nous l'avons déjà dit, la région couvrant l'Asie du Sud-Est et l'Ouest du Pacifique est une région extrêmement riche et variée en terme de biodiversité. Et pourtant, certaines « signatures » se dessinent lorsqu'il s'agit d'observer les combinaisons taxiques. On y voit une claire démarcation entre des taxons « asiatiques » et des taxons plutôt « australiens » de part et d'autre de la Wallacea (qui reste une zone de « mélange »). Cette observation ne date pas d'hier, Alfred Russel Wallace a été un des premiers observateurs et découvreurs de cette différence taxique. Dans son livre « The Malay Archipelago », il relate son périple en Malaisie, Indonésie et Nouvelle-Guinée entre 1854 et 1862 où il fût même parfois l'un des tout premiers européens à fouler le sol de certains lieux (van Oosterzee 1997). De ces différents voyages, Wallace tracera la célèbre ligne qui porte son nom lors d'un discours à la Royal Geography Society (Encadré 10).

Encadré 10 La ligne de Wallace

« In the archipelago, there are two distinct faunas rigidly circumscribed, which differ as much as those of South America and Africa and more than those of Europe and North America. Yet there is nothing on the map or on the face of the islands to mark their limits. The boundary line often passes between islands closer than others in the same group »... « We have here a clue to the most radical contrast in the Archipelago, and by following it in detail I have arrived at the conclusion that we can draw a line among the islands, which shall divide them that one-half shall truly belong to Asia, while the other shall no less certainly be allied to Australia ». **Wallace 1869, volume 1 :13.**

En 1863, à Londres, Wallace lu un article à la Royal Geography Society, sur la géographie de l'archipel Malais. Il traça alors une ligne rouge sur la carte passant sous les îles du détroit de Makassar. A l'ouest il écrivit « région Indo-Malaise » et à l'est il écrivit « région Australo-Malaise ». Cette ligne est devenue la « ligne de Wallace ».

La ligne de Wallace est la plus célèbre et en même temps la plus discutée de toutes les frontières biogéographiques jamais décrites. Elle est essentiellement basée sur la distribution d'oiseaux, principalement des perroquets mais également d'autres vertébrés. Longue d'environ 25 km, elle sépare les marsupiaux des tigres, les passereaux et les cacatoès des barbets et trogons (van Oosterzee 1997).

Curieusement, la ligne de Wallace ne fut pas la première ligne biogéographique tracée dans la région (Figure 17). En 1846, Salomon Muller définit sa ligne basée essentiellement sur l'écologie, faisant remarquer par exemple que des conditions arides étaient présentes à l'est depuis l'île de Java jusqu'à Tanimbar et que cela affectait la végétation. La ligne de Muller est similaire à celle de Wallace, elle passe à l'est de celle-ci entre Sumbawa et Flores dans les îles du Lesser Sunda. La ligne de Murray, énoncée en 1866, est basée sur la distribution de mammifères et montre une route assez similaire à la ligne de Wallace. La seule différence étant qu'elle passe entre Bali et Java alors que la ligne de Wallace passe entre Bali et Lombok.

En 1868, le terme de « ligne de Wallace » fut énoncé par Thomas Henry Huxley (1825-1895) dans un article basé sur la distribution des oiseaux. Huxley (1868) modifia quelque peu le tracé de la ligne la faisant passer entre Palawan et les Philippines. Cependant, Wallace n'accepta jamais cette extension de sa ligne et il est donc plus correct de l'appeler « ligne de Huxley ».

En 1896, Lydekker traça sa ligne qui correspond avec le bord du continent Australien, tandis que Weber, en 1902, dessine la sienne principalement basée sur la distribution de poissons d'eaux douces.

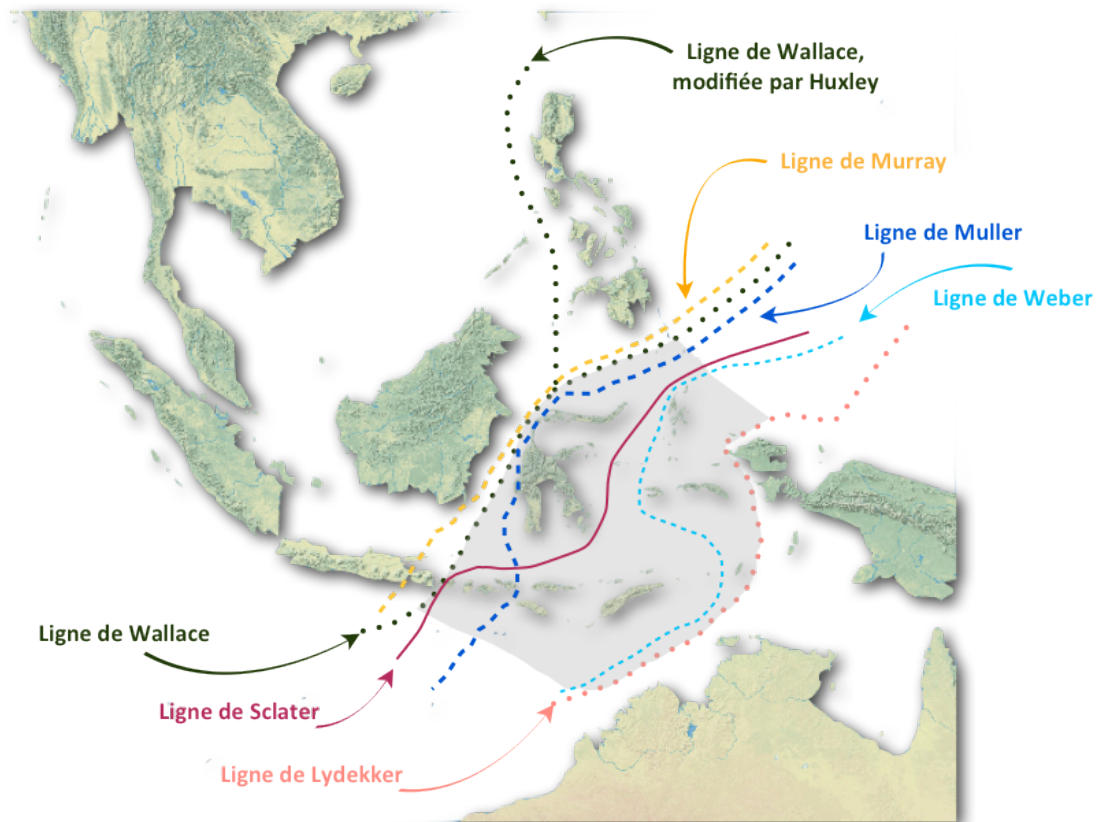


Figure 17 Lignes biogéographiques tracées dans la région (d'après van Oosterzee 1997)

In fine, seule la ligne de Wallace perdure réellement et est toujours discutée par les biogéographes. Seulement, existe-t-elle vraiment ? Peut-on dire qu'il y a une barrière biogéographique délimitant des taxons d'origine asiatique et des taxons d'origine australienne ? Où se trouve cette frontière ?

3. Climat et changements du niveau de la mer au Quaternaire

Comme nous venons de le voir, l'histoire géologique est déterminante pour justifier de la Biogéographie de la région. S'ajoute à cela une composante climatique non des moindres et une modification des niveaux des mers ayant eu lieu au Quaternaire et qui, sans nul doute, a dû avoir d'importantes conséquences sur les taxons tant aquatiques que terrestres présents dans la région à cette époque (De Bruyn 2012).

Par exemple, Cannon en 2012 écrit :

« The major land masses on the shelf, including the major islands of Borneo, Sumatra and Java, have been connected to the Asian mainland through land bridges for the majority of the Quaternary period (started at 2,58 Ma). This connectivity should have provided abundant opportunity for dispersal throughout the region and yet a large amount of evidence suggests that strong biogeographic barriers exist, particularly between Sumatra and Borneo or east to west. Apparently, the barrier was environmental and not geographic, with several authors proposing a 'savanna corridor' running north-south through the western portion of Sundaland». **Cannon, 2012 :118.**

Concernant la faune de Nouvelle-Guinée, Richardson et collaborateurs évoquent la connection avec l'Australie pour expliquer une colonisation de la Nouvelle-Guinée par des taxons australiens :

« Zoologists classify New Guinea with the Australian region (Darlington, 1957) because it forms part of the Australian plate and was connected to mainland Australia when sea levels were lower (e.g. during the last glacial maximum LGM). This allowed faunal migrations to occur from Australia to New Guinea. Few of these animals were able to disperse further west and virtually none made it upon Sulawesi. This faunal discontinuity was the basis for Wallace's line ». **Richardson et al. 2012 :144.**

Ils rajoutent que l'absence de fossiles de plantes en Nouvelle-Guinée est aussi la preuve qu'il n'y a pu avoir de colonisation par les plantes au milieu du Miocène à cause du niveau plus élevé de la mer qui rendait ces territoires submergés, territoires qui n'auraient émergés que vers -10 Ma selon Hall (Hall 2009).

De Bruyn et collaborateurs dans une grande analyse portant sur la faune d'eau douce de l'archipel Indo-Australien indiquent que la plus grande partie des Philippines et des îles wallacéennes sont reconnues comme étant des îles océaniques, donc, sans connection avec aucune des îles environnantes depuis leur émergence (van Oosterzee 1997), la présence de faune sur ces îles est donc, selon eux, due à de la dispersion.

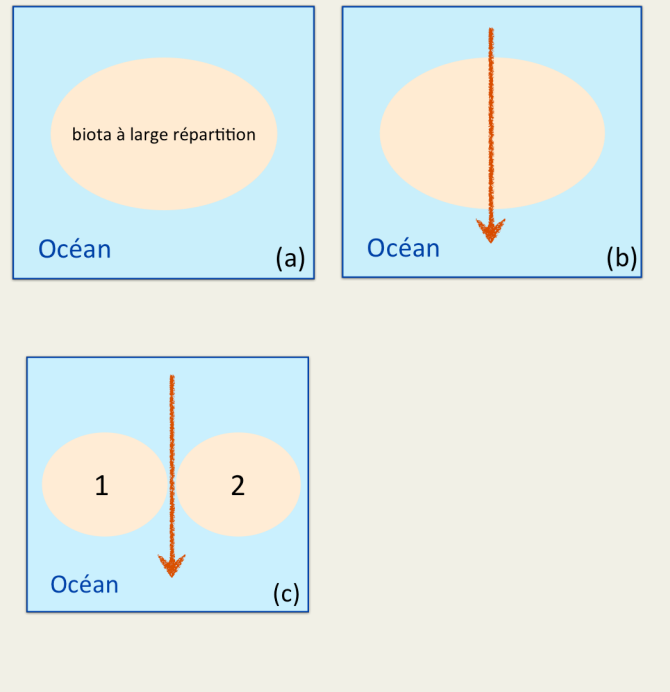
Ces quelques extraits tirés de la littérature exposent le principal problème lié à la Biogéographie : la distribution des taxons ne peut être expliquée que par un modèle de dispersion à partir d'un centre d'origine. Or il existe, parmi les causes explicatives possibles, un autre processus majeur possiblement responsable de la présence d'une telle biodiversité : la vicariance. Pour répondre aux questionnements relatifs à la présence de faune sur des îles océaniques et/ou volcaniques, Parenti et Ebach, dans un très récent article paru en janvier 2013 (Parenti et Ebach 2013), expliquent que ce processus de vicariance peut très bien se produire sur des îles volcaniques. Si l'on regarde l'exemple de l'Encadré 11, en (a), nous avons un biota avec des taxons à large répartition dans ce biota, en (b), la formation d'une chaîne volcanique dans le biota et en (c) la séparation de ce biota en deux biotas sœurs et la croissance de la chaîne volcanique.

Parenti et Ebach affirment que le biota marin évolue en parallèle de l'émergence de chaînes volcaniques et que donc, lorsque les îles volcaniques émergentes deviennent suffisamment hautes pour que des ruisseaux d'eaux-douces puissent s'établir, les taxons qui vivaient exclusivement en mer peuvent appartenir ensuite au biota terrestre dans la mesure où ils se trouvent sur-élevés sur le milieu terrestre.

« *This is a gradual, co-evolutionary process* ». **Parenti et Ebach 2013 :2.**

En définitive, la discussion autour de l'explication relative à la distribution des taxons en revient toujours à l'opposition dispersion *versus* vicariance.

Encadré 11 représentation simplifiée de la formation d'une chaîne volcanique divisant un biota au milieu de l'océan (modifié d'après Parenti et Ebach 2013)



Nous venons de finir un long préambule qui a posé les contextes géologique, climatique et naturel de notre région d'étude. Nous renvoyons le lecteur au chapitre I pour la partie « Méthodes » et à l'article suivant pour le « Matériel » ainsi que la « Discussion/Conclusion » qui synthétise l'analyse que nous avons mené.

4. « Comparative biogeography of Southeast Asia and the West Pacific region »

4.1 Contexte de l'article

La communauté des biogéographes se divise généralement en deux « groupes » : ceux qui recherchent des patterns généraux et ceux qui recherchent des explications narratives (Ebach et Humphries 2002). Nous nous plaçons dans le premier groupe et nous voulons rechercher les relations de parenté entre aires biogéographiques sur la base de la présence concomittente de taxons non apparentés. Pour cela, nous avons

choisi de sélectionner un échantillonnage non exhaustif mais suffisamment important pour obtenir un signal biogéographique.

Notre analyse est détaillée dans l'article suivant « Comparative biogeography of Southeast Asia and the West Pacific region », actuellement en révision pour la revue *Journal of Biogeography*.

4.2 Résumé

Nous utilisons la méthode 3ia pour rechercher l'existence (ou non) d'un pattern général concernant la région du Sud-Est asiatique et de l'Ouest du Pacifique. Pour cela, 35 cladogrammes sources de taxons non-apparentés ont été sélectionnés. Le résultat obtenu est ensuite comparé aux études géologiques faisant références pour vérifier la congruence de notre pattern global à l'histoire géologique. Enfin, nous discutons de la pertinence d'une telle approche à l'interprétation de patterns généraux.

4.3 Conclusions

Les relations de parenté entre aires localisées dans le Sud-Est asiatique et la région de l'Ouest du Pacifique sont toujours débatus à cause de l'histoire géologique complexe de cette région ainsi que l'énorme diversité des taxons présents. Des méthodes cladistiques ont déjà été utilisées pour reconstruire les relations de parenté entre les aires de cette région mais jamais avec un nombre si important de cladogrammes de taxons non-apparentés (35). Des patterns congruents devraient être le reflet de causes communes. L'extraction de l'information sur les relations de parenté entre aires contenues dans les analyses phylogénétiques des taxons constitue un test d'homologie pour les aires homologues. Nous avons obtenu un aréagramme général basé sur un échantillonnage empirique qui donne un nouveau regard sur la classification des aires dans cette région du monde et nous permet de faire de nouvelles inférences sur l'histoire évolutive de cette région.

ORIGINAL ARTICLE

Number of words: 6877 words

Comparative biogeography of Southeast Asia and the West Pacific region

Visotheary Ung^{1*}, René Zaragueta-Bagils^{2,1} and David M. Williams³

¹ CNRS UMR 7207 (CNRS-UPMC-MNHN), 57 rue Cuvier CP43 75005 Paris, France.

² UPMC Univ Paris 06

³ Department of Life Sciences, the Natural History Museum, Cromwell Road, London, SW7 5 BD, United Kingdom.

* Correspondence: Visotheary Ung, CNRS-UMR 7207 (CNRS-UPMC-MNHN), 57 rue Cuvier CP43 75005 Paris, France.

Email: visotheary.riviere-ung@snv.jussieu.fr

Running title: Comparative biogeography of Southeast Asia

AIM

To (1) use a compilation of phylogenies to investigate area relationships among Southeast Asia and the West Pacific region, (2) use recently published geological reconstructions of the region in order to compare the results gained from the comparative analysis; (3) discuss the relevance of such an approach to the interpretation of general pattern.

The main questions addressed are: how to explain actual distributions of taxa in Southeast Asia and the West Pacific region? Is there an emerging common pattern?

LOCATION

Southeast Asia (from India to New Guinea) and the West Pacific region (from Australia to Pacific islands).

METHODS

We (1) used three-item analysis (3ia) to find area relationships, (2) used geological literature to compare results.

RESULTS

Three-item analysis found 27 most optimal trees. An intersection tree, i.e. an area cladogram, reconstructed from the taxon-area statements (common to all 27) had an overall retention index 84.8% and retrieved 13 nodes with two major branches congruent with a separation between Southeast Asia and the West Pacific region.

MAIN CONCLUSIONS

The relationships between areas located in Southeast Asia and the West Pacific region, is still debated because of its complex historical geology and the enormous diversity of taxa occupying the region. Cladistic methods have previously been used to reconstruct the relationships between areas in the region but never with such a high number of unrelated taxa (35). Congruent patterns revealed by the combination of unrelated taxa should reflect a common cause. The extraction of information on area relationships contained in phylogenetic analyses of taxa consists of testing for area homologs.

We obtained an area cladogram from this region based on an empirical dataset which give account for new insights regarding area classification in the region.

KEY-WORDS

Areas of endemism, Cladistic, Comparative biogeography, Southeast Asia, West Pacific, Three-item analysis (3ia), Pattern, Process.

INTRODUCTION

Comparative biology provides rigorous empirical theory coupled with methods of analysis to assist in determining Earth history (Parenti and Ebach 2009). The ichthyologist Donn Eric Rosen (1929—1986; Nelson *et al.* 1987) first proposed the idea, he stated that biological and geological patterns can shed light on one another (by the process of “reciprocal illumination”) but cannot test, and therefore cannot reject, one another (Rosen 1978; see also Parenti 2006).

The Southeast Asia and the West Pacific region (SEA-WP hereafter) have long been noted as a centre of biodiversity for both marine and terrestrial taxa (Michaux 2010; Woodruff 2010). The complex geological history of the SEA-WP region makes it a compelling area for biogeographical studies (Hall 1998, 2002, 2011). To this end, we have compiled a large dataset composed of 35 non-related taxa and applied cladistic principles to retrieve a general area cladogram of the region since this sort of analysis is currently lacking (Ung 2013).

In general, cladistics makes no *a priori* assumptions about the process (or processes) responsible for any observed patterns of distribution. However, vicariance – the diversification of biotas following geological fragmentation – may more parsimoniously explain the congruence of distributional patterns found among unrelated taxa rather than any other processes.

De Boer wrote that “when several monophyletic groups of species comply with one and the same generalized area cladogram, we can safely assume that these groups did not acquire their distribution patterns independently by chance dispersals, but that they responded similarly to the same geological events. Area cladistic analysis should therefore always be based on two or more, preferably unrelated, groups” (de Boer 1995d).

Although historical biogeography is sometimes known as vicariance biogeography, the method does not intend to provide explanations about any processes involved in taxon diversification. That is, the aim is not to explain events that have led to the current distribution of taxa. Nonetheless, finding a certain level of congruence implies that a process has affected a large number of non-related taxa alike. Of course, dispersal may be an explanation but the problem remains as to how to interpret congruent patterns of non-related taxa which have different dispersal capabilities.

The complexity of geo-tectonic movements makes SEA-WP a challenging area for historical biogeographical studies, since many of the changes that occurred in the relative positions of the various islands and their parts suggest that a great amount of vicariant speciation has taken place. Therefore, as in systematics, where characters are discussed according to their discovered status (synapomorphy versus homoplasy), the obtained areagram will be interpreted in terms of the combination of taxa used to reconstruct it.

The purpose of this study is to perform a comparative biogeographical analysis of a compiled dataset composed of cladograms of taxa (plants and animals). The comparison with recently published geological reconstructions will give new clues concerning relationships between areas of endemism in the region. Finally we will discuss the relevance of such an approach to the interpretation of the general pattern retrieved.

MATERIAL

Data

Areas of endemism

As SEA-WP biogeography is a highly debated topic, we have chosen to begin by focusing on the study of Turner *et al.* (2001). These authors collected (for the first time) a large collection of both plants and animals cladograms. The dataset consists of distribution patterns and phylogenies for 29 monophyletic groups, selected according to the following criteria:

1. Availability of a cladistic phylogenetic hypothesis at species level, constructed using the method of parsimony (*sensu* Wagner parsimony, Kluge & Farris 1969);
2. Availability of detailed information for the distribution patterns of all terminal species;
3. Species exclusively – or at least predominantly – occurring in the region of interest;
4. The degree of confidence in the accuracy of the cladogram.

From this collection of 29 cladograms of taxa, 22 were retained according to the number of informative characters generated. We have selected 13 recently published cladograms of taxa distributed over SEA-WP to increase the dataset following the same criteria as described above.

The areas of endemism used are illustrated in Fig. 1. They have been delimited by the presence of a unique combination of taxa (Axelius, 1991).

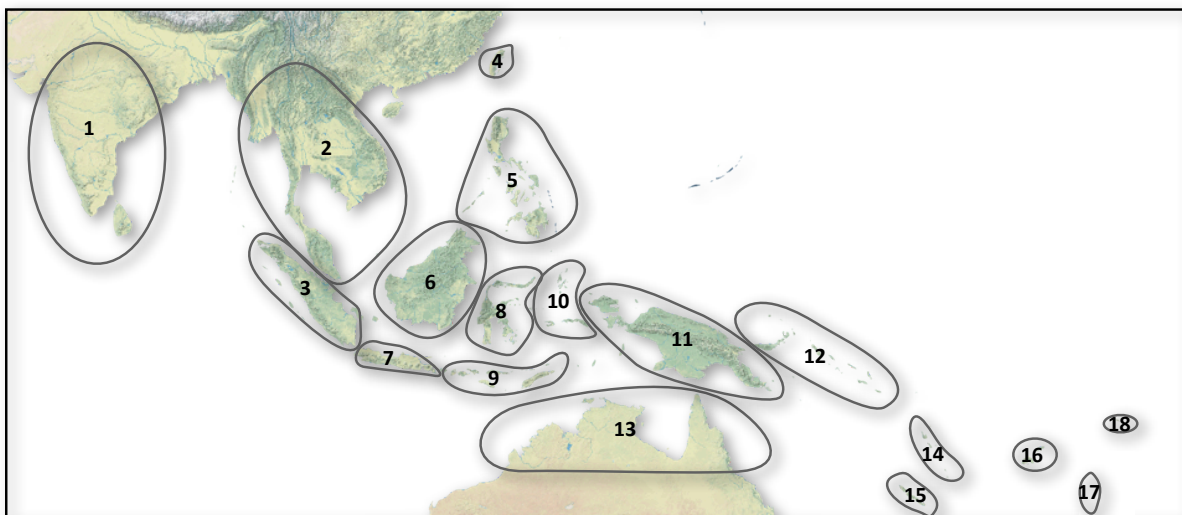


Figure 1: Areas of endemism of the biogeographical study. (1) India, (2) Southeast Asia, (3) Sumatra, (4) Taiwan, (5) Philippines, (6) Borneo, (7) Java, (8) Sulawesi, (9) Lesser Sunda Islands, (10) Moluccas, (11) New Guinea, (11a) Palawan, (12) Bismark archipelago and Solomon Islands, (13) North Australia, (14) Vanuatu, (15) New Caledonia, (16) Fiji, (17) Tonga, (18) Samoa.

Following recommendations made in Turner *et al.*, all taxa occurring outside the region of interest were discarded, i.e. the terminal taxa was merely deleted from the cladogram. In total, 18 areas of endemism were defined (Table 1).

Areas of endemism	Code
<i>India</i>	India
<i>Southeast Asia</i>	SEAsia
<i>Taiwan</i>	Taiwan
<i>Sumatra</i>	Sumatra
<i>Java</i>	Java
<i>Borneo</i>	Borneo
<i>Philippines</i>	Philippines
<i>Sulawesi</i>	Sulawesi
<i>Lesser Sunda Islands</i>	LSI
<i>Moluccas</i>	Moluccas
<i>New Guinea</i>	New_Guinea
<i>Bismark archipelago and the Solomon Islands</i>	Bismark_Solo
<i>Vanuatu</i>	Vanuatu
<i>New Caledonia</i>	New_Caledonia
<i>Fiji</i>	Fiji
<i>Samoa</i>	Samoa
<i>Tonga</i>	Tonga
<i>North Australia</i>	North_Aust

Table 1: Areas of endemism defined and their code coding

Taxon cladograms

Tables 2 presents a list of taxa used by Turner *et al.* (2001), all details are available here: <http://www.blackwellpublishing.com/products/journals/suppmat/jbi/jbi526/jbi526sm.htm>.

Table 3 presents our additional dataset. The entire dataset is composed of 35 cladograms of taxa which includes 839 species (see Appendix S1 for all distributions).

Clades	Clades Coding	Number of informative TACs (paralogy- and MAST- free)
<i>Arytera</i>	ARY	1200
<i>Bromheadia</i>	BRO	27104
<i>Casnoidea</i>	CAS	1824
<i>Chlorocistini</i>	CHL	121
<i>Crocidura</i>	CRO	4
<i>Cupaniopsis</i>	CUP	24
<i>Dundubia jacoona</i> assemblage	JAC	25
<i>Erismanthaeae</i>	ERI	12
<i>Fordia</i>	FOR	1
<i>Genianthus</i>	GEN	334
<i>Guioa</i>	GUI	30132
<i>Halobates princeps</i> group	HALC	16
<i>Halovelgia malaya</i> group	HAL	4
<i>Haloveloides</i>	HALO	160
<i>Hydrophilidae</i>	HYD	32
<i>Lepidopetalum</i>	LEP	23
<i>Megarthus</i>	MEGA	1
<i>Neoscortechinia</i>	NEO	24
<i>Parkia</i>	PAR	2272
<i>Rhysotoechia</i>	RHY	22
<i>Xanthophytum</i>	XAN	36
<i>Xenobates</i>	XEN	904

Table 2: Taxa extracted from Turner *et al.*'s study (2001) used for biogeographical analysis. The 3rd column documents the numbers of TACs (taxon-area cladograms) that are both paralogy- and MAST-free. High numbers were excluded from the final analysis, in this case *Guioa* (30132 TACs) and *Bromheadia*. Please refer to the original article to have full details regarding the references of the cladograms of taxa.

Clades	Clades Coding	Number of informative TACs (paralogy- and MAST-free)	References
<i>Calicnemiinae</i>	ODO	1	Gassman (2005)
<i>Chlorocyphidae</i>	CHLO	103	Van Tol & Gassman (2007 after van Tol, 1998)
<i>Cosmopsaltriina</i>	COS	1296	Duffels & Turner (2002)
<i>Cycas</i>	CYC	4	Keppel et al. (2008)
<i>Cyrtandra</i>	CYR	94	Cronk et al. (2005)
<i>Gallicolumba</i>	GAL	15	Jønsson, K.A. et al (2011)
<i>Gehyra</i>	GER	18	Heinicke, M.P. et al. (2011)
<i>Halobates</i>	HAL	4	Andersen (1998)
<i>Megapodidae</i>	MEG	1	Dekker (2007, redrawn from Birks and Edwards, 2002)
<i>Pittosporum</i>	PIT	1	Gemmill et al.(2002)
<i>Tropidophorus</i>	TRO	2	Honda et al. (2006)
<i>Varanus</i>	VAR	2	Ast (2001)
<i>Vireya</i>	RHO	21728	Brown et al. (2006)

Table 3: Additional taxa for the analysis. The 3rd column documents the numbers of TACs (taxon-area cladograms) that are both paralogy- and MAST-free.

METHODS

The computer program LisBeth (Zaragueta-Bagils *et al.* 2012) (<http://infosyslab.fr/downloadlisbeth/LisBeth.exe>) has been developed in order to use biogeographical characters obtained through a new implementation of three-item analysis (3ia hereafter) (Nelson and Ladiges 1991a, b, 1992; Nelson and Platnick 1991). LisBeth finds optimal trees, by applying compatibility analysis to suites of three-item statements (Cao 2008; Zaragueta *et al.* 2012).

In biogeography, there are 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) are respectively resolved using Paralogy-free Subtree analysis (Nelson and

Ladiges 1996) and the 'transparent method' (Ebach *et al.* 2005) which implements a version of Assumption 2 (Nelson and Platnick 1981).

LisBeth builds a summary tree, called an intersection tree (Cao *et al.* 2009), by combining the three-item statements present in all optimal trees; the intersection tree can be viewed as the minimal tree (for more details, see Zaragueta *et al.* 2012) and is by essence rooted since it is reconstructed from 3is (i.e. rooted trees).

LisBeth accepts taxon-area cladograms (TACs hereafter) as direct input information (i.e. the character) rather than via a tabular matrix. That is, LisBeth utilises the paralogy-, MAST-free TACs (Cao *et al.* 2007). This particular representation of knowledge gives new information and allows new assumptions regarding area evolution.

Traceability of the characters

Although this article is not meant to emphasize the LisBeth suit of programs, for the sake of clarity we have to introduce here a new tool we recently implemented in LisBeth which informs the users of which distributions of taxa "support" each node, a, information that can be related to the concept of synapomorphy.

RESULTS

Fig. 2 shows the tree yields from the analysis of the 35 independent taxa selected for this study, 1016 characters have been extracted and 27 most compatible trees have been found.

The overall retention index for the intersection tree is 0,848 and Compl = 56,6% (please, refer to Zaragueta Bagils *et al.* 2012 for the meaning of the completeness index Compl.)

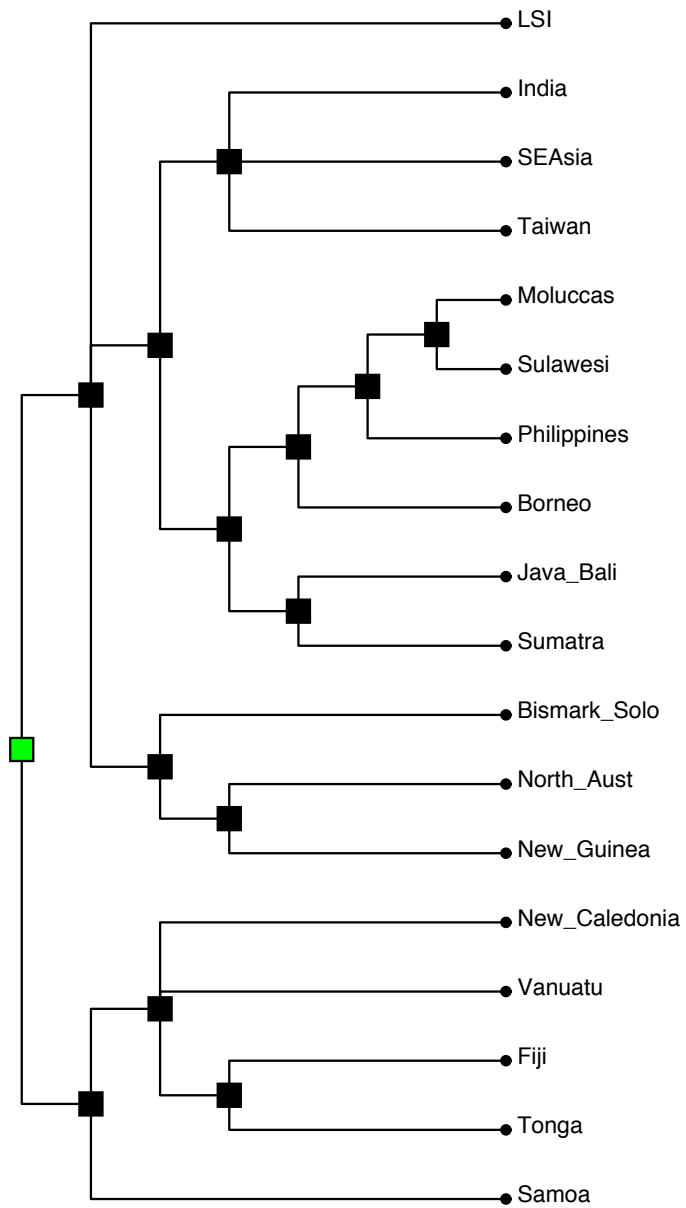


Figure 2: Intersection tree from 27 MCT.

RI= 31895/37607=0,848

Compl=431/762=56,6%

DISCUSSION

Our aim is to look for a general pattern, which is illustrated by the areagram found. A recent paper from Parenti and Ebach (2013) fairly explained that biogeographers should make explicit the difference between “evidence” and “hypothesis”:

“Treating a hypothesis as if it were evidence or an empirical observation confounds many biogeographical analyses” (Parenti and Ebach, 2013: 1).

We conducted our analysis as a way of testing biogeographical hypotheses.

The 3ia analysis yields a quite resolved cladogram of the areas defined in the SEA-WP. The retrieved pattern shows the relationships between each of them. Thirteen nodes are identified, revealing 13 clades (Table 4).

Nodes	Clades
1	A: (((India,SEAsia,Taiwan),((((Moluccas,Sulawesi),Philippines),Borneo),(Java-Bali,Sumatra))),Bismark-Solomons,(North-Aust,New-Guinea)),LSI)
2	B: ((India,SEAsia,Taiwan),((((Moluccas,Sulawesi),Philippines),Borneo),(Java-Bali,Sumatra)))
3	C: (India,SEAsia,Taiwan)
4	D: (((Moluccas,Sulawesi),Philippines),Borneo),(Java-Bali,Sumatra))
5	E: (((Moluccas,Sulawesi),Philippines),Borneo)
6	F: ((Moluccas,Sulawesi),Philippines)
7	G: (Moluccas,Sulawesi)
8	H: (Java-Bali,Sumatra)
9	I: (Bismark_Solomons, (North_Australia, New_Guinea))
10	J: (North_Australia, New_Guinea)
11	K: ((New_Caledonia,Vanuatu,(Fiji,Tonga)),Samoa)
12	L: (New_Caledonia,Vanuatu,(Fiji,Tonga))
13	M: (Fiji,Tonga)

Table 4: Nodes of the intersection tree and the clades revealed

Two major clades (A and K) are found, separating clearly the west pacific (clade K) from Wallacea (i.e. the region inbetween Wallace and Weber’s lines, see Fig. 4) and the rest of Southeast Asia (clade A). Inside the clade A, 2 clades are found with the Lesser Sunda Islands placed at the base (i.e. in polytomy) of this clade; respectively, clade B which groups ((India, SEAsia, Taiwan), (((Moluccas, Sulawesi), Philippines), Borneo),(Java-Bali,

Sumatra))) and clade I ((Bismark_Solomons, (North_Australia, New_Guinea)). Clade B included clade C (India, SEAsia, Taiwan) and clade D (((((Moluccas, Sulawesi), Philippines), Borneo), (Java_Bali, Sumatra))).

Java_Bali is grouped with Sumatra and constitutes the clade H inside clade D; clade F is composed of Borneo and clade E, itself composed of Philippines and clade G (Moluccas and Sulawesi).

A general pattern emerges from the 3ia analysis (Fig. 3). The tree is clearly divided in two parts along a line that could be traced separating the Pacific islands from the rest of Southeast Asia. We have called it the Pacific clade simply in reference to its location. The second part of the tree shows a separation between what we have called the "Australian" clade, i.e. clade I from the rest of the Archipelago. Apart from the unresolved position of the Lesser Sunda Islands (due to the few number of taxa distributed there), Lydekker's (1902) line is clearly identified. It traces "the boundary of the North Moluccas, separating first Timor and Australia and then, between the islands of the Babar and Tanimbar groups, west around Buru and Halmahera and to the Pacific" (van Oosterzee, 1997) (Fig. 4). Five monophyletic clades are distinguishable: a "Pacific" (clade K), an "Australian" (clade I), an "Indonesian" (clade H), a "Southeast Asian" (clade C) and a "Wallacean" (clade G). The latter is closer to the "Indonesian" clade than to the "Southeast Asian". Both of them are included in the clade D which is a sister clade to the "Southeast Asian" clade. The "Australian" clade (clade I) is closer to clade B than to the "Pacific" clade despite its geographical proximity. This is due to the distribution of the taxa.

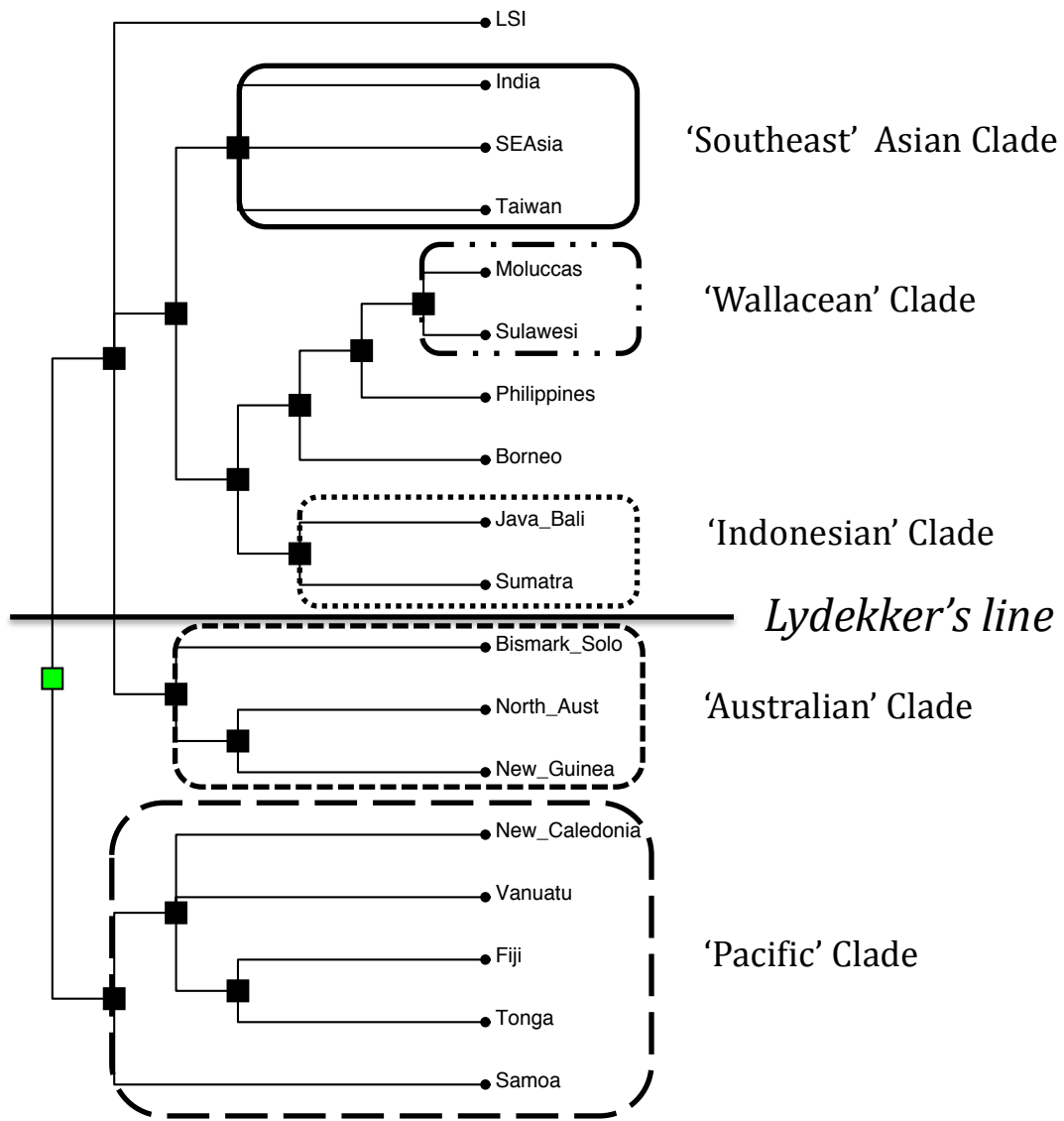


Figure 3: Intersection tree and the clades highlighted

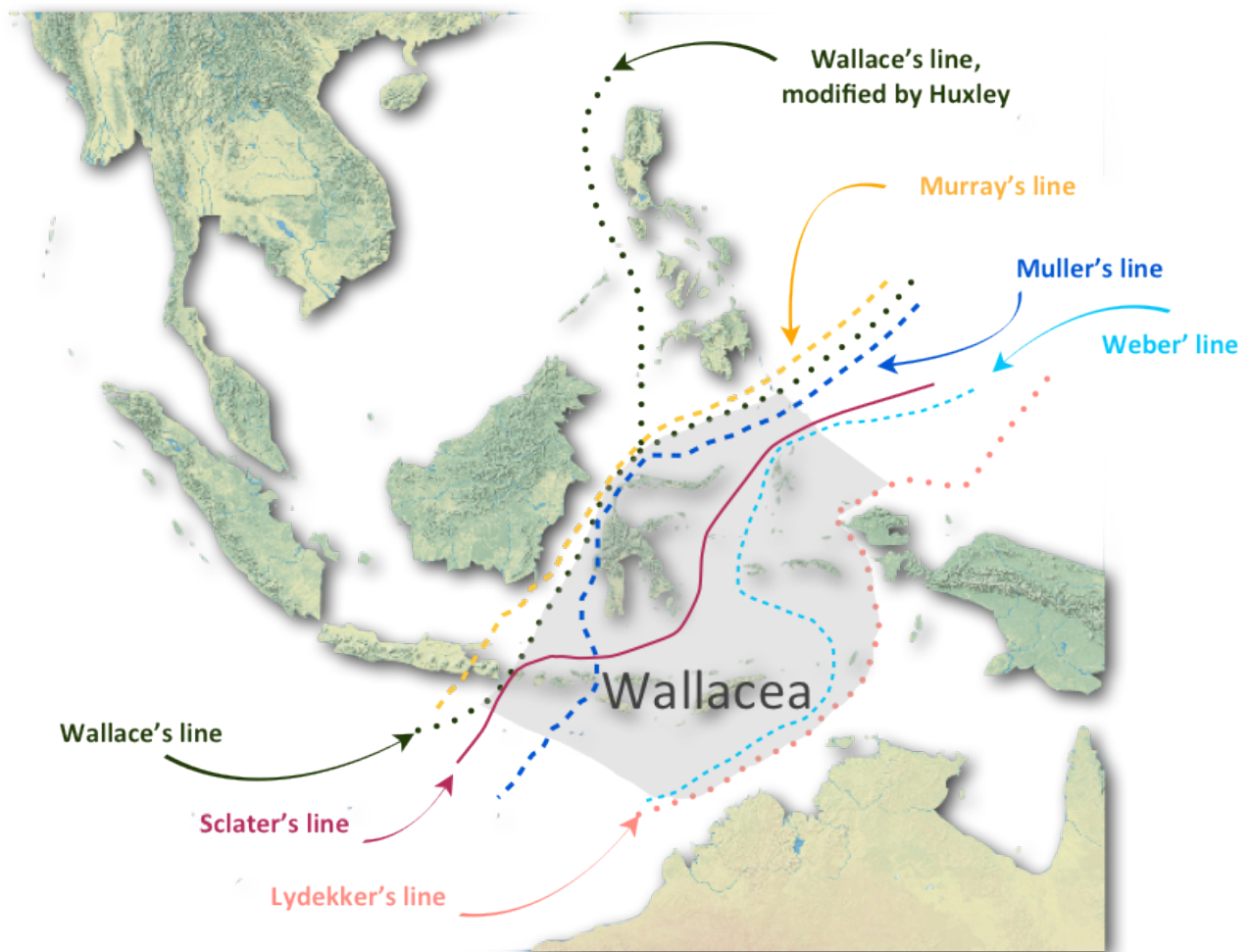


Figure 4: Examples of lines drawn in the Indo-Pacific archipelago based on the distributions of different taxa. Wallace's line (1863) is based on the distribution of birds, mostly parrots and various vertebrates; Huxley's line (1868) is based on the distribution of birds; Sclater's line focussed on birds as well ; Lydekker's line corresponds with the edge of the Australian continental shelf and Weber's line (1902) is related to the distribution of freshwater fishes (van Oosterze 1997)

van Welzen et al in 2003 reassessed their dataset in a unrooted analysis and they gained a result similar to our. They concluded that the pattern retrieved was related more to proximity than to cladistics relatedness. Although their result is similar to our general pattern, we cannot interpret them similarly. With an unrooted tree, nothing

can be said regarding relationships between the areas as there is no root. Nonetheless, their result is still relevant for our purpose since it gives the same results.

Traceability of characters history

The clades are found thanks to specific combination of taxa, e.g. clade G (grouping the Moluccas islands and Sulawesi, i.e. the Wallacean clade) is supported by the combined presence of species from the following taxa: *Megarthus*, *Rhododendron*, *Rhysotoechia*, Chlorocystini, *Xenobates* and *Parkia* (3 plants and 3 animals, precisely insects) (Fig. 5, Table 5).

Clades	Supported by
A	COS, CYR, GAL, GEH, MEG, CUP, ARY, RHY, XEN (9)
B	CYR, GEH, RHO, CAS, HAL, LEP, HALI, XEN, ERI , NEO, PAR (11)
C	RHO, CAS, HAL, GEN, PAR (4)
D	CHLO, MEG, MET, RHO, CAS, HAL, MEGA, LEP, XEN, GEN, NEO, PAR, XAN, HYD, JAC (15)
E	CHLO, RHO, TRO, CAS, RHY, XEN, GEN, NEO, FOR , PAR, XAN, HYD, CRO (13)
F	CHLO, MEG, MET, RHO, TRO, CAS, RHY, XEN, GEN, NEO, PAR, CRO (12)
G	MEG, RHO, RHY, CHL, XEN, PAR (6)
H	RHO, NEO, PAR, XAN (4)
I	ODO , COS, CYC , GAL, RHO, VAR, CAS, CUP, MEGA, ARY, CHL, HALO , HALI (13)
J	COS, CYR, GEH, MEG, VAR, CAS, CUP, ARY, CHL, HALI, XEN (11)
K	COS, CYR, GEH, CUP, HALC (4)
L	GEH, MEG, PIT, CUP, MEGA, ARY, HALC (7)
M	CYR, GEH, PIT, HALC (4)

Table 5: Taxa supporting each nodes of the intersection tree (in bold are indicated taxon that support only one node)

- 1:** COS, CYR, GAL, GEH, MEG, CUP, ARY, RHY, XEN
- 2:** CYR, GEH, RHO, CAS, HAL, LEP, HALI, XEN, **ERI**, NEO, PAR
- 3:** RHO, CAS, HAL, GEN, PAR
- 4:** CHLO, MEG, MET, RHO, CAS, HAL, MEGA, LEP, XEN, GEN, NEO, PAR, XAN, **JAC**
- 5:** CHLO, RHO, TRO, CAS, RHY, XEN, GEN, NEO, **FOR**, PAR, XAN, HYD, CRO
- 6:** CHLO, MEG, MET, RHO, TRO, CAS, RHY, XEN, GEN, NEO, PAR, CRO
- 7:** MEG, RHO, RHY, CHL, XEN, PAR
- 8:** RHO, NEO, PAR, XAN
- 9:** **ODO**, COS, **CYC**, GAL, RHO, VAR, CAS, CUP, MEGA, ARY, CHL, **HALO**, HALI
- 10:** COS, CYR, GEH, MEG, VAR, CAS, CUP, ARY, CHL, HALI, XEN
- 11:** COS, CYR, GEH, CUP, HALC
- 12:** GEH, MEG, PIT, CUP, MEGA, ARY, HALC
- 13:** CYR, GEH, PIT, HALC

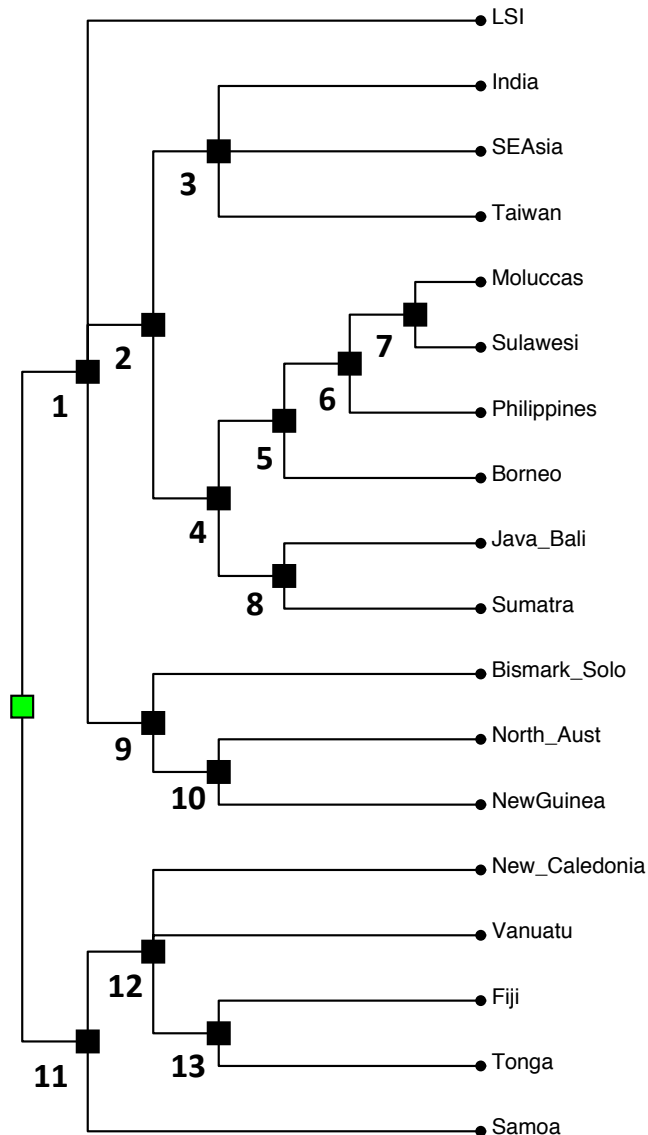


Figure 5: IT and taxa that supported each nodes mapped on the tree

The best-supported nodes are those with the highest number of taxa, which we can qualified as “synapomorphic taxa”. Hence, clade D is the best supported with the presence of 15 taxa; clade I, the “Australian” clade, and clade E are each supported by 13 taxa; 12 taxa support clade F; clades B and J represent the distribution of 11 taxa; clade A shows 9 taxa; clade L has 7 taxa; clade G with 6 taxa and clades C, i.e. the “Southeast asian” clade, K, i.e. the “Pacific” clade, and M are less well supported (compared to the others) with only 4 “synapomorphic” taxa.

Clade K is interesting because, despite the number of taxa distributed over those areas (13: *Arytera*, Chlorocystini, Cosmopsaltriina, *Cupianupsis*, *Cycas*, *Cyrtandra*, *Erimanthus*, *Gallicolumba*, *Gehyra* and *Halobates princeps*, and with a total of 86 species, see Appendix S1 for more details), only five support it: *Cosmopsaltriina*, *Cyrtandra*, *Cupianupsis*, *Gehyra* and *Halobates princeps*.

In Table 5, we have indicated in bold taxa that support only one node: *Erismanthus*, *Dundubia jacoona* assemblage, *Fordia*, whose distributions are unambiguous “synapomorphies” of clades B, D, E. *Caliceniinae*, *Cycas* and *Haloveloides* support clade I. Table 6 shows distributions of taxa and nodes supported.

Taxa	Clades
ARY	A; I; J
CAS	B; C; D; E; F; I; J
CHL	A; G; I; J
CHLO	D; E; F;
COS	A; I; J; K
CRO	E; F
CUP	A; I; J; K; L
CYC	I
CYR	A; B; J; K; M
ERI	B
FOR	E
GAL	A; I
GEH	A; B; J; K; L; M
GEN	C; D; E; F
HAL	B; C; D
HALC	K; L; M
HALI	B; I; J
HALO	I
HYD	D; E
JAC	D
LEP	B; D
MEG	A; D; F; G; J; L
MEGA	D; I; L
MET	D; F
NEO	B; D; E; F; H
ODO	I

PAR	B; C; D; E; F; G; H
PIT	L; M
RHO	B; C; D; E; F; G; H; I
RHY	E; F; G
TRO	E; F
VAR	I; J
XAN	D; E; H;
XEN	A; B; D; E; F; G; J;

Table 6: Taxa and clades supported

Fig. 6 illustrates their distributions to which we add the *Varanus* distribution because it supports clades J and I (J being included in I). It is noteworthy that although these taxa are broadly distributed over the region, each distribution supports only one area, *Erismanthus* appears as a “synapomorphy” of area B, the *Dundubia jaccona* assemblage of area D, *Fordia* distribution supports area E, and I has the distributions of Chlorocyphidae, *Cycas* and *Halobates* as unambiguous “synapomorphies”. These results highlight the fact that, in spite of what seems complex biogeographical relationships, the distributions of these taxa may be assigned to a single area. The identification of distribution of taxa as “synapomorphies” of biogeographical areas allows focusing the discussion on the pattern of relationships.

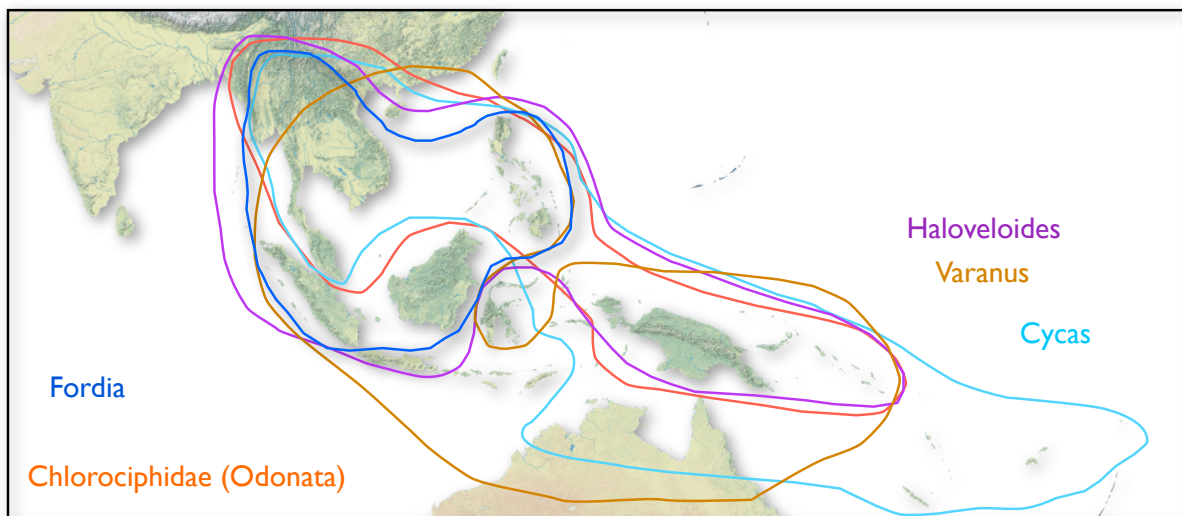


Figure 6: Distribution of taxa supporting only one clade

Comparison with the geology

Exploring cladistics relationships between areas of endemism leads to a consideration of the geological history of a region of interest. In our case, regarding SEA-WP, different interpretations regarding its evolution because of its complex geology were discovered (e.g. Hamilton 1979; Holloway 1979; Duffels 1986; Hall 1996, 1997, 1998, 2002, 2011 among others). Some of the complexity is captured by Hall:

“...it is clear from the geology of the region that the snapshot we see today is no less complicated than in the past. The region has developed by the interaction of major lithospheric plates, principally those of the Pacific, India-Australia and Eurasia, but at the present day a description only in terms of these three plates is a very great oversimplification.” (Hall 1998: 99)

“ The abrupt division between faunas and floras in Indonesia first recognized by Wallace in the nineteenth century, has its origin in the rapid plate movements and reorganisation of land-masses in SE Asia” (Hall 1998: 100)

According to Hall, therefore, much of the evidence that must be used in a regional tectonic model of SE Asia is based on the *interpretation* of geological data from the small ocean basins, their margins, and from the geologically more complicated land areas around them. He goes on to note:

“The reader should be aware that, as in other areas of science, geologists differ in their interpretations of these data, and much of the information does not lend itself to unambiguous reconstruction. Nonetheless, a complete tectonic history can only be deduced from the geology on land combined with data from the oceans.” (Hall 1998: 105)

In another paper from the same book, Holloway and Hall (1998) claim that to a geologist the dismissal of the role of dispersal is odd. Judging from the geological history of SE Asia it seems highly unlikely that any understanding of the biogeographic patterns can be achieved without considering both vicariance and dispersal. Certainly, the Cenozoic development of the region is characterised more by amalgamation than fragmentation. Nevertheless, this reasoning misses that the source of relationships of areas is based on relationships of distribution of taxa. These are tree-like independently of the geological history of the region. Moreover, a tree of areas derived from biological evidence could reasonably be interpreted as: the fragmentation of a previously continuous land area such as Gondwanaland, or Pleistocene division of Sundaland through marine transgression (Ruedi 1995); the approach and accretion of terranes (and thus dispersal to) onto a larger land mass; or the slow dispersal of organisms, with speciation, through an archipelago with stable geography.

The Gondwana origins of all component continental blocks of SE Asia, i.e. the core of Sundaland, is now widely accepted (Hall 2009, 2011, 2012; Hall et al. 2011; Metcalfe 2011). Sundaland was assembled from continental

blocks that separated from Gondwana in the Paleozoic and amalgamated with Asian blocks in the Triassic (Hall 2011; Hall et al. 2011; Metcalfe 2011). The first event that shaped the region (about 180 Ma ago) is the break-up of the Gondwana which led to the separation of India and Australia-Antarctica from Africa. The Indian subcontinent moved northward towards Asia and collided later (at about 50 Ma ago) (de Boer and Duffels 1996 and references herein). Australia, which had become separated from Antarctica by the opening of the Tasman Sea (95 Ma ago), changed its course northwards (de Boer and Duffels 1996). The floor of the Thethys sea was forced to subducted under the Pacific plate which gave rise to a volcanic island arc (the West Pacific island arc, also named the Outer Melanesian arc by Duffels (1986) and Holloway (1979) although not entirely recognized as the same (Michaux 1994; de Boer and Duffels 1996) and simply the Melanesian arc system by Hall (2002)). Remnants of this West Pacific island arc have been recognized and are as follows: the central Philippines, northern, central, and southeastern New Guinea, and the Bismarck Archipelago. The collision between the Pacific plate and the Asian continent must have occurred about 40-42 Ma ago and caused the fracture of the West Pacific island with its northern western part rotating clockwise which made the central Philippines collide with the continental western Philippines (Rangin *et al.* 1990a, b; Daly *et al.* 1991; Honza 1991).

To the east, the continuous South-West Pacific island arc (also known as the East Melanesian arc) is composed of current Vanuatu linked to Solomon and Fiji (de Boer 1995d and references therein). At about 9-12 Ma ago, it collided with the Australian continent in the Solomon area and simultaneously in the New Guinea area, which makes it broken up giving rise to Vanuatu, Fiji and the Tonga-Kermode (de Boer 1995d). By 3 Ma ago Fiji was totally isolated. The palaeogeographic reconstruction presented here can be summarized in the “cladogram-like” graph for the West Pacific island arc (Fig. 7).

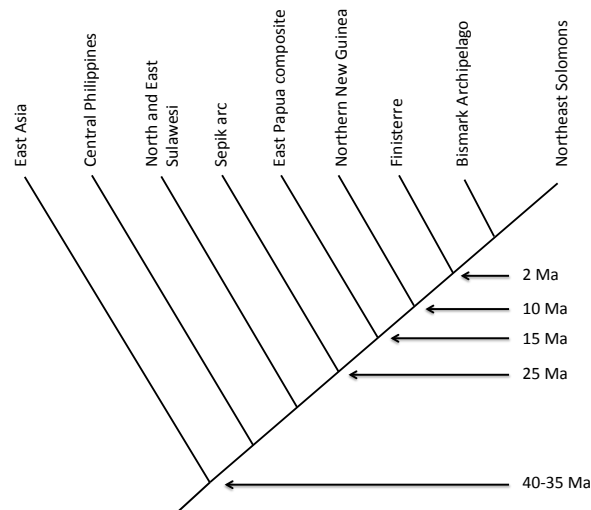


Figure 7: Geological area cladogram showing the fragmentation sequences in the West Pacific island arc and ages of nodes redrawn after de Boer, 1995d.

Some geological events can be drawn from examination of the geological area cladogram (Fig.7) and the intersection tree (Fig. 2). The main one is the vicariant event that separated the areas emerged from the East Melanesian Arc (sensu de Boer, 1995d).

That scheme is consistent with our areagram. Node 11 of Fig. 2 shows that after the first vicariant event that has separated Samoa, successive events may be reconstructed: Vanuatu and New Caledonia, then Fiji and Tonga.. Furthermore, this clade is revealed by the presence of only four taxa: *Cosmopsaltriina*, *Cyrtandra*, *Gehyra*, *Cupaniopsis* and *Halobates princeps* whose dispersal abilities varied a lot from one to another. Hence, we could hypothesize that those taxa were present on the Melanesian arc and that its fracture led to this distribution.

Nonetheless, the rest of the general areagram suggest relationships emerging from geographical proximity (clade I, clade C, clade H and clade G). Thus, dispersal events would be a more probable explanation despite the different dispersal capabilities of the taxa under study. However, one has to keep in mind that the areagram shows relationships between areas of endemism that are *biogeographic* areas and not geographic areas even if they are given names from geography. Many islands in that region should be considered as separate fragments because they are geologically composite. For instance Sulawesi should be separated in three distinct areas: north, west and south (Hall 1996; Michaux 2010; Hall 2011), New Guinea in at least two areas, north and south (Hall 2002). The main impediment remains the availability of precise distributional data since most published cladograms of taxa do not present accurate distributions.

Cladistic biogeography

Biogeography and geology are often opposed, although these two disciplines are definitively complementary (Holloway and Hall 1998; Hall 2002, 2011; Michaux 2009). If there is a contradiction between them, then one discipline should not prevail. Congruence between geological reconstructing/modelling history and biogeographical reconstruction of relationships between areas of endemism emphasize the fact that both can indeed be true.

From 35 different taxa, including both animals and plants, we have shown that a pattern does emerge from the analysis and, to a certain extent, fits with the most common geological reconstructions of SEA-WP evolution (Hall 2002, 2011, Hall et al. 2011; Metcalfe 2011). This result emphasizes the accuracy of the 3ia method in revealing a biogeographical pattern that does not only consider dispersal as the main process of organism distribution.

CONCLUSIONS

Our analysis indicates that a vicariant event separated Fiji, Tonga, Vanuatu, New Caledonia and Samoa from the rest, which corroborates the results of de Boer and Duffels who examined the historical biogeography of the cicadas of Wallacea, New Guinea and the West pacific (de Boer and Duffels, 1996). They compared each cicada area-cladogram with the geological cladogram (Fig. 7) and concluded that:

“from the high degree of congruence between the geological cladogram and the two taxon-area cladograms we conclude it is highly probable that vicariance caused by the fragmentation of a West and a South-West Pacific island arc is responsible for most of the present-day generic diversification” (de Boer and Duffels 1996:174).

They intended at first to use the cladistic method of Platnick and Nelson (1978) and Humphries and Parenti (1986), i.e. combining taxon-area cladograms based on several unrelated groups into a general area cladogram. Nevertheless, they considered that the area cladograms for the two groups of cicadas provided an insufficient basis for the construction of such a general cladogram. We agree: congruence among two cladograms may provide the weakest support. Indeed, their general argument strengthens our case, as we combined 35 different taxon cladograms from both plants and animals and obtained the general areagram shown in Fig. 2 where relationships between the areas of endemism can be discussed in term of combination of taxa that supported the nodes of the tree. Our analysis has recovered five « monophyletic » clades : “Pacific” (clade K), “Australian” (clade I), “Indonesian” (clade H), “Southeast asian” (clade C) and “Wallacean” (clade G) supported by a certain assemblage of taxa. Furthermore, Weber’s Line is revealed which confirms that drawing lines is informative.

Systematic biogeography is similar to phylogenetic systematics in that proposal of area relationships to be tested and accepted or rejected in an effort to form relevant area classifications (Parenti and Ebach, 2009). 3ia is a straightforward method that serves historical biogeographic purposes. The hierarchical representation of knowledge available in LisBeth authorise new discussions in term of relationships between areas towards a natural classification.

ACKNOWLEDGEMENTS

Visotheary UNG is grateful to Peter van Welzen for reviewing a draft of the manuscript and for fruitful discussions about the biogeography of the region.

The authors would like to thank several anonymous referees for constructive critiques of the manuscript.

This research is supported by a grant (I3iaSEAB, AAP-IN-2009-033) from the Foundation for Research on Biodiversity (FRB).

REFERENCES

Andersen, N.M. (1998) Marine water striders (Heteroptera, Gerromorpha) of the Indo-Pacific: cladistic biogeography and Cenozoic paleogeography. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 341-354. Backhuys Publishers, Leiden.

Ast J. C. (2001) Mitochondrial DNA evidence and evolution in Varanoidea (Squamata). *Cladistics*, **17**, 211-226.

- Axelius, B. (1991) Areas of distribution and areas of endemism. *Cladistics*, **7**(2),197-199.
- Birks, S. M., and Edwards S. V. (2002) A phylogeny of the megapodes (Aves: Megapodiidae) based on nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **23**, 408-421.
- Brown G. K., Nelson G. and Ladifés P. (2006) Historical biogeography of *Rhododendron* section *Vireya* and the Malesian Archipelago. *Journal of Biogeography*, **33**, 1929-1944.
- Cao N. (2008) Analyse à trois éléments et anatomie du bois des Fagales, Engl. Thèse de doctorat. Université Pierre et Marie Curie.
- Cao N., Zaragüeta-Bagils, R. & Vignes-Lebbe R. (2007) Hierarchical representation of the hypotheses of homology. *Geodiversitas*, **29** (1) : 5-15.
- Cao, N., El Azawi, M. and Zaragüeta-Bagils, R. (2009) Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example, *Bulletin de la Société Géologique de France*, **180**, 13-15.
- Cronk, Q. C. B., Khien M., Wagner, W. L. and Smith J. F. (2005) Evolution of *Cyrtandra* (Gesneriaceae) in the Pacific Ocean: the origin of a supertramp clade. *American Journal of Botany*, **92** (6), 1017-1024.
- Daly, M. C., Cooper, M. A., Wilson, I., Smith, D. G. and Hooper, B. G. D. (1991) Cenozoic plate tectonics and basin evolution in Indonesia. *Marine and Petroleum Geology*, **8**, 2-21.
- de Boer, A. J. (1995) Islands and cicadas adrift in the West-Pacific. Biogeographic patterns related to plate tectonics. *Tijdschrift Voor Entomologie* **138**: 169-244.
- de Boer, A. J. and J. P. Duffels (1996) Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **124**(1-2), 153-177.
- Dekker R. W. R. J. (2007) Distribution and speciation of Megapodes (Megapodiidae) and subsequent development of their breeding behaviour. In: *Biogeography, Time, and Space: Distributions, Barriers, and Islands*, 45-91. W. Renema (eds) Springer.
- Duffels, J. P. (1986) Biogeography of indo-pacific cicadoidea: a tentative recognition of areas of endemism. *Cladistics*, **2**, 318-336.
- Duffels, J.P. and Turner, H. (2002) Cladistic analysis and biogeography of the cicadas of the Indo-Pacific subtribe *Cosmopsaltriina* (Hemiptera: Cicadoidea: Cicadidae). *Systematic Entomology*, **27**, 235-261.
- Ebach M. C., Humphries C. J., Newman R. A., Williams D. M. and Walsh S. A. (2005) Assumption 2: opaque to intuition? *Journal of Biogeography*, **32**, 781-787.
- Ewart, A. (1988) Geological history of the Fiji-Tonga-Samoan region of the S.W. Pacific and some palaeogeographic and biogeographic implications. In: J.P. Duffels, *The Cicadas of Fiji Samoa and Tonga Islands, their Taxonomy and Biogeography* (Homoptera, Cicadoidea). *Entomograph*, **10**, 15-23.
- Gasmann, D. (2005) The phylogeny of Southeast Asian and Indo-Pacific Calicnemiinae (Odonata, Platycnemididae). *Bonner Zoologische Beiträge*, **53**, 37-80.
- Gemmill, C. E. C., Allan G. J., Wagner W. L. and Zimmer E. A. (2002) Evolution of insular pacific *Pittosporum* (Pittosporaceae): origin of the Hawaiian radiation. *Molecular Phylogenetics and Evolution*, **22** (1), 1-42.

- Hall, R. (1996) Reconstructing Cenozoic SE Asia in *Tectonic Evolution of SE Asia*. Hall, R. & Blundell, D. (eds.). 106. p. 153-184. Geological Society of London Special Publication.
- Hall, R. (1997) Cenozoic plate tectonic reconstruction of SE Asia. In Fraser A.J., Matthews S.J. & Murphy R.W. (eds) *Petroleum geology of Southeast Asia*. Geological Society of London Special Publication, **26**, 11-23.
- Hall, R. (1998) The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. *Biogeography and Geological evolution of SE Asia*, pp.99-131.
- Hall R. (2002) Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, **20**, 353-434.
- Hall, R (2009) Southeast Asia's changing palaeogeography. *Blumea: Journal of Plant Taxonomy and Plant Geography*, **54**, No. 1-3, 148-161.
- Hall, R. (2011) Australia-SE Asia collision: plate tectonics and crustal flow. The SE Asian gateway: history and tectonics of Australia-Asia collision (eds R. Hall; M. A. Cottam; M. E. J. Wilson), Geological Society, London, Special publications, **355**, 75-109.
- Hall, R. (2012) Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In: *Biotic Evolution and Environmental Change in Southeast Asia*. ed. / D. J. Gower; J.E. Richardson; B. R. Rosen; L. Rüber; S. T. Williams. Cambridge University Press, 2012. p. 32-78.
- Hall R., Cottam M.A. and Wilson M. E. J. (2011) The SE Asian gateway: history and tectonics of the Australian-Asia collision. (eds R. Hall; M.A. Cottam; M.E.J. Wilson), Geological Society, London, Special publications, **355**, 1-6.
- Hamilton, W. (1979) Tectonics of the Indonesian region. *United States Geological Survey Professional Paper*, p. 1078
- Heinicke, M. P., Greenbaum E., Jackman T. R. and Bauer A. M. (2011) Phylogeny of a trans-Wallacean radiation (Squamata, Gekkonidae, Gehyra) supports a single early colonization of Australia. *Zoologica Scripta*, **40**, 584-602
- Holloway, J. D. (1979) A survey of the Lepidoptera, Biogeography and Ecology of New Caledonia. Series Entomologica 15. The Hague: W. Junk.
- Holloway, J. D. and Hall, R. (1998) SE Asian geology and biogeography: an introduction. In: *Biogeography and Geological Evolution of SE Asia*. ed. / R. Hall ; J.D. Holloway. Leiden: Backhuys Publishers, p. 1-23.
- Honda, M., Ota H., Murphy R. W. and Hikida T. (2006) Phylogeny and biogeography of water skinks of the genus *Tropidophorus* (Reptilia: Scincidae): a molecular approach. *Zoologica Scripta*, **35**, 2006, 85-95
- Honza, E. (1991) The Tertiary arc chain in the western Pacific. *Tectonophysics*, **187**, 285-303
- Humphries, C. J., and Parenti, L. R. (1986) Cladistic biogeography. Oxford monographs on biogeography no. 2. Clarendon Press, Oxford.
- Jønsson, K. A. Iresredt M. Bowie A. C. K. Christidis L. and Fjeldså (2011) Systematics and biogeography of Indo-Pacific ground-doves. *Molec. Phylog. Evol.* **59**, 538-543.
- Keppel G., Hodgskiss P. D. and Plunkett G. M. (2008) Cycas in the insular South-west Pacific: dispersal or vicariance? *Journal of Biogeography*, **35**: 1004-1015.

- Kluge, A. G., and Farris, J. S (1969) Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **18**, 1-32.
- Metcalf I. (2011) Paleozoic-Mesozoic history of SE Asia. The SE Asian gateway: history and tectonics of Australia-Asia collision (eds R. Hall; M.A. Cottam; M.E.J. Wilson), Geological Society, London, Special publications, **355**, 7-35.
- Michaux, B. (1994) Land movements and animal distributions in east Wallacea (eastern Indonesia, Papua New Guinea and Melanesia). *Palaeogeography, Paleoclimatology, Paleoecology*, **112**, 323-343.
- Michaux, B. (2009) Reciprocity between biology and geology: Reconstructing polar Gondwana. *Gondwana research*, **16**, 655-668.
- Michaux, B. (2010) Biogeology of Wallacea: geotectonic models, areas of endemism, and natural biogeographical units. *Biological Journal of the Linnean Society*, **101**, 193-212.
- Nelson, G., Atz, J.W., Kallman, K.D. & Smith, C.L. (1987) Donn Eric Rosen, 1929-1986 *Copeia*, 541-547.
- Nelson, G. J. and Ladiges, P. Y. (1991a) Standard assumptions for biogeographic analysis, *Australian Systematic Botany*, **4**, 41-58.
- Nelson, G. J. and Ladiges, P. Y. (1991b) Three-area statements: Standard assumptions for biogeographic analysis, *Systematic Zoology*, **40**, 470-485.
- Nelson, G. J. and Ladiges, P. Y. (1992) Information Content and Fractional Weight of Three-Item Statements, *Systematic Biology*, **41**, 490-494.
- Nelson, G. J. and Ladiges, P. Y. (1996) Paralogy in Cladistic Biogeography and Analysis of Paralogy-Free Subtrees, *Am. Mus. Novit.*, **3167**, 1-58.
- Nelson, G. and Platnick, N. I. (1981) Systematics and Biogeography. Cladistics and Vicariance. Columbia University Press, New York.
- Nelson, G. J. and Platnick, N. I. (1991) Three-taxon statements: A more precise use of parsimony? *Cladistics*, **7**, 351-366.
- Parenti, L. R. (2006) Common cause and historical biogeography. In Ebach, M. C. and Tangney, R. (eds.) Biogeography in a Changing World. Fifth Biennial Conference of the Systematics Association, UK. CRC Press: pp. 61-82
- Parenti, L. R. and Ebach, M. C. (2009) Comparative Biogeography: Discovering and Classifying Biogeographical Patterns of a Dynamic Earth. University of California Press, Berkeley.
- Parenti, L. R. and Ebach, M. C. (2010) Wallacea Deconstructed. In D.M. Williams & S. Knapp (eds.), Beyond Cladistics: The Branching of a Paradigm. University of California Press, Berkeley, pp. 303-318.
- Parenti, L. R. and Ebach M. C. (2013) Evidence and hypothesis in biogeography. Guest editorial. *Journal of Biogeography*, **40**, (5), 813-820.
- Pigram, C. J. and Davies, P. J. (1987) Terranes and the accretion history of the New Guinea orogen. B.M.R.J. Aust. Geol. Geophys., **10**, 193-212.
- Platnick, N. I. & Nelson, G. (1978) A method of analysis for historical biogeography. *Systematic Zoology*, **27**, 1-16.

- Rangin, C., Jolivet, L. and Pubellier, M. (1990a) A simple model for the tectonic evolution of southeast Asia and Indonesia region for the past 43 Ma. *Bull. Geol. Soc. France* **8**, **VI**, 889-905.
- Rangin, G., Pubellier, M., Azema, J., Briais, A., Chotin, P., Fontaine, H., Huchon, P., Jolivet, L., Maury, R., Muller, C., Rampnoux, J.-P., Stephan, J.-F., Tournon, J., Cottreau, N., Dercourt, J. and Ricou, L. E., (1990b) The quest for Tethys in the western Pacific. 8 palaeogeodynamic maps for Cenozoic time. *Bull. Soc. Geol. Fr.*, **(8)** 6, 907-913.
- Riddle, B. R. and Hafner, D. J. (2010) Integrating pattern with process at biogeographic boundaries: the legacy of Wallace. *Ecography*, **33**, 321-325.
- Rosen, D. E. (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*, **27**, 159-188.
- Ruedi, M. (1995) Taxonomic revision of shrews of the genus *Crocidura* from the Sunda Shelf and Sulawesi with description of two new species (Mammalia: Soricidae). *Zool. J. Linn. Soc.*, **115**, 211-265.
- Turner H., Hovenkamp P. and van Welzen P. C. (2001) Biogeography of Southeast Asia and the West Pacific. *Journal of Biogeography*, **28**, 217-230.
- Ung V. (2013). Biotic Evolution and Environmental Change in Southeast Asia. -- Edited by David Gower, Kenneth Johnson, James Richardson, Brian Rosen, Lukas Ruber and Suzanne Williams. *Systematic Biology*, Book review (in press).
- van Oosterzee P. (1997) Where worlds collide: the Wallace line. Ithaca and London: Cornell University Press, xiv + 234 pp.
- van Tol, J. (1998) The Odonata of Sulawesi and adjacent islands. Part 4. A new genus and species of Chlorocyphidae from South-East Sulawesi, *Zoologische Verhandlungen* **323**, 441-448.
- van Tol, J. and Gasmann D. (2007) Zoogeography of freshwater invertebrates of Southeast Asia, with special reference to Odonata. In: *Biogeography, Time, and Space: Distributions, Barriers, and Islands*, 45-91. W. Renema (eds) Springer.
- Weber M. C. W. (1902) *Der Indo-Australische Archipel und die Geschichte Seiner Tierwelt*, Gustav Fischer, Jena, 46 p.
- Woodruff, David S. (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and future of the remaining refugial-phase biodiversity. *Biodivers Conserv*, **19**, 919-941.
- Zaragueta-Bagils R., Ung V., Grand A., Vignes-Lebbe R., Cao N. and Ducasse J. (2012) LisBeth: new cladistics for phylogenetics and biogeography. *C. R. Palevol*, **11**, 563-566.

SUPPORTING INFORMATION

Appendix S1: Distributions of all taxa in the corresponding areas.

Appendix S2: Algorithm for traceability of characters.

BIOSKETCHES

Visotheary Ung's research focuses on historical biogeography of Southeast Asia and integrated temporal framework in cladistic biogeography through diachronic events analysis. Her PhD is supported by a grant from the French Foundation of Research on Biodiversity (FRB).

René Zaragüeta Bagils's research concerns the theory of cladistic phylogenetics and biogeography and its philosophical implications, particularly interested in tree-item analysis.

David Williams' research includes systematics and biogeography of diatoms, and the history, philosophy and theories of comparative biology, particularly cladistics (three-item analysis).

Author contributions: VU selected the data, performed the analysis and led the writing, DMW had the original idea, made corrections and style improvement to the original text, RZB contributed to the design of the analysis.

Editor: Malte Ebach

Appendix S1

Distribution of species

	ARY	CAS	CHL	CHLO	COS	CRO	CUP	CYC
India	1	2		1	5		1	1
Bismark Archipelago and the Solomons	1	1						
Lesser Sunda Islands	1							
North Australia	9	4		7	1		12	1
New Guinea	10	3		20	8		11	1
The Moluccas	1			3	6		1	
Sulawesi	1	1	3	1	2	6	2	
The Philippines	1	2	1			3		1
Borneo	1	3	3			2		
Java and Bali	1	2	1			3		
Sumatra	1	3	2			5		
Southeast asia	1	3	2	1		4		1
Taiwan		2						
New Caledonia	6						27	1
Vanuatu	2			1	1		1	1
Fiji	1				1		4	1
Tonga	1			1	1			1
Samoa	1			1	1		1	
Total	40	24	12	36	26	23	60	9

Appendix S1

	CYR	ERI	FOR	GAL	GEN	GER	HAL
India					3		1
Bismark Archipelago and the Solomons				3			1
Lesser Sunda Islands							
North Australia	1			4		44	1
New Guinea	1	2				9	
The Moluccas						3	
Sulawesi				2	1		1
The Philippines	4		1	4	2	2	
Borneo	1	3	3		4		
Java and Bali	1				3		1
Sumatra		3	2				
Southeast asia	4	2	12		10	7	2
Taiwan	1						
New Caledonia						2	
Vanuatu						2	
Fiji	1			1		2	
Tonga	2			1		2	
Samoa	1			1		2	
Total	16	10	18	16	23	75	7

Appendix S1

	HALC	HALI	HALO	HYD	JAC	LEP
India	1	2	2	2	2	1
Bismark Archipelago and the Solomons	1					1
Lesser Sunda Islands	1	6				
North Australia	1	3	2			1
New Guinea	1		1			4
The Moluccas	1		1	1		
Sulawesi	1	1	3	1		1
The Philippines			1	5	2	
Borneo	1		1	1		
Java and Bali			1	1		
Sumatra			1	1	1	2
Southeast asia	1	1	1	3	4	
Taiwan		1	1	1		
New Caledonia	1					
Vanuatu	1					
Fiji	1					
Tonga	1					
Samoa	1					
Total	13	13	13	14	9	10

Appendix S1

	MEG	MEGA	MET	NEO	ODO	PAR	PIT
India						3	
Bismark Archipelago and the Solomons		1		1	1		
Lesser Sunda Islands						1	1
North Australia	5		2				5
New Guinea	10	5	5	1	7	2	
The Moluccas	6					1	
Sulawesi	2		23	1		2	
The Philippines	1	1	4	2	2	1	
Borneo			4	5		4	
Java and Bali	1		1	1		2	
Sumatra	1	1	1	3		4	
Southeast asia			1	4	2	4	
Taiwan			2				
New Caledonia		4					5
Vanuatu	1						
Fiji		1					1
Tonga	1	1					2
Samoa							
Total	28	14	41	18	12	24	14

Appendix S1

	RHO	RHY	TRO	VAR	XAN	XEN	Total
India						1	16
Bismark Archipelago and the Solomons	3			1		1	28
Lesser Sunda Islands							3
North Australia	27	6		32	4	2	139
New Guinea	3	8		1		2	153
The Moluccas	10	1				3	30
Sulawesi	6	2	2			3	68
The Philippines	8	1	3		1	2	51
Borneo	6	3	2		18	2	75
Java and Bali	8				1		26
Sumatra	6				2	1	42
Southeast asia	1		14	17	4	2	113
Taiwan							7
New Caledonia						1	47
Vanuatu							10
Fiji					1		15
Tonga							14
Samoa							9
Total	77	21	21	51	31	20	839

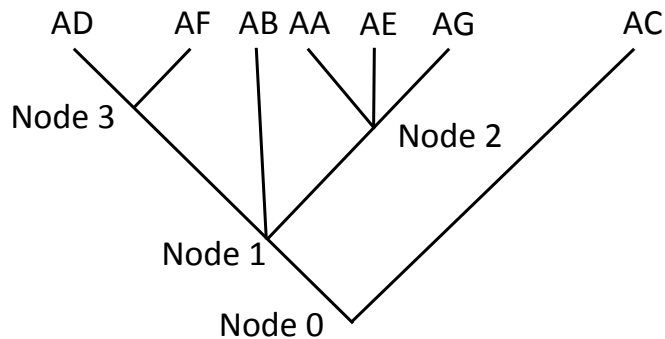
APPENDIX S2

Algorithm for traceability of the characters

LisBeth applies the paralogy-free sub-tree analysis to remove paralogous nodes, and the « transparent method » to remove redundant information due to widespread taxa.

1. For each source cladograms of taxa,
Paralogy-free subtree analysis -> paralogy-free TACs;
Then, Transparent method -> paralogy and MAST-free TACs
Store in file .3ia
2. From .3ia file:
For each paralogy- and MAST-free TAC,
Calculation of 3is and their FW (important for computing of the “compatible trees”)
3. Calculation of the most compatible trees with the 3is list, and reconstruction of the intersection tree;
Store in file .3iz;
4. Creation of the list L0 of all the distinct 3is of the intersection tree, each with its “associated” node.
The “associated” node of the 3is (A (B C)) is the most distal node including B and C but not A. (*)

For instance, in the tree below, the 3is (AD (AA AE)) is associated to Node 2.



Also, each 3is in L0 is associated with an (initially empty) list of source cladograms.

5. In the .3ia file

For each source cladograms of taxa,

all paralogy- and MAST-free TACs are listed;

For each paralogy- and MAST-free TACs,

all 3is are calculated

For each 3is:

If the 3is is present in L0, then the source cladogram is added in the list of source cladograms associated with the 3is in L0.

If the 3is is not present in L0, it is simply ignored.

Then, at the end of this step, each 3is in L0 (that is each 3is of the intersection tree) is associated:

- with a node of the intersection tree,

- and with the list of the source cladograms of taxa which generated a TAC including this 3is.

6. Finally, the relation is reversed: for each node in the intersection tree, we collect all the 3is in L0 associated with this node, then all the source cladograms associated with these 3is. Thus, each node of the intersection tree is associated with a list of source cladograms of taxa ; this list is made non redundant.

En résumé:

1 intersection tree -> N nodes.

For each node -> N 3is.

For each 3is -> N paralogy- and MAST-free TAC containing this 3is.

For each paralogy- and MAST-free TAC -> 1 source cladogram of taxa.

Conclusion et perspectives

1. Conclusions

Nous avons pu mettre en application, pour la première fois, les principes de la 3ia sur un échantillonnage assez large dont les résultats ont mis en évidence l'existence de 5 « clades monophylétiques » : « Pacifique », « Australien », « Indonésien », « Sud-Est asiatique » et « Wallacéen » qui sont soutenus par la combinaison particulière de taxons, que nous appelons « taxons synapomorphiques ». Seule la ligne de Lydekker est retrouvée, le « clade Australien » est plus proche de l'ensemble formé par les « clades Sud-Est asiatique, Indonésien et Wallacéen » qu'il ne l'est du « clade du Pacifique » alors qu'on aurait pu le penser si l'on s'en tient à leur proximité géographique. En revanche, nos résultats montrent clairement que des événements de vicariance ont eu lieu et séparés Fidji, Tonga, Vanuatu, la Nouvelle-Calédonie du reste (des aires étudiées), ce qui corrobore les résultats de De Boer et Duffels qui ont examinés l'histoire biogéographique des cicadas de Wallacea, Nouvelle-Guinée et du Pacifique Ouest (de Boer et Duffels 1996).

Bien entendu, nous devons réserver au sujet de ces inférences qui ne sont que préliminaires et qui doivent être étoffés par des analyses complémentaires, avec plus de cladogrammes de taxons et aussi, sans doute une rediscussion des aires d'endémisme à considérer.

2. Perspectives

Si nous n'avons pu explorer plus avant l'analyse biogéographique de cette région nous avons, cependant, identifié certaines pistes qui mériteraient d'être poursuivies. C'est ce que nous proposons de faire dans les paragraphes suivants.

2.1 Sélection de taxons selon leur âge

Une des premières pistes à suivre consisterait à faire une sélection des taxons à analyser d'une façon un peu plus ciblée.

Tenir compte de leur âge nous permettrait d'être sûr qu'ils aient eu à faire face à des événements de vicariance s'ils sont assez anciens. Penny van Oosterzee souligne que « *the first step in the search for general biogeographic patterns, then, is the search for an appropriate animal and plant group that will unambiguously reveal past geologic and climatic changes* ». Elle continue en proposant le groupe des insectes comme groupe le plus adéquat « *In a story of drifting armada-like continents and magical faunal boundaries, which group could be more appropriate*

than 'looking-glass' insects to provide a finely tuned mirror to the scene : they are ancient, have a worldwide distribution, are highly endemic and are poor dispersers. As such, they mirror geological history ».

Effectivement, on trouve dans la littérature d'autres références présentant les insectes comme de parfaits candidats pour les desseins de la Biogéographie. Ainsi, selon van Tol et Gassman, les Odonates présentent un registre fossile débutant au Permien et sont donc pertinent pour la recherche de pattern généraux « *Most orders of aquatic insects have a fossil record starting in the early Permian, or Middle carboniferous (Odonata), making even the break-up of Gondwana (late Jurassic) relevant to understanding present distributional patterns* » (2007 : 45). Ils citent également d'autres groupes intéressants pour ce genre d'étude « *Many orders of aquatic insects, or at least those with aquatic larval stages, are known in the fossil record as early as the Early Permian, namely in the terminology of Rasnitsyn and Quicke (2002), Ephemera, Hemiptera, Coleoptera, Neuroptera and Trichoptera. The earliest fossils are those of Libellulida (i.e. Odonata) from the middle Carboniferous (c. 325 Ma)* » (**van Tol et Gassman, 2007 : 50**).

Du côté des plantes, nous ne sommes pas en reste non plus, Milne en 2004, estima que le sous-genre Rhododendron est âgé de -46 -32 Ma suggérant de la sorte que l'origine du genre *Vireya* est plus récente. Un chronogramme plus récent publié par Alex Twyford et James Richardson utilisant le marqueur nucléaire RPB2 basé sur un travail de Goetsch et collaborateurs (2005) date le groupe basal de *Vireya* à -13 Ma seulement (Webb et Ree, 2012).

Les Cycas aussi sont de très bons candidats car c'est un groupe très ancien d'environ 300 espèces actuelles dont les genres sont datés de -250 Ma (Hill et Stevenson, 1998). Les Cycas sont les plus répandus et les plus divers, Norstog et Nicholls en 1997 suggèrent, qu'en plus, leur morphologie a très peu changée.

Ces quelques exemples illustrent bien la nécessité de connaître les âges des taxons étudiés. Or, notre analyse préliminaire regroupait déjà 16 plantes (dont le genre *Vireya* et les Cycas) et non moins de 14 insectes (dont des Odonates et des coléoptères), ce qui est un argument supplémentaire pour dire que l'aréogramme que nous obtenons repose sur des données sources pertinentes.

2.2. Wallacea : une zone tampon majeure

Revenons sur la Wallacea, cette zone si « étrange » où s'entremêlent taxons asiatiques et australiens. Une autre perspective intéressante consisterait à se focaliser sur les taxons maritimes. En effet, l'archipel Indo-Australien est le plus spectaculaire des archipels mondiaux de part sa richesse (Bellwood et al. 2012). Il est défini comme étant la région marine entre les 15° N et 15° S et 100-150 ° E, globalement toute la partie marine de l'Asie du Sud-Est et de l'Ouest du Pacifique et forme un ensemble caractérisé, cohérent pour des analyses de

Biogéographie (Bellwood et al. 2012). Une littérature assez importante est désormais accessible et centrée sur des taxons marins dans la région (Santini et al. 2002, Lourie et Vincent 2004, Teske et al. 2004, Williams et Duda 2008, de Bruyn et al. 2012, Lohman et al. 2012 parmi d'autres), il est donc possible de mener une analyse de Biogéographie historique à partir de taxons marins seuls.

De plus, nous avons vu au point 2 de ce chapitre que le niveau de la mer avait changé durant le Quaternaire, ce qui a, indéniablement impacter fortement les taxons locaux. De Bruyn et collaborateurs en 2012 ont attiré l'attention sur le fait que concernant les gastéropodes d'eaux douces « *the probability of passive dispersal across sea water is extremely low for this group, prompting a vicariance hypothesis involving the drift of terranes, forming present day Sulawesi from the New Guinea margin in the Miocene, to explain the disjunct distribution on Sulawesi and the Torres Straits Islands.* ». Ils continuent ainsi à propos des crabes potamides d'eaux-douces « *one study found 3 deep lineages of pre-Pleistocene origin (Miocene to Pliocene) within genus Johora on the Malay Archipelago and Tioman island. They suggested that the origin of the entire group might be linked to the submergence of the Isthmus of Kra, subsequent diversification to the orogeny of the main range for the mainland clades and sea-level changes for the Tioman clade/species respectively. A more extended study (Shih et al 2009) revealed the existence of two clades within the largest subfamily with a large distribution in the IAA. Shih et al. suggested that this major split occurred when sea-level highstands in the early Miocene isolated the Sunda islands from mainland Southeast Asia. Interestingly, the Philippines have been colonised twice, the north Philippine taxa originate from Southeast Asia while Isolapotamon from Mindanao must have come from Borneo possibly via the Sulu islands. The diversification, thus, seems to be largely driven by pre-Pleistocene vicariant events, with possibly just a single instance of late Miocene dispersal into the Philippines.* ».

Somme toute, il apparaît indispensable de se pencher sur ce problème particulier que constituent les taxons aquatiques (à la fois marins et d'eaux-douces), d'autant plus que cela rejoint l'interrogation soulevée dans les conclusions générales (voir chapitre I) au sujet des incitations à mettre en place pour préserver cette biodiversité via la création d'aires protégées.

2.3. Sulawesi : une île modèle pour tester la vicariance

Un autre point intéressant que nous aimerions soulever ici, se rapporte au cas très singulier de l'île de Sulawesi.

Wallace qualifiait déjà cette île d'unique en 1881 :

« *there is no other example on the globe of an island so closely surrounded by other islands on every side, yet preserving such a marked individuality in its forms of life ; while, as regards the special features which characterize its insects, it is, so far yet known, absolutely unique.* ». **(Wallace 1881).**

Plusieurs études font état de son statut si particulier que ce soit sur le plan biologique (Evans et al. 2003, Stelbrink et al. 2012 par exemple) que géologique (Metcalf 2002, 2011, 2012, Hall 2002, 2011, 2012 par exemple).

Si l'on s'attarde sur les reconstructions géographiques, de toute évidence, Sulawesi est une île composite formée d'assemblage de terrains provenant du Sundaland, du plateau philippin et du plateau australien (Figure 18). C'est donc une parfaite candidate pour tester des événements de vicariance, on s'attendrait à voir sur l'aréogramme que le Nord de Sulawesi soit plus proche des Philippines, l'Ouest plus proche des îles Indonésiennes (Bornéo, Java) et le Sud de Sulawesi plus proche de l'Australie. La question reste à ce jour toujours ouverte.

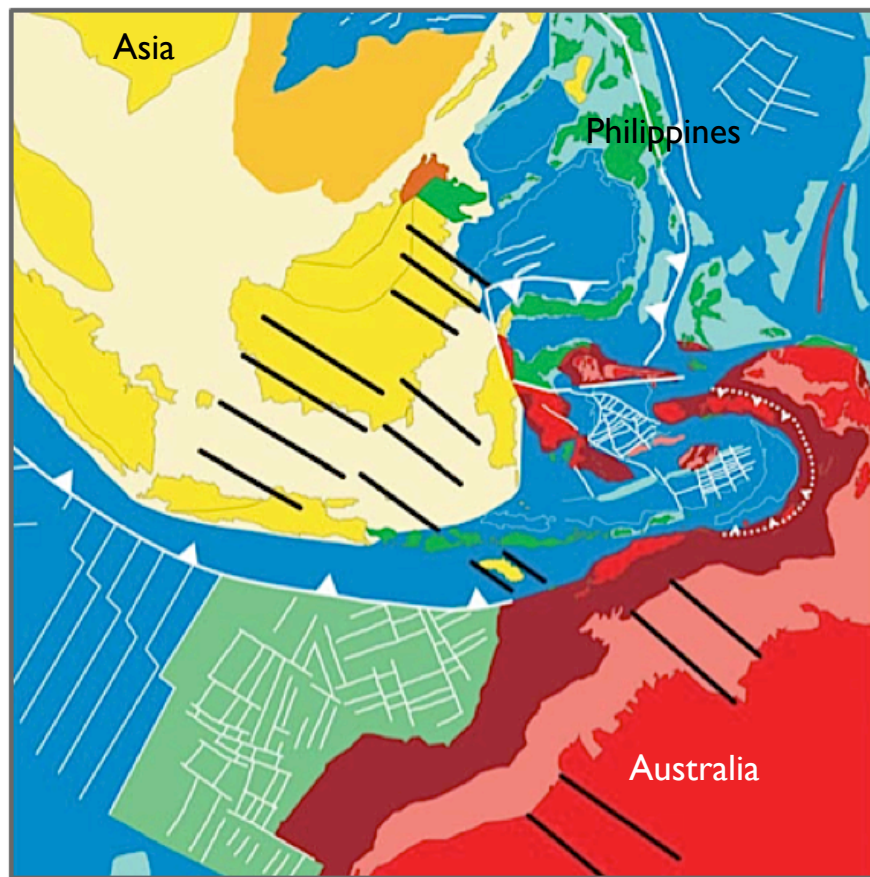


Figure 18 Reconstruction de la région de Banda à 4 Ma. (modifiée d'après Hall 2013, avec autorisation de reproduction de l'auteur)

2.4. Tester l'existence de « Pandora »

En 2010 Parenti et Ebach ont publié un article « Wallacea deconstructed » dans lequel ils testent l'existence de la région Wallacea en tant que région biogéographique unique. Pour cela ils utilisent une approche de Biogéographie comparative, tout en se servant de la classification/codification des aires d'endémisme sur toute la région Indo-Ouest Pacifique établit par Silder et Silder en 1938-1939.

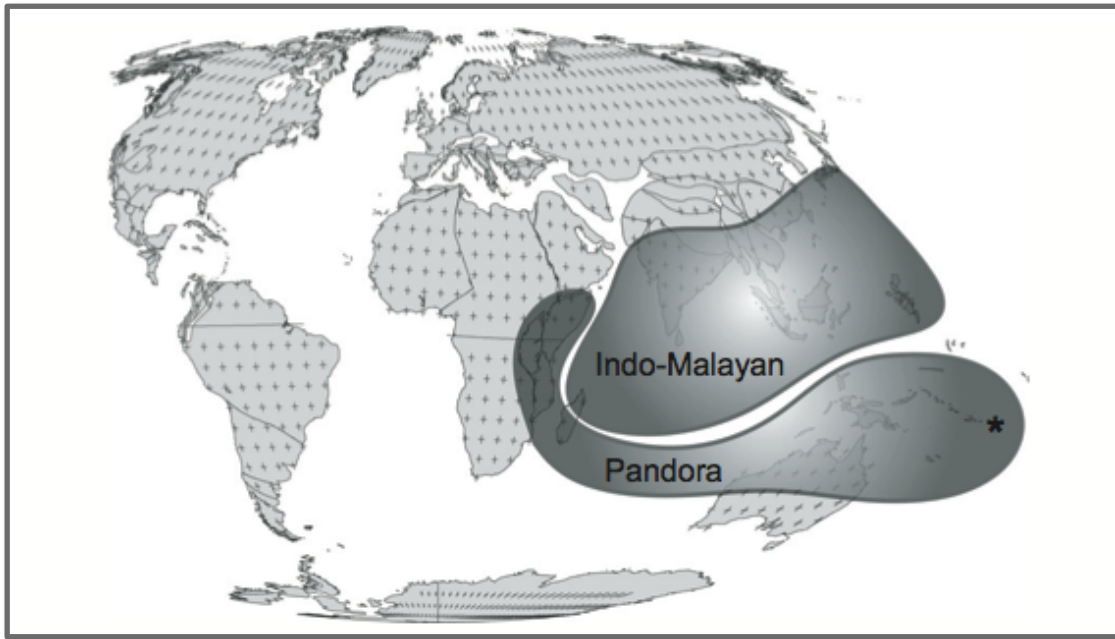


Figure 19 Sous-régions biogéographiques Pandora et Indo-Malésienne représentées sur une reconstruction globale du monde à -30 Ma (carte provenant de l'Université de l'Institut du Texas pour la Geophysique) (Parenti et Ebach 2012, avec autorisation de reproduction de l'éditeur).

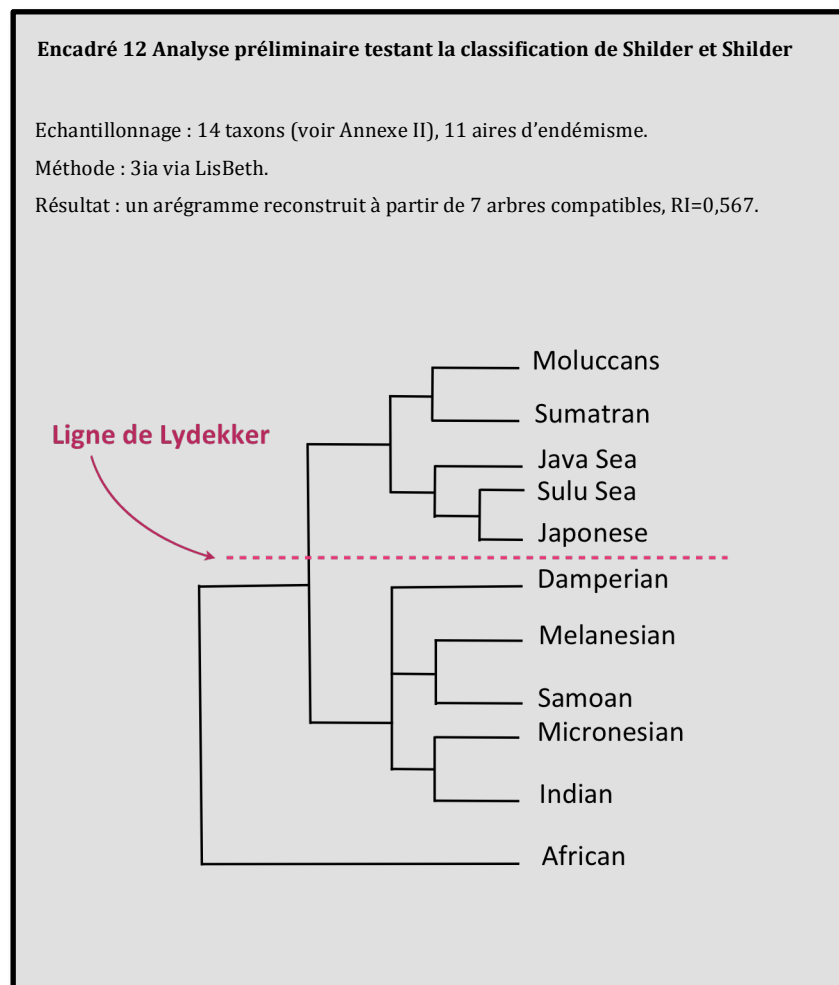
Suite à leurs résultats ils en arrivent à la conclusion que la Wallacea n'EST pas une région biogéographique unique et ils proposent deux sous régions basées sur leur code de nomenclature pour les aires (ICAN Ebach et collaborateurs, 2009) : la région Indo-malaisienne et Pandora (Figure 19).

Nous ne discuterons pas ici de la pertinence de l'ICAN car nous avons déjà soulevé certaines critiques à son égard (Zaragüeta Bagils et al. 2009) ; plutôt, nous préférons discuter de l'existence de ces deux régions biogéographiques. Pour cela, il suffira de collecter des cladogrammes de taxons présents à la fois dans l'Asie du Sud-Est, dans le Pacifique et sur la côte Est de l'Afrique afin de reconstruire un aréagramme et voir si (1) un pattern général se dessine et si (2) les deux régions sont retrouvées (ou l'un des deux).

2.5. Tester la classification de Silder et Silder 1938-1939

En parfaite liaison avec ce que nous venons de dire, une autre analyse intéressante serait de tester la pertinence de la classification de Silder et Silder 1938-1939. Nous pourrions dans un premier temps reprendre l'échantillonnage de l'article en révision sur le Sud-Est asiatique, mettre en correspondance nos aires d'endémisme et celles de Silder et Silder et voir si nous retrouvons les mêmes relations de parenté que Parenti et Ebach qui les a amené à définir les deux sous-régions Indo-Malaisienne et Pandora.

Nous avons déjà effectué une analyse préliminaire sur un échantillonnage réduit (Encadré 12) non publiée encore mais dont les résultats tendent à réfuter les deux sous-régions, en revanche la ligne de Lydekker est à nouveau retrouvée.



Cette dernière approche permettrait de répondre à la difficile question de l'existence d'une classification naturelle des aires biogéographiques tout comme nous recherchons une classification naturelle des organismes en Systématique.

2.6 Nouvelles pistes pour établir des règles de délimitations d'aires de conservation

Nous avons vu que l'utilisation des principes de la cladistique nous amenait à déterminer des relations de parenté entre aires biogéographiques sur la base de la présence conjointe de taxons. Mais finalement, dans quel but ? Outre la satisfaction personnelle de tout chercheur à faire avancer l'état des connaissances, cette approche fournit de nouvelles bases pour la délimitation d'aires de protection tant terrestres que marines (voir chapitre I 3.2 l'article « To add a little semantics can help for conservation policies »). Parenti et Ebach (2009) insistent sur le fait que le concept d'endémisme a un rôle particulier à jouer en biologie de la conservation. Ce que l'on définit comme des « points-chauds » de biodiversité (Myers et al. 2000) sont des aires biogéographiques présentant un nombre extrêmement élevé de taxons endémiques ainsi qu'un habitat dégradé par les activités anthropiques. De ce fait, les initiatives majeures de conservation ont été concentrées sur ces « points-chauds » qui présentent à la fois une grande diversité et un fort taux d'endémisme. Si nous faisons le parallèle avec la Systématique, cette approche, basée sur une similarité globale, se rapproche le plus des méthodes phénétiques, rien n'est dit sur les relations de parenté entre aires biogéographiques. Or, la recherche d'une classification naturelle des aires d'endémisme occupées par des groupes naturels de taxons est le principal objectif de la Biogéographie systématique, historique.

Myers et collaborateurs en 2010 énoncent que « *Biogeography and conservation are linked inexorably by the relationships between habitat area, primary productivity, earth history, and species richness. This linkage is especially strong in Southeast Asia where the areal extent of the land has repeatedly fluctuated twofold in the last few million years* » (Myers et al. 2010 :920).

Un des conséquences pourrait donc, concerner l'aide aux décideurs dans leur politique de protection environnementale. La faune et la flore exceptionnelles du Sud-Est asiatique et de l'Ouest du Pacifique sont victimes de la surexploitation forestière sur ces îles et déclinent indiscutablement, s'ajoute à cela le commerce animalier qui déciment toujours plus d'espèces aujourd'hui: les plus emblématiques étant les tigres, singes et certaines espèces de tortues, chassés à des fins alimentaires et médicinales. Il en va de la survie de nombreuses espèces : les populations d'Orang-Outang que l'on ne trouve plus que là-bas affichent un déclin dramatique, certains des derniers refuges de deux espèces de Rhinocéros se trouvent sur les îles de Sumatra et de Java³.

¹www.conservation.org/where/priority_areas/hotspots/asia-pacific/Sundaland/Pages/default.aspx

La carte de la Figure 20 indique les « points-chauds » de biodiversité à l'échelle mondiale parue en 2011 par l'organisation **Conservation International**. On y voit aisément que le Sundaland ainsi que tout le Sud-Est asiatique font partis des sites à protéger prioritairement. En 2005, déjà, cette même organisation listait 34 « points-chauds » de biodiversité dont le Sundaland, la Wallacea, les îles Solomons et de l'archipel de Bismark ainsi que toute la péninsule sud asiatique. De récentes études publiées dans un ouvrage édité en 2012 (Gower et al. 2012) font état cette situation tragique. Au travers de plusieurs exemples : commerce des reptiles, fragilité des écosystèmes tourbiers et des récifs coralliens, les différents auteurs soulignent les conséquences graves des activités anthropiques sur l'environnement. Inutile de préciser que ce problème, sans être nouveau, est loin d'être résolu. Il nécessite des incitations internationales collaboratives de protection de ces zones afin de réduire la déforestation par exemple ou le commerce des animaux. Notamment des contrôles bien plus drastiques aux frontières pour éviter la contrebande. De même, une des premières solutions repose sur l'éducation et la prise de conscience des communautés locales de la nécessité de sauvegarder leur environnement. Seulement, cela ne peut être possible que par la prise de décisions politiques des différents gouvernements concernés. Et là encore, l'action conjointe internationale est plus que souhaité. Comment laisser l'entière responsabilité de la sauvegarde d'un patrimoine mondial à des pays qui ont déjà tant de mal à garantir un certain niveau de vie à leurs concitoyens ?

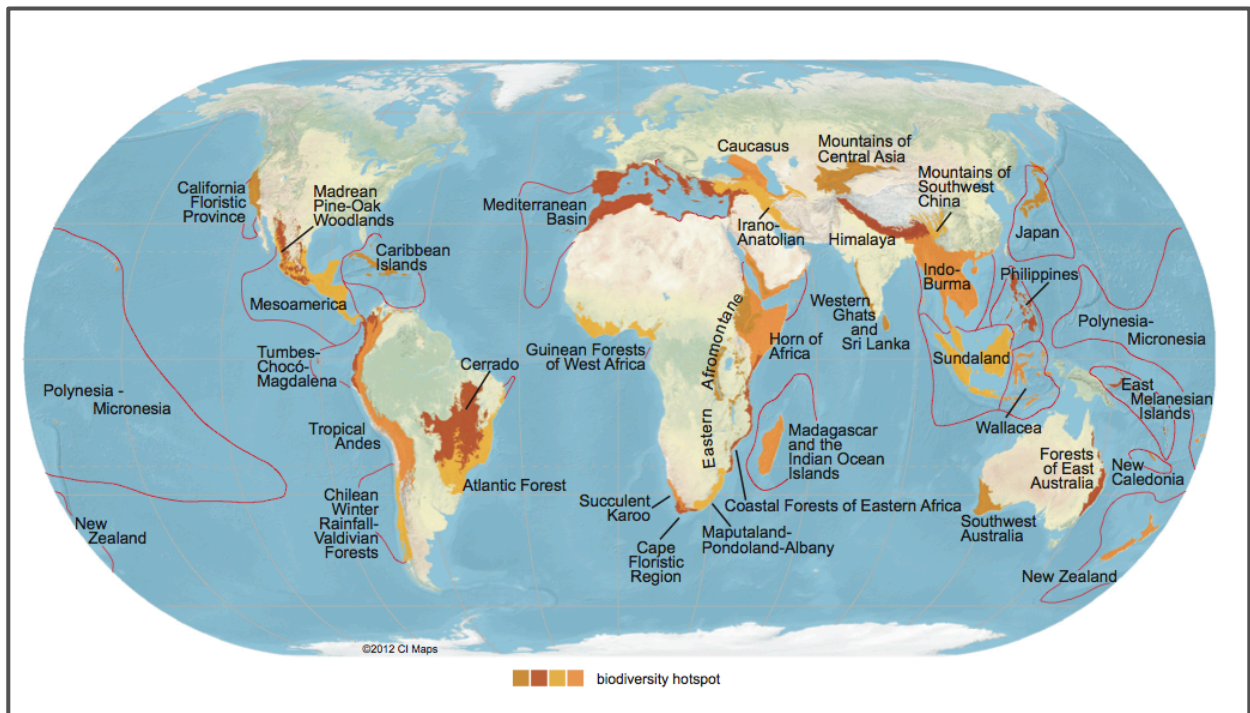


Figure 20 Carte mondiale des régions présentant des « points-chauds » de biodiversité (2011, www.conservation.org)

Nous clôturons maintenant ce long chapitre sur le Sud-Est asiatique qui nous a permis d'étudier un cas empirique et nous allons retourner vers des considérations méthodologiques concernant un sujet non moins intéressant : la Biogéographie et le temps.

Chapitre III : Réflexion sur une approche intégrative de la dimension temporelle en Biogéographie historique

*« What can paleontology offer to phylogenetic and historical biogeographic (particularly vicariance
biogeographic) studies ? »*

Lance Grande, 1985 : 234

Introduction

À la question « Pourquoi les taxons sont-ils distribués là où ils le sont aujourd'hui ? », Platnick et Nelson en 1978, répondent qu'ils le sont pour deux raisons, que deux processus peuvent être responsables de cette distribution. Le premier, la vicariance, implique que les taxons ont évolué et se sont spécialisés là où ils vivent, prédisant ainsi que les taxons se trouvant dans une ou plusieurs aires et les barrières entre ces aires sont du même âge. Le deuxième, la dispersion, implique une spéciation postérieure à la mise en place d'une barrière prédisant que la barrière est donc d'un âge antérieur à la spéciation.

La principale différence repose donc sur la relation existant entre l'âge du (des) taxon (s) et l'âge du (des) barrière (s) qui limite(nt) les aires biogéographiques.

La Biogéographie doit donc tenir compte du temps, et ce, encore plus lorsque l'on se place dans le paradigme de la Biogéographie historique qui recherche des patterns généraux (voir chapitre I et chapitre II). Il est bien évident que les distributions taxiques actuelles ont été façonnées au fil du temps, des années d'évolution entremêlant évolution géologique et évolution biologique. Le cadre est ainsi posé : il s'agit de trouver un moyen de tenir compte de façon non ambiguë de la dimension temporelle et d'arriver à une véritable Biogéographie historique intégrative.

Ce chapitre a pour objectifs : (1) de faire un état de l'art sur l'« intégration » du temps dans les études biogéographiques actuelles et (2) d'analyser « où » se trouve le temps réellement et enfin (3) de présenter une possible solution afin de rendre la Biogéographie cladistique telle que nous la pratiquons (au sens hiérarchique) véritablement intégrative.

1. Etat de l'art

La Biogéographie historique doit par essence intégrer la dimension temporelle. Or, nous remarquons que dans la littérature, peu d'articles en font véritablement état. Nous avons décidé de ne retenir ici, que quelques études qui, à notre sens, donnent effectivement une vision alternative et intégrative du temps et que nous allons présenter successivement par ordre chronologique.

1.1 « The time control refinement »

Lance Grande fut un des premiers auteurs à introduire la notion temporelle dans l'un de ses articles intitulé « **The use of paleontology in systematics and biogeography, and a time control refinement for historical biogeography** », publié en 1985 dans la revue *Paleobiology*. Il y dressait déjà les avantages à utiliser des fossiles pour la Systématique et la Biogéographie, qui sont de quatre ordres selon lui : le premier, est d'apporter des

taxons supplémentaires avec de nouvelles données morphologiques et ontogénétiques (lorsqu'ils sont bien conservés); le deuxième concerne l'extension de l'étendue biogéographique d'un taxon; le troisième est d'établir un âge minimal pour un taxon et enfin, le quatrième, de présenter des biotas fossiles qui peuvent être inclus dans des patterns biogéographiques non reconnaissables dans les biotas plus récents ou plus vieux. Ce qui est intéressant dans son article, c'est cette proposition d'une solution pour intégrer des données de fossiles (caractères relatifs aux fossiles eux-mêmes mais également leurs habitats) dans des analyses de Systématique (ce qui était déjà couramment appliqué) et de Biogéographie, le 'time control refinement' permettrait de tenir compte de la différence temporelle entre fossiles et actuels en se basant sur l'approche biogéographique cladistique énoncé par Platnick et Nelson en 1978, Nelson et Platnick en 1980 et 1981 et également Rosen en 1975 et 1978. Le lecteur est prié de se référer à l'Encadré 2 du chapitre 1 pour plus de détails. En quelques mots, cette méthode consiste en l'énoncé que lorsqu'il y a une perte de résolution dans la topologie d'un arbre d'aires, *id est* un aréagramme ou cladogramme d'aires suite à l'incorporation d'informations (dispersion de taxons) provenant de biotas plus récents, on peut conclure qu'un plus vieux (donc en position plus basale sur l'arbre, c'est à dire plus inclusif) biota endémique a été identifié (biota dont proviennent les taxons dispersant) pour lequel ces mêmes taxons auraient une distribution à large répartition (Figure 21).

« If we find that a biota's affinities change from indicating a more resolved pattern to indicating a less resolved pattern, we might conclude that we have identified an older basal endemic biota ».

Grande, 1985 : 238.

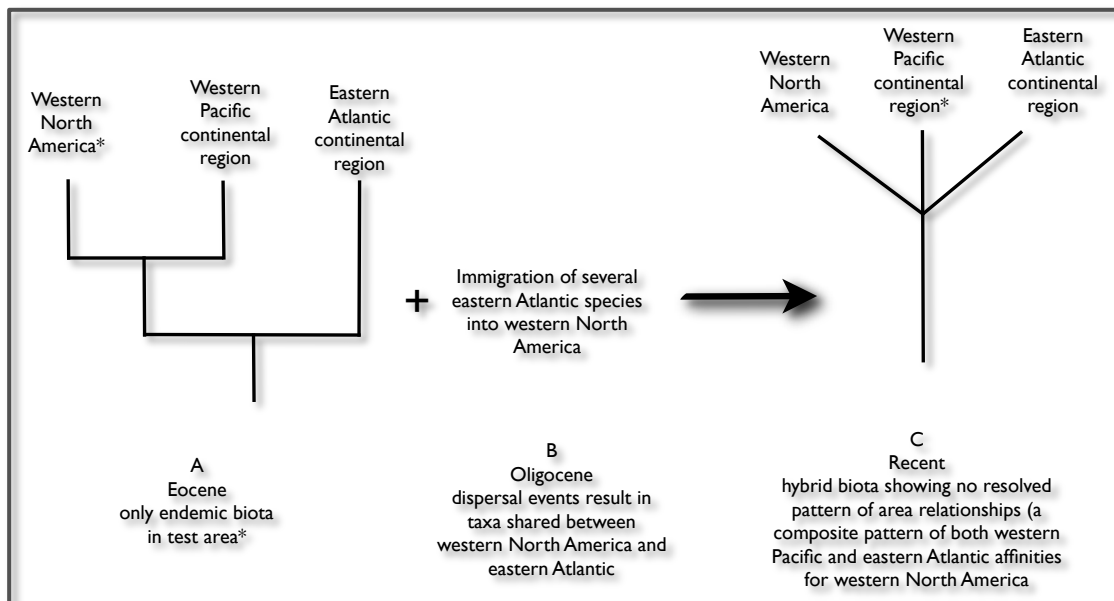


Figure 21 Modèle hypothétique présentant comment une histoire biogéographique complexe (non strictement vicariante) pour une aire test (ici l'ouest de l'Amérique du Nord) pourrait donner un pattern irrésolu pour les aires à une période récente. Les descendants des taxons ayant dispersés en B pourraient indiquer des relations de parenté transatlantiques, et obscurcir la précédente relation transpacifique de l'aire test (modifié d'après Grande 1985).

Cette proposition de méthode, illustrée sur un exemple de téléostéens (Grande est un paléoichtyologue) permet en quelque sorte de faire un ré-ajustement temporel lorsque des événements se produisent à des échelles de temps différentes. Au regard des avancées tant théoriques que méthodologiques de la Biogéographie historique, nous ne pouvons que conclure que cette première approche, bien qu'étant avant-gardiste pour son époque, reste très timide quant à ses conclusions. Elle n'est ni généralisée ni généralisable, pourtant, elle présente cet avantage de proposer une intégration « vraie » du temps en ce sens qu'elle combine des taxons d'âges différents.

1.2 « Temporal congruence and cladistic analysis biogeography and co-speciation »

En 1990, Roderic Page introduisit le « Temporal congruence and cladistic analysis biogeography and co-speciation » (Page, 1990). Dans ce modèle, Page fait le parallèle entre la Biogéographie et la co-spéciation. En effet la relation existant entre un taxon et son aire d'endémisme peut être assimilée à celle existant entre un parasite et son hôte. Partant de ce postulat, l'auteur cherche à « réconcilier » des arbres de parasites et d'hôtes lorsque ceux-ci présentent une topologie non congruente. L'exemple de la Figure 22 illustre comment deux arbres non congruents peuvent être « réconciliés » pour rendre les topologies congruentes (les relations entre chaque terminal ne doit en aucun cas être modifiées). Ce faisant, ce que l'on observe, c'est la création de nouveaux nœuds dans l'arbre (Figure 22 **B**) et donc l'introduction de paralogie taxique, les relations entre a, b, c et d sont dédoublées sur l'arbre résultant, ce qui est contraire aux premières intentions de Page : tester des hypothèses de cospéciation et par homologie, de vicariance. En Biogéographie cladistique, lorsque l'on recherche des relations de parenté entre aires d'endémisme, on cherche à tester des hypothèses d'homologie (en terme d'âge minimale) sur des aires que l'on considère homologues (Figure 29). Or dans ce cas précis il faudrait auparavant supprimer la paralogie pour procéder à l'analyse même. Ce qu'il faut souligner toutefois, c'est qu'à ce moment là, les développements autour de la 3ia implémentant des principes théoriques n'en étaient encore qu'à leurs balbutiements, il faudra attendre 1991 pour voir apparaître les publications phares de Gareth Nelson, Norman Platnick et Pauline Ladiges sur le sujet. Ceci étant, Page, lui aussi, à ce moment précis songeait déjà à une Biogéographie intégrative en cherchant à mettre en relation cladogrammes et échelle temporelle (Figure 23).

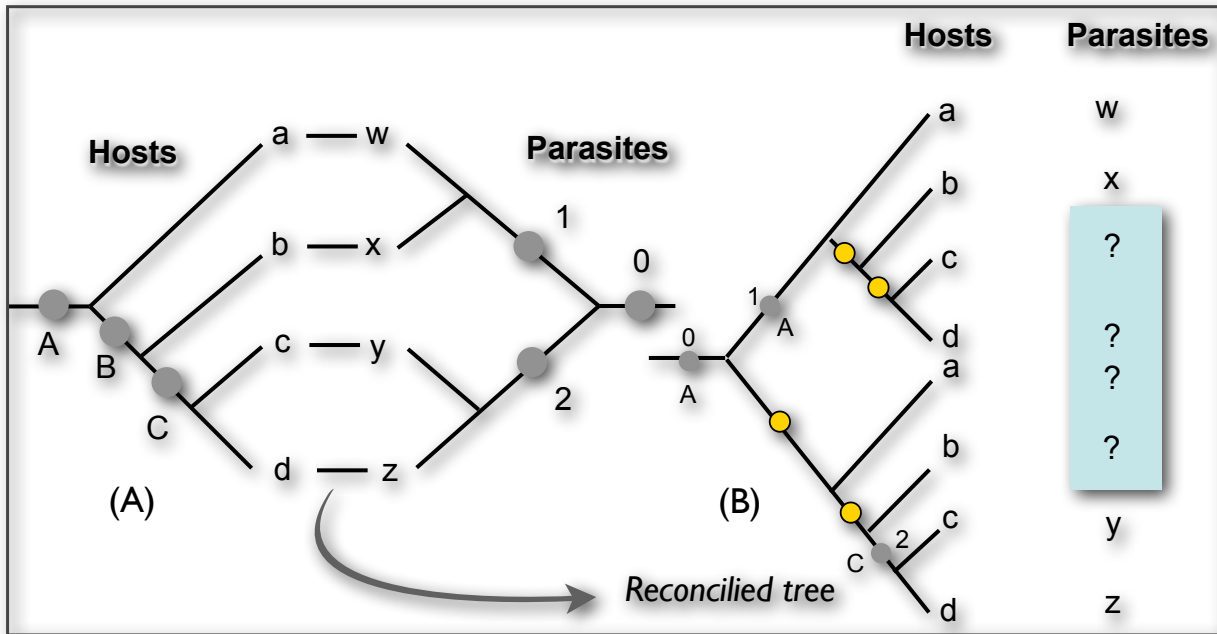


Figure 22 « Réconciliation » entre des patterns incohérents d'hôtes et de parasites (A) pour les rendre congruents (B). Les cercles jaunes et les points d'interrogation indiquent les composants et termes nécessaires à la réconciliation (modifié d'après Page 1990).

La Figure 23 montre la datation (en terme d'âge minimale) d'évènements cladogénétiques sur la phylogénie des hôtes en (a) ; en (b) et (c) figurent les phylogénies des parasites et les datations correspondantes. On remarque que la phylogénie en (b) est cohérente d'un point de vue datation avec l'arbre réconcilié c'est-à-dire que l'évènement de différenciation entre w et x est contemporain à celui de a par rapport au reste. De même pour y et z, la datation est la même que pour c et d. Ce qui n'est pas le cas sur la Figure 23(c) ou seul y/z est cohérent avec l'arbre réconcilié, l'évènement cladogénétique séparant w et x étant postérieur à celui de leur hôte a et b. De ce constat, Page conclut que l'information apportée par la datation des évènements de cladogénèse autorise plus d'hypothèses de cospéciation que l'information apportée par le cladogramme seul. Etant entendu qu'une cospéciation est forcément contemporaine, s'il n'y a pas de congruence temporelle entre les cladogrammes d'hôtes et de parasites on ne peut pas en conclure qu'il y a eu cospéciation.

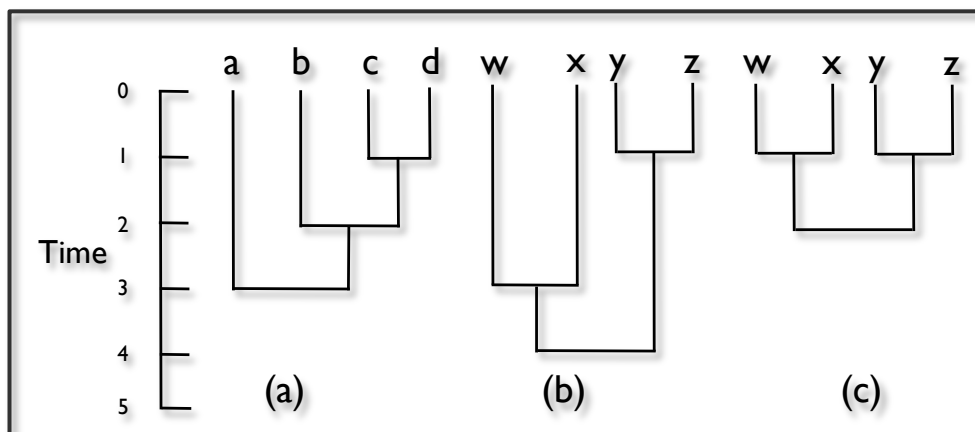


Figure 23 Exemples de phylogénies et de leur adéquation temporelle. (a) une phylogénie d’hôtes cohérente avec le cladogramme d’hôtes de la figure 12. En (b) et (c) deux phylogénies de parasites cohérentes avec le cladogramme de parasites de la figure 12. Les âges des évènements cladogénétiques dans les phylogénies de parasites sont tous cohérents en (b) avec l’arbre réconcilié tandis que en (c) seul un événement avec cohérent avec l’arbre réconcilié de la Figure 22 (modifié d’après Page 1990).

1.3 Le découpage en « tranches temporelles » ou « Time-Slicing »

Nous l’avons déjà dit, la Biogéographie cladistique concerne les aires biogéographiques et leurs relations de parenté. Les distributions de taxons sont le résultat de processus agissant à différents niveaux temporels et spatiaux. En conséquence, la Biogéographie est souvent différenciée en Paléobiogéographie (lorsque les taxons sont fossiles) et Néobiogéographie lorsque les taxons sont actuels (Cecca et al. 2011). Dans cette distinction, la Paléobiogéographie intègre de fait le temps alors que la Néobiogéographie n’inclue pas de données temporelles. Cette vision particulière a amené certains chercheurs à effectuer des « découpages temporels » ou « time-slicing » lorsqu’ils étudient des taxons fossiles d’âges différents. C’est ce que nous allons développer ci-après.

1.3.1 « The Chronobiogeographic paradigm »

En 2001, Hunn et Upchurch ont introduit un ensemble d’approches qui utilisent le temps et l’espace simultanément afin de contraindre les interprétations des causalités phylogénétiques. C’est ce qu’ils appellent une « approche douce » qui traite les topologies d’arbre et les données stratigraphiques comme des jeux de données séparées par opposition à la « stratocladistique » énoncée par Wagner en 1998 et Fox et al. en 1999, qualifiée « d’approche dure » aux données temporelles dans laquelle les étendues stratigraphiques des taxons sont utilisées comme des caractères qui peuvent directement influencer la topologie du cladogramme.

Dans le « Chronobiogeographic paradigm » de Hunn et Upchurch, la congruence entre la structure des cladogrammes et leur occurrence stratigraphique est estimée par une variété de métriques. L'hypothèse retenue est que le cladogramme peut être interprété comme un caractère représentant l'histoire évolutive (Hennig 1966), c'est-à-dire que les cladogrammes en plus d'être une séquence de cladogénèse, contiennent aussi la séquence de transformation des caractères dans le temps et l'espace. Or si les cladogrammes retracent l'histoire des caractères, c'est-à-dire, une séquence de l'évolution des caractères, alors ils doivent contenir un axe temporel (de la racine vers les terminaux) que l'on qualifie d'indicateur de polarité temporelle (« temporal polarity »). Deux conséquences peuvent être tirées de ce raisonnement (les auteurs parlent d'axiomes) :

- (1) Un nœud fils ne peut se trouver en position plus basale que son nœud ancêtre sur l'axe temporel et conséquemment,
- (2) Un nœud ancêtre ne peut se trouver à une position plus apicale que l'un de ses nœuds fils sur l'axe temporel.

L'introduction de cette notion d'axe temporel vient souligner le fait qu'il n'est pas possible de faire de même pour l'espace. En effet, les aires ne peuvent pas être ordonnées le long d'un axe comme c'est le cas pour le temps.

« Without information on the geographical histories of the areas concerned, the multistate area character can be constrained only by information obtained from the spatial distributions on the taxon cladogram ». **Hunn et Upchurch 2001 : 393.**

En conséquence, la contrainte imposée par les aires est moindre comparativement à celle imposée par le temps. Le but des auteurs est de développer une méthode qui utilise les relations de parenté phylogénétique en combinaison avec le temps et l'espace afin de produire une « estimation » la plus simple de l'histoire biogéographique et surtout exempte au possible de toute hypothèse a priori. Les auteurs présentent leur modèle théorique, considérant un premier cas d'étude pour lequel toutes les données sont disponibles et le modifient ensuite en le complexifiant avec des données manquantes.

Comme illustration voici un exemple tiré de leur article dans lequel ils schématisent comment un événement de dispersion influe sur la topologie des cladogrammes d'aires (Figure 24).

Soit une aire unique (XYZ) à t_0 ; à t_1 , une première fragmentation sépare l'aire (X) de l'ensemble (YZ) suivie d'une seconde fragmentation qui sépare ensuite (Y) de (Z). A droite de la figure est représenté le cladogramme d'aires (ou aréagramme) correspondant (Figure 24A). En t_3 , si des événements de dispersion ont lieu de l'aire (X) vers l'aire (Z), l'aréagramme correspondant devient celui en B, avec (XZ) plus proche entre eux qu'ils ne sont de Y soit en écriture parenthésée : (Y(XZ)).

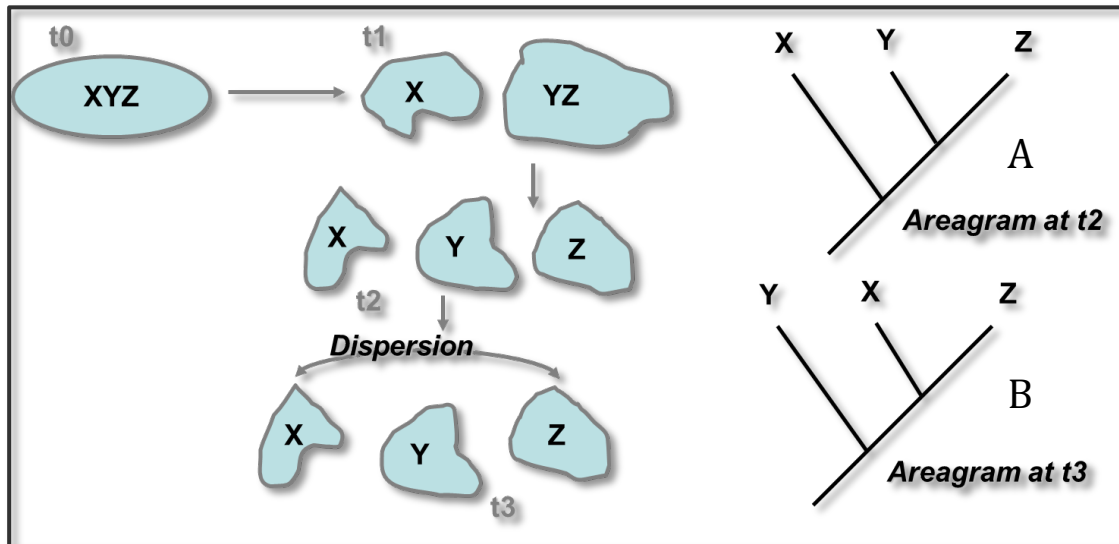


Figure 24 Séquence de fragmentation hypothétique de l'aire XYZ (t1 et t2) suivie d'une dispersion à t3 (modifié d'après Hunn et Upchurch 2001).

Seulement ils ne présentent aucune solution à l'intégration de la dimension temporelle. Ils affirment la nécessité d'avoir un algorithme approprié (Encadré 13) et ne font que soulever des concepts et procédures que cet « algorithme » (en fait, juste le problème reformulé sous forme de liste) devrait suivre sans finalement en donner une réelle implémentation.

Les auteurs concluent en disant que les données sur la distribution temporelle des taxons peuvent ajouter une importante contrainte sur les analyses biogéographiques et que les méthodes les plus rigoureuses et « philosophiquement justifiées », rationnellement, réalisant un diagnostic de causalité des évènements phylogénétiques sont celles qui prennent en compte les données de temps et d'espace simultanément. Nous ne pouvons qu'acquiescer à cette dernière conclusion. Cependant, aucune solution n'est encore énoncée pour intégrer le temps et le paradigme chronobiogéographique reste de l'ordre de l'énoncé théorique.

Encadré 13 Algorithme du « paradigme chronobiogéographique », « From Paradigm to Practice » (Hunn et Upchurch, 2001)

- (1) les hypothèses phylogénétiques (qui doivent être les briques de base de toutes analyses phylogénétiques) doivent être reconstruites en utilisant des techniques de cladistique ou de maximum de vraisemblance. Là encore, les fossiles peuvent être utilisés pour calibrer des horloges moléculaires et/ou pour leur dimension temporelle.
- (2) Un tel algorithme doit être appliqué à un jeu de données après apposition de la dimension temporelle et des ensembles spatiaux sur les nœuds des cladogrammes.
- (3) Ensuite, une analyse des composantes (ou d'une réconciliation) devrait être implémentée pour découvrir un cladogramme commun d'aires.
- (4) Ensuite la procédure devrait inclure un test des données indépendantes, ici la recherche d'une corroboration géologique.
- (5) L'idéal biologique est expliqué ci-après : « des postulats impliquant des causalités de spéciation peuvent être fait en considérant chaque instance de la relation de parenté ancêtre/descendant dans le temps et l'espace sur le cladogramme final dans leur contexte paléogéographique ».

1.3.2 « The Temporally Partitioned Component Analysis »

En 2002, ces mêmes auteurs, Hunn et Upchurch, finalisent leur paradigme chronobiogéographique en présentant la «Temporally Partitioned Component Analysis» (TPCA). La méthode intègre des données temporelles à des techniques cladistiques classiques (Nelson et Platnick 1981, Page 1988) via COMPONENT (Page 1993) et TreeMap (Page 1995). En premier lieu, la méthode recherche des « optimal cladograms of areas » ou OCA, en déterminant quelles relations de parenté entre aires produisent la « meilleure explication » (sous certaines conditions optimales) pour la distribution spatiale observée dans un cladogramme de taxons. Ensuite, un test de randomisation est utilisé pour déterminer si le degré d'adéquation entre les aires et la topologie des cladogrammes de taxons est meilleur qu'attendue par chance (Page 1991). La particularité de la méthode est de partitionner une période donnée en intervalles de temps (Figure 25). Un exemple tiré d'un échantillonnage de Dinosauriens est donné (composé des relations phylogénétiques et des distributions stratigraphiques/géographiques pour environ 150 genres, voir Upchurch et al 2002 pour plus de détails) et pour lequel ils montrent un patron incohérent lorsqu'ils appliquent leur méthode de TCPA. Leurs résultats montrent que les relations de parenté entre les taxons diffèrent totalement d'une tranche temporelle à une autre. Et ce, d'autant plus pour des tranches étendues (comme le « Mésozoïque » ou le « Jurassique »). Aussi, ils trouvent des OCA statistiquement significatifs pour le « Jurassique moyen », « Jurassique supérieur » et le « Crétacé inférieur » présentant des topologies incompatibles. Ceci les amène à la conclusion que les histoires biogéographiques peuvent parfois présenter des structures réticulées rendant compte de « fusion » d'aires ou de coalescence d'aires. Ce phénomène est aussi connu sous le nom de « géodispersion » (Liebermann 2000) auquel les auteurs

préfèrent le terme « Mass Coherent Dispersal » (MCD). Il rend compte du fait que lorsqu'une barrière physique entre deux aires est supprimée, les taxons peuvent potentiellement se disperser et ainsi changer les relations de parenté entre aires d'une façon similaire et au même moment.

Ainsi, un cladogramme d'aires unique ne peut pas être interprété sans ambiguïté.

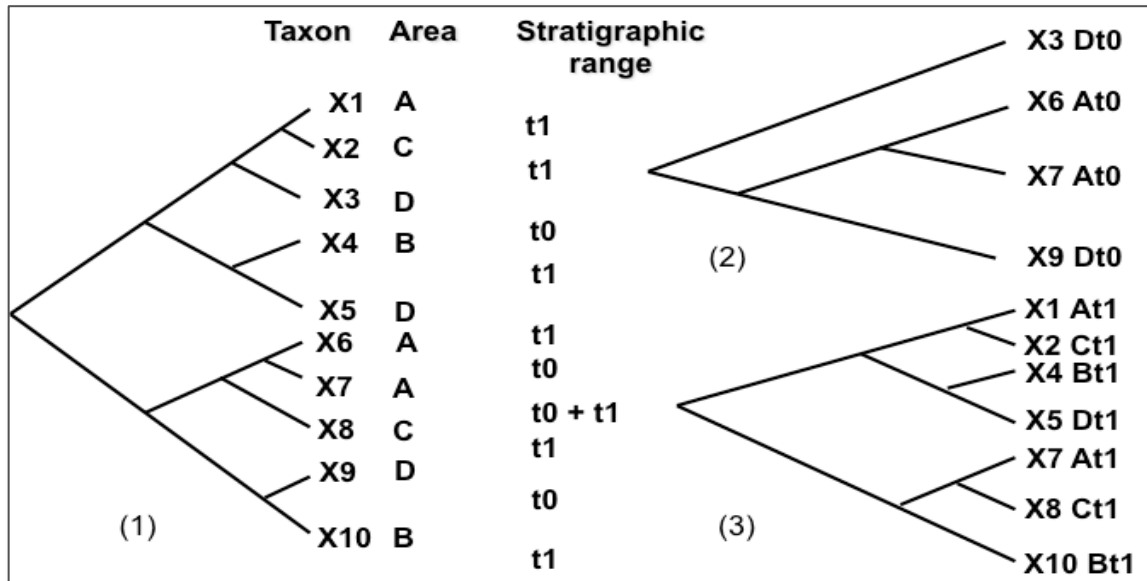


Figure 25 Représentation schématique d'une partition temporelle pour un cladogramme de taxons. (1) le cladogramme original contient 10 taxons (X1 - X10) distribués sur 4 aires (A - D) et présent à deux temps différents (t1 et t0). (2) et (3) sont les cladogrammes temporellement partitionnés contenant seulement les taxons présents à t0 ou à t1 (modifié d'après Hunn et Upchurch 2002).

Finalement, en dépit d'une conclusion qui va à l'encontre de ce que la Biogéographie historique recherche, à savoir des patterns généraux, leur approche présente un intérêt certain dans leur volonté d'intégrer la dimension temporelle. Seulement, ce que nous observons ce n'est ni plus ni moins que du partitionnement en tranches temporelles, en cela la méthode n'est plus intégrative.

1.3.4 « The Test of Time »

En 2003, Donoghue et Moore ont souligné la nécessité d'une « Biogéographie intégrative » dans laquelle la dimension temporelle devrait être intégrée contrairement à ce qui est habituellement fait dans les approches biogéographiques cladistiques. S'appuyant sur deux cas d'études classiques des deux hémisphères nord et sud,

ils ont interprété leurs résultats à la lumière du modèle de « congruence/pseudo-incongruence ⁴» énoncé par Cunningham et Collins en 1994, connu aussi sous l'appellation de « test du temps en Biogéographie historique ». Ce phénomène se produit quand deux arbres ont les mêmes topologies mais divergent dans le temps. Les auteurs ont définis 4 cas différents conformément à ce concept (Figure 26) :

1. « Congruence », lorsque la topologie des arbres et le temps sont congruents (i.e. coïncidence temporelle) ;
2. « Pseudo-congruence », lorsque les topologies sont congruentes mais pas le temps ;
3. « Incongruence », lorsque les topologies ne sont pas congruentes même si le temps reste correct ;
4. « Pseudo-incongruence » lorsque ni les topologies ni le temps concordent.

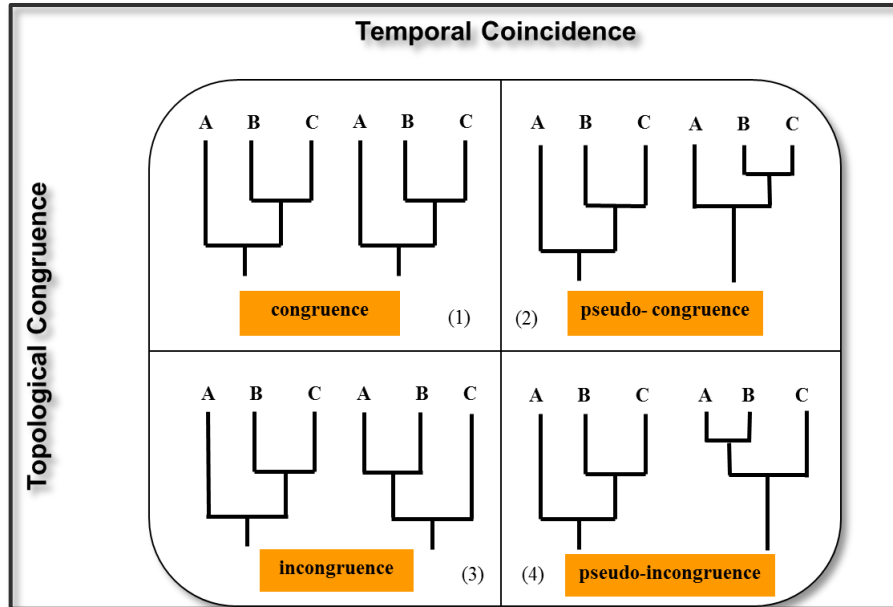


Figure 26 « The test of Time » en Biogéographie historique. (1) Congruence, (2) Pseudo-congruence, (3) Incongruence et (4) Pseudo-incongruence (modifié d'après Donoghue et Moore).

⁴ Nous attirons l'attention sur le fait que nous utilisons ici les termes anglais pour congruence et incohérence soit, « congruence » et « incongruence » afin de respecter le formalisme initial des auteurs.

Concernant l'exemple de l'hémisphère Sud, tiré de l'étude de Linder et Crisp (1995) sur le *Nothofagus* et la Biogéographie du Pacifique (voir Encadré 14 pour plus de détails), les auteurs concluent que si nous voulons établir l'impact d'évènements de vicariance particuliers, telle que la fragmentation du Gondwana, alors nous devrions choisir des groupes d'âges appropriés. Ce qui requiert une évaluation indépendante des temps de divergence. En effet dans cet exemple, ils soulignent qu'il est impossible de comprendre la signification des résultats quant aux causes sous-tendues en l'absence d'information sur la période à laquelle les taxons se sont diversifiés.

Encadré 14 « *Nothofagus* and Pacific biogeography » (Linder et Crisp 1995)

Les auteurs recherchent les relations de parenté entre l'Amérique du Sud (SA), l'Australie (AU) et la Nouvelle-Zélande (NZ) basées sur la distribution et les cladogrammes de *Nothofagus* et de 12 autres plantes à fleurs endémiques du Pacifique Sud. Leurs résultats (obtenus après une analyse implémentée dans COMPONENT, Page 1993a) indiquent que l'Australie et la Nouvelle-Zélande sont plus proches entre eux qu'elles ne le sont de l'Amérique du Sud : **(AS,(AU, NZ))**. Or ce pattern est conflictuel avec l'analyse de plusieurs groupes d'insectes et avec les preuves géologiques concernant l'histoire évolutive de ces mêmes terrains, à savoir, une plus proche parenté entre l'Amérique du Sud et l'Australie : **(NZ, (AU, SA))**. Selon les interprétations classiquement admises, la Nouvelle-Zélande se serait isolée il y a 80 millions d'années, tandis que l'Amérique du Sud et l'Australie étaient encore connectés (Antarctica) jusqu'à 40 millions d'années.

Le second exemple concerne la Biogéographie de l'hémisphère Nord, d'après une étude de Donoghue et collaborateurs publiée en 2001 (voir Encadré 15 pour plus de détails).

Les auteurs ont représenté leurs résultats dans une figure à trois dimensions montrant les cladogrammes d'aires, les temps de divergences et les patterns de mouvements inférés pour chaque taxon étudié sur le modèle de congruence-pseudo-congruence. Il apparaît que les disjonctions intercontinentales ont eu lieu à des périodes différentes impliquant des itérations d'épisodes de vicariance et de dispersion. Ils concluent :

“at this stage, truly integrative methods are likely to emerge, since cladistics has laid a very solid phylogeny-based framework, and there is absolutely no danger that temporal data will somehow displace the use of phylogenies”.

Donoghue et Moore, 2003 : 268.

Nous sommes bien évidemment en accord avec leur propos, malgré tout, il faut souligner le fait qu'encore une fois, la dimension temporelle n'est pas réellement “intégrée” puisqu'elle est prise en compte en “parallèle” aux

arbres de relations de parenté entre aires étudiées. L'analyse des cladogrammes de taxons renseigne sur les relations de parenté entre aires dont la topologie est ensuite mise en regard des temps de divergence de chaque taxon.

Gardons pour la suite de notre exposé que, comme le précise Donoghue et Moore (p 262): "*ignoring temporal information obscures the connection between biogeographic patterns and their underlying causes.*"

Encadré 15 « Phylogenetic patterns in Northern hemisphere plant biogeography » (Donoghue et al. 2001)

Les auteurs s'intéressent à 7 groupes de plantes présentant des disjonctions intercontinentales et des espèces distribuées sur 3 à 4 aires d'endémisme principales. Le placement de fossiles a fourni des points de calibration pour les datations moléculaires pour chaque à l'exception d'une pour laquelle un taux de substitution a été pris dans la littérature. Une analyse de Dispersion-Vicariance a été conduite avec DIVA (Ronquist 1996, 1997a) pour inférer des aires ancestrales et le sens de la dispersion pour chacun des clades. Cette méthode replace une ou plusieurs aires (en fonction des contraintes imposées) à chaque nœud interne de sorte que les coûts associés à la dispersion et aux événements d'extinction soient minimisés (Ronquist 1997a, b).

Leurs résultats montrent que pour les analyses de 5 clades présentant des espèces en Chine (CH), Japon (JA) et l'ouest de l'Amérique du Nord (WNA), chaque possibilité est mise en évidence, à savoir :

1. **(CH, (JA, WNA))** pour *Hamamelis* et *Weigela-Diervilla* ;
2. **(JA, (CH, WNA))** pour *Triosteum* et *Buckleya* ;
3. **(WNA, (CH, JA))** pour *Torreya*.

Triosteum et *Buckleya* sont vraiment congruents, avec les mêmes relations de parenté entre aires et dans le même délai (divergence au Miocène supérieur). En contraste, *Hamamelis* et *Torreya* apparaissent clairement incohérents, avec des conflits de relation de parenté entre aires mais dans une même période (divergence Miocène inférieur).

Comparativement, *Hamamelis* et *Weigela-Diervilla* (divergence à la limite Miocène-Pliocène) montrent une instance de pseudo-congruence avec des relations de parenté entre aires identiques mais à des périodes différentes. Enfin, la comparaison de *Weigela-Diervilla* avec *Triosteum* ou *Buckleya* révèle une instance de pseudo-incongruence, avec à la fois des patterns conflictuels à des temps différents.

Les deux autres clades (*Liquidambar* et *Cercis*) utilisés incluent des espèces dans l'est de l'Asie, l'Europe et l'Amérique du Nord mais pas le Japon ni l'ouest de l'Amérique du Nord. Dans les deux cas, les espèces Européennes apparaissent plus proches des espèces Nord américaines **(Asia, (Europe, North America))** et la dispersion aurait eu lieu via l'Atlantique Nord. Néanmoins, dans le cas du *Liquidambar*, la disjonction Europe-Amérique du Nord est daté de 35 millions d'années alors qu'elle est de 15 millions d'années pour le clade *Cercis*, indiquant ainsi, une autre instance de pseudo-congruence.

1.3.5 « The Cladistic Biochronologic Analysis »

La dernière approche d'intégration temporelle que nous avons choisi de présenter est celle développée par Peter J. Makovicky dans un article paru en 2008 "*Telling time from fossils: a phylogeny-based approach to chronological ordering of paleobiotas*". Il y formule une nouvelle méthode (la "Cladistic Biochronologic analysis" ou CBA) pour déterminer l'ordre biochronologique entre des paleobiotas basé sur les relations phylogénétiques des taxons les constituant. Cette méthode, adaptée de la Brooks Parsimony Analysis (BPA) (voir Encadré 16 pour plus de détails) crée un ordonnancement temporel des paléobiotas. Le raisonnement sous-jacent est que le temps peut être imaginé comme un caractère continu, irréversible sur un cladogramme correctement enraciné, interprété comme un arbre évolutif. L'ordre des nœuds de la racine aux feuilles de l'arbre représente l'information temporelle du cladogramme en cela que le nœud le plus apical est toujours postérieur à tous les autres nœuds (Figure 27).

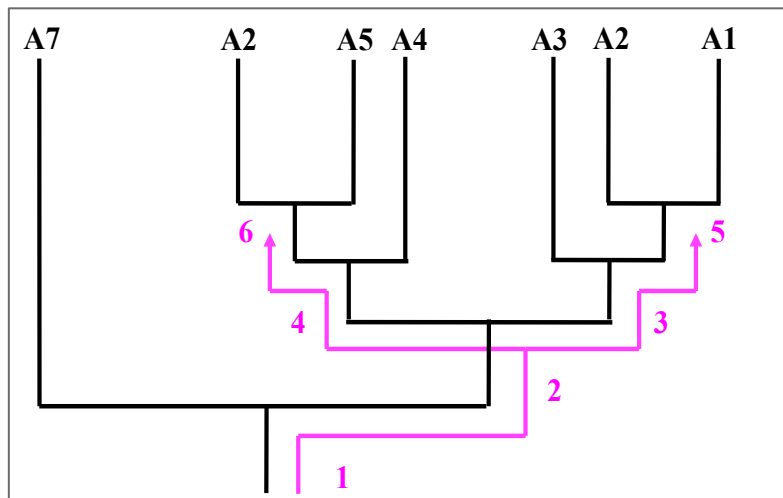


Figure 27 Relation entre le temps et les nœuds (taxon) d'un cladogramme enraciné. Le temps peut être vu comme un caractère irréversible, continu (flèches roses) dont les valeurs peuvent être suivies par l'ordre hiérarchique des nœuds du cladogramme (modifié d'après Makovicky 2008).

Une des exigences évidente de la CBA pour être précise est que le registre fossile soit généralement congruent avec les schémas évolutifs déterminés à partir des informations intrinsèques aux organismes autrement dit leurs relations de parenté.

L'unité d'analyse du CBA est le paléobiota, c'est-à-dire un biota éteint ayant sa propre histoire évolutive. Ceux-ci peuvent être délimités de différentes manières (par exemple, selon la présence ou l'absence de taxons

sélectionnés au sein de nombreux types de chronozones⁵ utilisés en biochronologie, ou par la combinaison de paramètres biotiques, géographiques et géologiques tels que ceux utilisés pour délimiter des localités fossilifères) et à différentes échelles (par exemple des localités/carrières, différent types de chronozones) mais doit être définis de façon équivalente dans une même étude. De ce fait, les paléobiotas ont des plages temporelles distinctes au cours du temps.

Sa méthodologie est la même que celle de la BPA de premier ordre (Brooks 1981, Kluge 1988, Wiley 1988 a, b, voir Encadré 16). Une matrice est construite avec les paléobiotas en place et lieu des taxons et les « syntaxons » (les nœuds internes) en place des caractères d'une analyse cladistique traditionnelle (Figure 28). À ceci près que les taxons terminaux ne sont pas inclus dans la CBA parce que l'information temporelle n'est pas hiérarchique à travers eux. Comme pour la BPA, le codage des « syntaxons » suit la méthode additive : un « syntaxon » est codé présent si l'un ou plus de ses descendants (y compris le « syntaxon » apical) est présent et reconnu à l'intérieur d'un paléobiota. Ensuite, la matrice codée est analysée par la méthode de maximum de parcimonie dans le but de trouver un arbre ou des arbres qui maximise(nt) la congruence entre les informations apportées par chaque syntaxon. Ces arbres chronologiques, ou cladochronogrammes, sont enracinés par un ancêtre entièrement nul comme en BPA ou alors par un paléobiota reconnu pour être antérieur à tous les autres biotas de l'analyse.

Makovicky affirme que sa méthode présente des avantages certains dans des situations où des données stratigraphiques traditionnelles sont insuffisantes ou non disponibles pour la plupart ou tous les paléobiotas étudiés. Ceux montrant des données stratigraphiques pauvres peuvent être analysés en même temps que ceux pour lesquels l'ordre temporel est connu ou mieux contraints. En cela, la CBA est une adaptation de la BPA pour l'analyse du temps plutôt que des relations de parenté biogéographique historique entre les biotas.

⁵ En stratigraphie, division plus petite que l'étage (Dictionnaire de géologie, Foucault et Raoult, 2010).

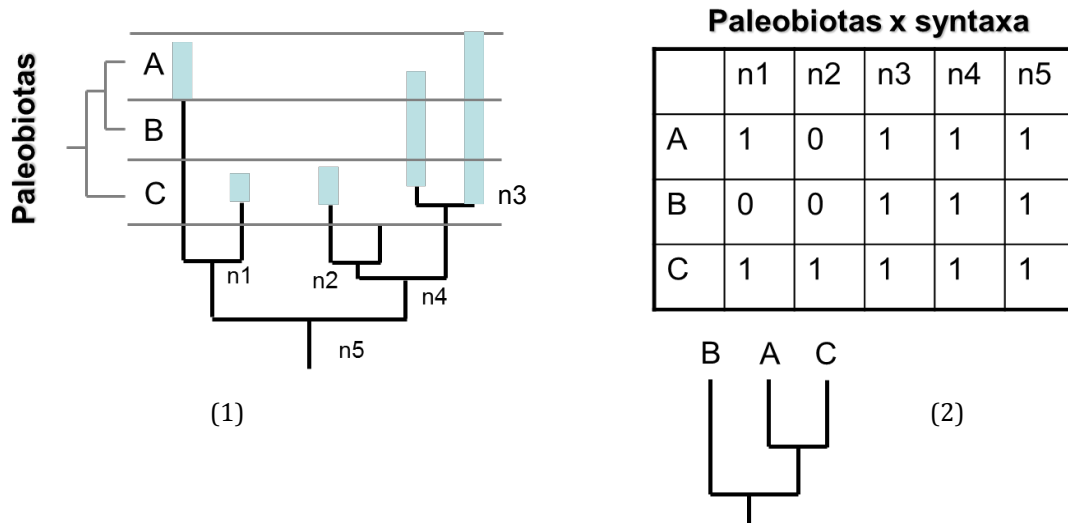


Figure 28 Phylogénie calibrée d'un clade hypothétique à 6 taxons dont les distributions couvrent 3 paléobiotas (A, B et C) en (1) et en (2) la matrice paléobiota/syntaxon correspondante ainsi que le cladochronogramme résultant.

Que pouvons-nous dire de cette approche ?

Premièrement, que l'auteur a tout à fait raison dans sa compréhension de la dimension temporelle, à savoir que le temps est hiérarchique et procède de la racine aux terminaux et que seuls les nœuds internes sont pertinents pour l'analyse à cause de la présence possible de paralogie temporelle au sein des cladogrammes (Zaragüeta-Bagils *et al*, 2004) ; ainsi, les terminaux deviennent inutiles.

Du reste, nous devons souligner que l'interprétation de l'auteur au sujet de la diachronie (p 352) n'est pas convaincante. En effet, il soutient que c'est un non-sens de combiner des données diachroniques par la seule raison que des événements locaux d'extinction et de vicariance/endémisme et de dispersion diachroniques peuvent causer des problèmes et qu'il est donc plus judicieux de contraindre l'aire géographique.

« it makes little sense to incorporate Cretaceous od Cenozoic fossil vertebrate faunas of North and South America into a single analysis, as these continents had nearly separate biotic histories following the break-up of Gondwana in the Late Jurassic (Scotese 2004), and thus share few clades that diversified in post-Jurassic times and would be informative with respect to temporal order of extinc biotas on both continents. ». **Makovicky 2008 : 352.**

Avec cette assertion, Makovicky prouve qu'il se place d'un point de vue d'énoncés narratifs quant aux processus géologiques (de transformation des terres au cours du temps) ayant conduits à ces topologies. Il ne se place plus du tout dans un schéma de recherche de patrons généraux mais des processus sous-jacents. Or il faut garder en tête que ce sont les cladogrammes de taxons qui nous informent sur les relations de parenté entre les aires d'endémisme considérées comme des aires biogéographiques et non simplement géographiques/géologiques.

Encadré 16 « Brooks Parsimony Analysis » (Brooks 1981, Kluge 1988, Wiley 1988 a, b)

La BPA a été proposée par Wiley (1987, 1988 a, b) et modifiée ultérieurement par Brooks (1990). La méthode est une analyse de parcimonie de cladogrammes taxons-aires codifiés comme des variables à deux états et traités comme des caractères (Morrone 2009 et références internes). Pour utiliser la BPA, une matrice aires-taxons est reconstruite à partir de cladogrammes de taxons et dans laquelle les noms des taxons terminaux sont remplacés par les aires dans lesquelles ils sont distribués. Chaque terminal et nœud interne est numéroté et codé présent (1) ou absent (0) dans la matrice. Les aires manquantes sont codées « inconnues » (?) et les taxons à large répartition sont codés comme des « synapomorphies » (selon l'implémentation de l'Assumption 0 (Zandee and Roos, 1987; Wiley, 1988a)). Une ligne additionnelle ne comportant que des « 0 » est rajoutée pour enraciner l'arbre. La matrice est ensuite analysée par un logiciel implémentant un algorithme de parcimonie dans le but d'obtenir un cladogramme général d'aires.

Rq : il existe 3 « Assumptions » en Biogéographie qui permettent de traiter les MASTs, l'assumption 0 ou A0, l'assumption 1 ou A1 et l'assumption 2 ou A2 (Nelson and Platnick 1981, Page 1988).

2. Synchronie, diachronie : le temps, une dimension oubliée ?

2.1. Pourquoi la Biogéographie historique est-elle un problème cladistique ?

Toute la problématique qui nous préoccupe aujourd'hui concerne la Biogéographie historique que l'on cherche à rendre « intégrative » en tenant compte du temps. Seulement, en quoi la Biogéographie historique est-elle adaptée à ce genre de procédure ?

Dans la Figure 29 ci-après, d'après une communication personnelle de René Zaragüeta Bagils, les différentes étapes d'une analyse phylogénétique cladistique classique sont mises en regard de celles nécessaires à une analyse biogéographique historique classique. On ne peut que constater la parfaite isométrie entre chacune de ces étapes. En Systématique, comme en Biogéographie, une analyse n'en est réellement une que parce que nous réalisons des tests d'hypothèses d'homologie des caractères que nous avons définis au préalable. Nous cherchons à mettre en évidence les relations de parenté entre différents taxons sur la base de caractères hérités d'un ancêtre commun (les synapomorphies). La première étape consiste à établir les hypothèses d'homologie qui peuvent être représentées sous forme d'arbre hiérarchique (avec des états de caractères aux terminaux et aux nœuds). Le parallèle est exactement le même en Biogéographie historique, les hypothèses d'homologie primaire sont les cladogrammes de taxons non apparentés ; lorsque les taxons terminaux sont remplacés par les aires d'endémisme dans lesquels les taxons sont distribués, nous obtenons une structure hybride le Taxon-Area-Cladogram (ou TAC) qui correspond véritablement au caractère (avec des aires terminales et des clades aux nœuds) et ensuite l'analyse de congruence permet de reconstruire un cladogramme d'aires ou aérogramme présentant des aires terminales et des aires biogéographiques internes vues comme des ensemble de taxons.

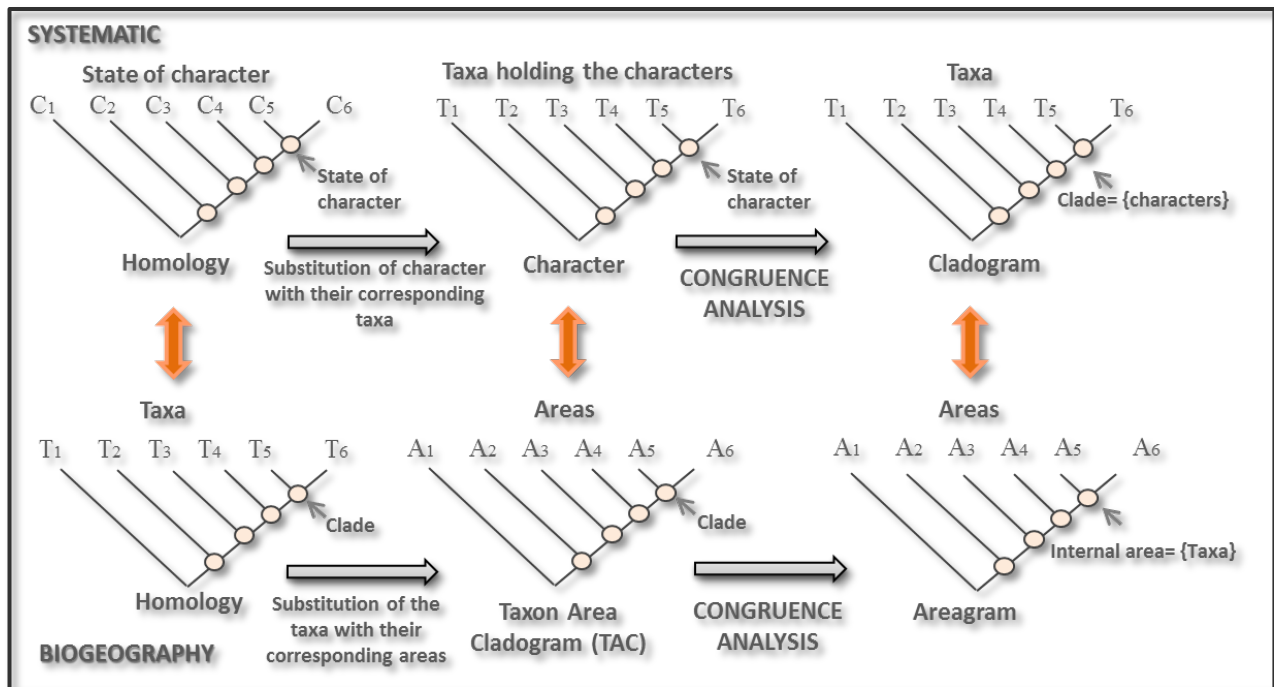


Figure 29 Equivalence entre la Systématique et la Biogéographie historique (Zaragüeta Bagils, comm. pers.)

Ce schéma illustre bien tout le propos dont il est question : la Biogéographie historique est un problème cladistique qui nécessite des méthodes cladistiques afin de reconstruire les relations de parenté entre aires d'endémisme.

La problématique de la dimension temporelle réside dans le fait qu'il n'existe à l'heure actuelle aucune méthode pour intégrer le temps : lorsque l'on analyse des cladogrammes de taxons dont certains présentent à la fois des taxons actuels et des taxons fossiles, comment considérer les aires des fossiles qui n'existent plus actuellement ? Autrement dit : comment traiter une analyse diachronique ?

2.2. Synchronie : analyse dans une même tranche temporelle

Nous l'avons vu dans la première partie de ce chapitre, plusieurs auteurs ont présenté des solutions à cette question or aucune n'est pleinement satisfaisante. Elles ont toutes en commun de proposer, *in fine*, des partitionnements en tranches temporelles et une analyse au sein d'une même tranche, ce qui équivaut en réalité, à une série d'analyses synchroniques. De la sorte, on s'abstrait de la dimension temporelle et de toute considération intégrative.

2.3 Diachronie : analyse d'une ou plusieurs périodes temporelles simultanément

La difficulté réside dans l'analyse d'une ou plusieurs périodes temporelles simultanément. Pour cela, il est impératif de se baser sur un formalisme sans faille, nous devons comparer ce qui est comparable ce qui requiert de se fonder sur un modèle qui autorise cette analyse multidimensionnelle. Et la représentation hiérarchique des unités opérationnelles qu'offre la 3ia en fait une méthode de choix pour l'intégration de la dimension temporelle.

3. Vers une Biogéographie historique intégrative

Le lecteur l'aura compris, nous allons, une fois de plus, nous baser sur la représentation hiérarchique pour proposer une solution intégrative du temps. Nous renvoyons le lecteur au chapitre I de ce manuscrit s'il veut avoir des détails sur le principe de la méthode que nous n'allons pas réexpliquer maintenant.

Cependant, nous voulons souligner, ici, la parfaite adéquation de la représentation hiérarchique et du modèle basé sur la conception sémantique (introduit au chapitre I) pour l'intégration du temps. Ce modèle permet de re-normaliser chacun des paramètres au travers de la hiérarchie, de plus, il est interdisciplinaire comme nous l'avons montré et a déjà été mis en avant dans une publication de 2006 de Hermsen et Hendricks. Ces auteurs ont souligné le fait qu'il était nécessaire d'utiliser la hiérarchie temporelle.

« Consideration of hypothesis of phylogeny within the context of the hierarchy of time has the potential to bridge research programs in systematic biology and paleontology by placing phylogenetic hypothesis in a temporal framework ». **Hermsen et Hendricks, 2006 :404.**

La solution que nous proposons repose sur le principe de traitement de la paralogie taxique temporelle énoncée par Zaraguëta Bagils et al. en 2004. Dans cet article, les auteurs exposent leur modèle dans lequel le temps est représenté de manière hiérarchique (Figure 30). Dans cette représentation, en lieu et place de l'échelle stratigraphique classique, où chaque période temporelle se suit, le temps n'est plus « borné » par chaque partition. C'est-à-dire que le Crétacé* constitue la classe la moins inclusive et est incluse dans le Jurassique*, lui-même inclut dans le Triassique* et ainsi de suite (Figure 30). Ce qui a pour conséquence, finalement, qu'une période ne se « finit » jamais et que le temps se « lit » aux nœuds, rendant ainsi possible, la détermination des âges minimum de chacun des nœuds internes d'un arbre.

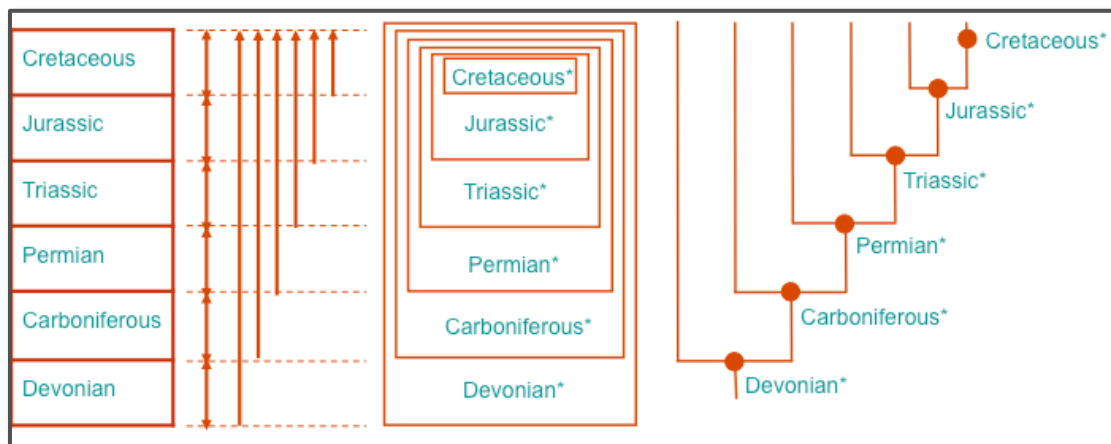


Figure 30 Représentation hiérarchique des intervalles temporelles (durées) (modifié de Zaragüeta Bagils et al. 2004)

Nous l'avons déjà dit, la Biogéographie historique recherche les relations de parenté entre aires d'endémisme à partir de l'analyse de congruence des cladogrammes de taxons. Dans la mesure où nous recherchons un pattern général provenant de plusieurs taxons non apparentés, il est primordial de pouvoir d'abord intégrer le temps ; ces taxons présentent, en effet, des histoires évolutives particulières qui peuvent avoir eu lieu à des temps différents.

L'exemple illustré dans la Figure 31 montre comment l'âge d'un nœud peut être déterminé. En Figure 31-1, 4 taxons dont les âges sont connus (A : 10 Ma, B : 5 Ma ; C : 50 Ma et D : 75 Ma) sont représentés ainsi que leur relation de parenté. L'âge du nœud interne (B, (C, D)) est de 75 Ma. L'introduction d'un nouveau taxon X d'âge 80 Ma ne change pas l'intervalle dans lequel se situe ce nœud interne qui reste bien compris entre 100 et 75 Ma et l'âge du nœud ((B, X), (C, D)) devient 80 Ma. Sur l'exemple Figure 31-2, le taxon introduit a un âge de 40 Ma et à nouveau, cela n'a pas d'incidence, le nœud reliant (B, X) est daté de 40 Ma et le nœud ((B, X), (C, D)) est daté de 75 Ma.

Conformément au principe décrit dans la paralogie taxique temporelle, un nœud est daté de l'âge du taxon le plus ancien qui le constitue.

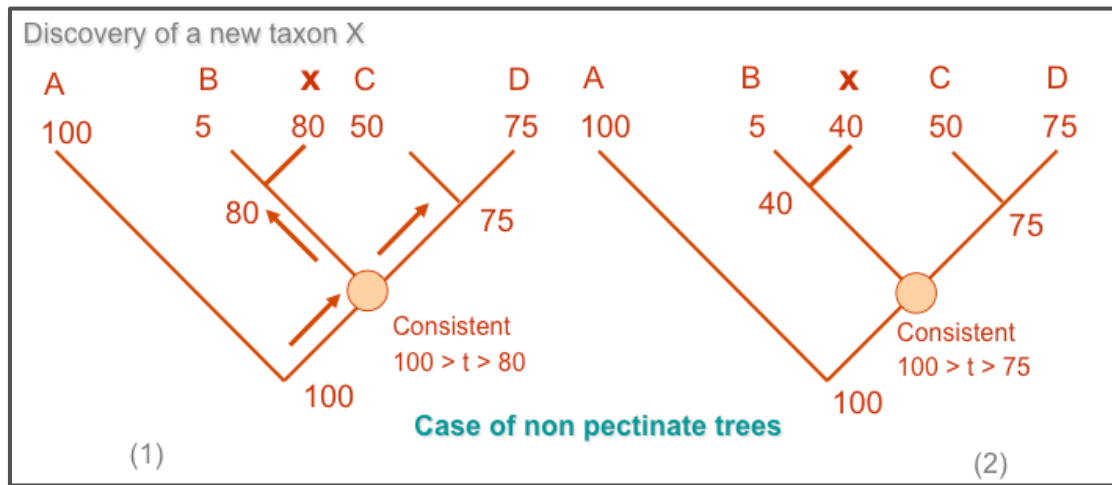


Figure 31 Deux exemples de cladogrammes non pectinés pour lesquels l'intégration d'un nouveau taxon X (d'âges différents : 80 Ma pour 1 et 40 Ma pour 2) dans l'échantillonnage n'ajoute pas d'ambiguïté concernant le temps (d'après Zaragüeta-Bagils et al. 2004).

Ainsi, il est possible de dater chaque nœud d'un cladogramme en partant des terminaux et en remontant jusqu'à la racine de façon à obtenir un arbre « libéré » de toute paralogie taxique temporelle.

Dans l'article suivant, nous présentons notre modèle d'intégration du temps en Biogéographie historique. Il s'agit d'un court commentaire faisant un état des lieux de ce qui est déjà fait et une proposition reposant sur nos développements en 3ia.

4. « Bridging the temporal gap in comparative biogeography »

(Journal of Biogeography, en préparation)

4.1 Contexte de l'article

Afin de répondre à cette question concernant la dimension temporelle, nous avons entrepris de faire une exégèse des principaux articles traitant de ce point précis, ceci nous a conduit à ce court commentaire dans lequel nous proposons un modèle d'intégration de la dimension temporelle.

4.2 Résumé

Malgré une littérature foisonnante sur le traitement du temps en Biogéographie, nous relevons que, finalement, aucune méthode intégrative (au sens où nous l'entendons) n'a été énoncé jusqu'à présent. Beaucoup de modèles

ont été développés mais ils reposent tous sur un partitionnement temporel, ce qui revient à un traitement du temps en synchronie. Nous proposons une approche basée sur la paralogie taxique temporelle, énoncée en 2004 et qui permet de dater les nœuds internes d'un arbre qui sont les seuls porteurs d'information.

4.3 Conclusions

Pour devenir pleinement intégrative, la Biogéographie doit intégrer la dimension temporelle, or, aucune méthode n'existe pour cela. Nous proposons un nouveau modèle/protocole qui, sans être complètement intégratif, présente toutefois, l'avantage de poser les premières bases vers un traitement du temps optimal.

Commentary

Bridging the temporal gap in Biogeography

UNG Visotheary

UMR 7207 (CNRS-UPMC-MNHN), 57 rue Cuvier CP43 75005 Paris.

Word count : 1516

Biogeography has become a discipline of strong interest since many years, it is a comparative science (Parenti & Ebach, 2009) which needs methods of analysing taxa distributions. An explosion of biogeographic methods and applications in the late 70's and early 80's coincided with a numerical revolution in systematics (Parenti & Ebach, 2009). Nevertheless, where is the time ?

Most studies don't consider properly the temporal dimension. To the question: « Why taxa are distributed where they are today? », Nelson and Platnick in 1978 answer it is for two reasons, i.e. two processes can be responsible for these distributions. First, vicariance, implies that taxa have evolved and specialised where they lived, predicting that taxa and the geological barriers between them are of the same age. Second, dispersal, implies a post-speciation after the appearance of a geological barrier thus predicting that the barrier is anterior to the speciation. The main difference lies on the possible age existing between taxa and geological barriers that limit biogeographical areas. Thus, biogeographical studies are often divided into “paleogeography” and “neogeography” (Rosen 1992) respectively dealing with fossil and extant taxa spatiales distributions. We do think this is not the point. Integrating temporal data implies analysing two temporal scales (with extinct and extant taxa) together in a diachronic analysis.

Since integrating the temporal dimension is crucial, several models have been published to this end. Here we have selected some of them deemed essential and we propose our model to achieve the integration of time without ambiguity making it, in a way, « absolut ».

The first one is the « Temporal congruence and cladistic analysis biogeography and co-speciation » model, published in 1990 by Roderic Page where he introduced the parallel between biogeography and co-speciation. Indeed, the relation between a taxon and its areas of endemism can be likened to that between a parasite and its host. Starting from this premise, he reconciles parasites tree to hosts tree when they are not congruent, thus, introduces redundant information (i.e. paralogy) when nodes are split into two nodes to retrieve the same relationships (see Page 1991 for more details).

In 2001, Hunn and Upchurch have stated a set of approaches combining time and space data to constrain interpretations of phylogenetic causality in a “chronobiogeographical paradigm”. Their method uses

phylogenetic relationships in combination with time/space data to produce the simplest and most assumption-free estimate of biogeographical history. They finalize their first model with the “temporally partitioned component analysis” (Upchurch and Hunn 2002). The method incorporates temporal data into conventional cladistic biogeography techniques (COMPONENT (Page, 1993a) and TreeMap (Page, 1994c). First, the method search for “optimal area cladograms” i.e. area relationships providing the “best” explanation for the spatial distributions observed in the taxon cladogram. And then, a randomisation test is used to determine whether the degree of fit between OAC and taxon cladogram topologies is greater that would be expect by chance. This « time-slicing » method makes partitions of the entire time-scale into different time-range and thus, avoids the temporal problem since each analysis is done within the same time-range (i.e synchronic analysis).

Donogue and Moore, in 2003, have stressed, again, the necessity of an “integrative biogeography”, in which the temporal dimension should be integrated as « *it is usually missing from cladistic biogeographic approaches* ». They used the “congruence/pseudo-congruence” model of Cunningham and Collins (1994), i.e. the « test of time ». This phenomenon occurs when two trees have the same topology but a divergence in the timing. Four cases are possible: (1), congruence : topology of trees and timing are congruent (i.e; temporal coincidence); (2), pseudo-congruence : topologies are congruent but not the timing; (3), incongruence : topologies are not congruent even if the timing is correct and (4) pseudo-incongruence : neither the topologies nor the timing are congruent. Here again, they are dealing with synchronic analysis.

In 2008, Peter J. Makovicky, introduced a new method for determining biochronologic order between paleobiotas based on the phylogenetic relationships of their constituent taxa, the Cladistic Biochronologic Analysis (CBA). This method adapted from Brook Parsimony Analysis of Endemicity create a temporal ordering of paleobiotas. The rationale is that time can be conceived as a continuous, irreversible character on a correctly rooted cladogram that is interpreted as an evolutionary tree. The root-to-tips order of nodes represents the temporal information of a cladogram and need a congruent fossil record with evolutionary patterns determined from organisms.

The unit of CBA is a paleobiota, with distinct temporal ranges through time.

A matrix is compiled with paleobiotas in place of taxa and syntaxa (i.e. internal nodes) in place of characters. With the difference that terminal taxa are not included in CBA because the temporal information accross terminal taxa is not hierarchical. As in BPA, coding of syntaxon information follows the additive binary method; such that a syntaxon is coded as present if one or more of its descendants (including apical syntaxa) is present and recognized within a paleobiota. Then, the coded matrix is analyzed using maximum parsimony in order to find the tree(s) that maximize the congruence between syntaxon information. These chronological trees, i.e. cladochronograms, are rooted either on an all-zero ancestor as in BPA or on an available paleobiota known to pre-date all other biotas. Makovicky claimed that his method may be advantageous in situations where traditional stratigraphic data is insufficient or unavailable for some or all of the paleobiotas under study.

Temporal paralogy

The genuine question of historical biogeography to be integrative is how to deal with diachronic data ? Historical biogeography should be integrative *per se*. However, all the aforementioned methods do not integrate the temporal dimension, on the contrary, their solution is to apply the « time-slicing » solution, i.e. reducing a large temporal scale into several ones according to the timing information available.

We would like to suggest a protocol based on the concept of temporal paralogy (Zaragüeta-Bgils et al. 2004). This approach has already been highlighted by Hermsen and Hendricks in 2006. This framework offers a rigorous conceptual model for time integration. Better than being a arrow (i.e. the geological scale), time is represented as a hierarchy where one classe (i.e. one time period) is included into another (except the less inclusive one, in the example of the Figure 1: Cretaceous*) and so on. This mathematical representation of time accounts for being a pectinate tree, thus continuous.

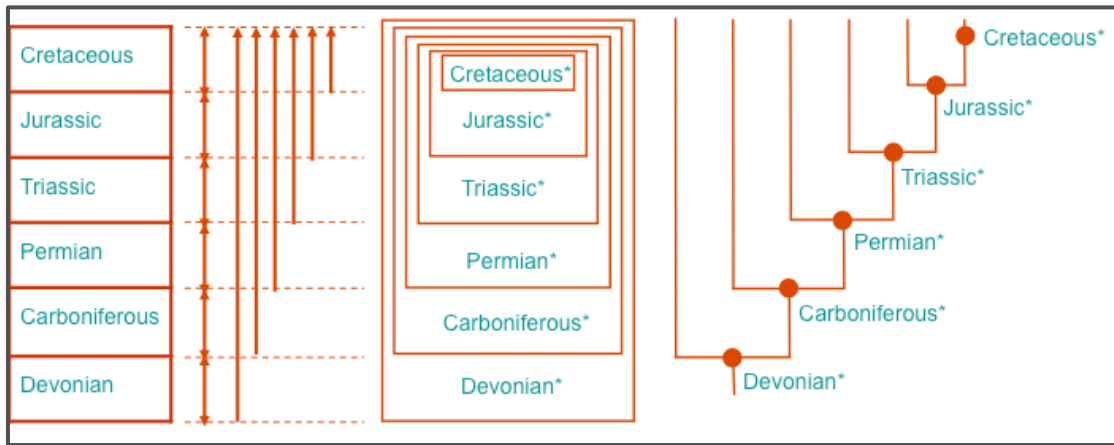


Figure 1 : Hierarchical representation of time (modified from Zaragüeta Bagils et al. 2004).

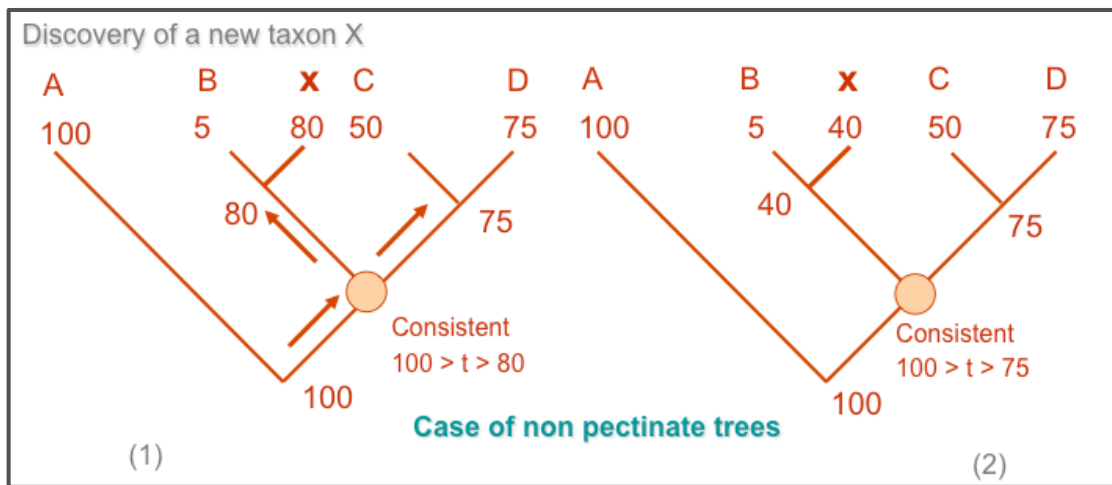


Figure 2 : Two cases of non pectinate trees for each a new taxon is added without changing the interval for the nodes (Zaragüeta-Bagils, comm. pers.). (1) a new taxon X of 80 Ma is closer to taxon B dated at 5 Ma, the age of the node (B, C) is 80 Ma so as the node ((B, X), (C, D)). (2) the age of taxon X is 40 Ma, and the age of the node ((B, X), (C, D)) is 75 Ma.

Figure 2 shows how to treat paralogous nodes presenting two additional taxa from different ages. Whenever two taxa from different ages are present at the node, the estimation of age of the most closely related node is the one of the oldest taxa. This procedure free the cladogram from any temporal paralogy rendering it comparable with others without splitting it into different time-slices.

The main advantage of this model is its knowledge representation which can be extended to any discipline using tree-like representation (e.g. Systematics).

Consequently, we suggest the following algorithm :

- (1) Harvesting and selection of cladograms distributed on the area of interest with (possibly) their ages plus their distribution;
- (2) Identification of areas of endemism;
- (3) Time treatment, following the temporal paralogy concept, ages of the nodes are determined then;
- (4) Replacement of the terminals by the corresponding areas;
- (5) Treatment of the geographic paralogy and widespread taxa via LisBeth (Zaragüeta-Bagils et al 2012) ;
- (6) Congruence analysis via LisBeth (Zaragüeta-Bagils et al 2012) ;
- (7) Interpretation of the reconstructed areagram in the light of the geological history of the region.

Since we do not present a solution yet, this protocol still need improvements and testing, nevertheless, we do think that Historical Biogeography needs to rely on the temporal paralogy paradigm (Zaragüeta-Bagils et al. 2004) in order to be fully integrative.

References

- Donoghue M.J. & Moore B.R. (2003) Toward an integrative historical biogeography. *Integr.Comp.Biol.*, **43**, 261-270.
- Hennig, W. (1966) Phylogenetic systematics. Univ. of Illinois Press, Urbana-Champaign.
- Hermesen E.J. and Hendricks J. R. (2006) The hierarchy of time. *Palaios*, **21**, 403-405.
- Hunn C.A. and Hupchurch P. (2001) The importance of time/space in diagnosing the causality of phylogenetic events: towards a "chronobiogeographical" paradigm? *Syst. Biol.* **50**(3), 391-407.
- Makovicky Peter J. (2008) Telling time from fossils : a phylogeny-based approach to chronological ordering of paleobiotas. *Cladistics*, **24**, 350-371.
- Page R.D. (1990) Temporal congruence and cladistic analysis of Biogeography and cospeciation. *Syst. Zool.*, **39** (3), 205-226.
- Page R.D. (1993a) COMPONENT: Tree comparison software for Microsoft windows, version 2. Natural History Museum, London.
- Page R.D. (1994c) TreeMap. Release 3.1 Oxford. University of Oxford.
- Parenti, L.R. & Ebach, M.C. (2009) Comparative Biogeography: Discovering and Classifying Biogeographical Patterns of a Dynamic Earth. University of California Press, Berkeley.
- Rosen B. R. (1992) Empiricism and the biogeographical black box : concepts and methods in marine paleobiogeography. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **92**, 171-205.
- Upchurch P. and Hunn C.A. (2002) "Time": the neglected dimension in cladistic biogeography? *Geobios*. Mémoire special n°24.
- Zaragüeta-Bagils, R. Lelièvre H. and Tassy P. (2004) Temporal paralogy, cladograms and the quality of the fossil record. *Geodiversitas* **26** (3), 381-389.
- Zaragüeta Bagils R., Ung, V., Grand, A., Cao, N. & Ducasse, J. (2012) LisBeth: New cladistics for phylogenetics and biogeography. *Comptes Rendus Palevol*, **11**, 563-566.

Conclusions et perspectives

1. Conclusions

Comme nous l'avons vu au début de ce chapitre, plusieurs tentatives d'intégration du temps ont été formulées, sans pour autant apporter de véritable solution au problème. Néanmoins, tous s'accordent à dire que cette entreprise exige un algorithme/modèle/protocole rigoureux pour le mener à bien.

Nous retiendrons que la dernière approche de Makovicky (la CBA) est celle qui se rapproche le plus de ce que nous proposons en ce sens qu'elle reprend le caractère hiérarchique du temps. Seulement, la méthode de la BPA n'étant pas optimale voire même sujette à des reconstructions fausses (Ebach et al. 2005), nous lui préférons la méthode 3ia qui est, à ce jour, la seule à proposer un cadre conceptuel d'analyse formel, contraint et interdisciplinaire.

Nous suggérons la procédure décrite dans l'Encadré 17 pour toutes analyses biogéographiques diachroniques pour lesquelles il est indispensable de prendre en considération la dimension temporelle avant toutes autres étapes.

Encadré 17 Procédure proposée pour une analyse biogéographique intégrative

- (1) Sélection de cladogrammes de taxons distribués sur l'aire d'intérêt avec si possible toutes les datations (âges de diversification des taxons) ainsi que leur distribution;
- (2) Détermination des aires d'endémismes;
- (3) Traitement du temps, suivant la paralogie temporelle (**Figure 31**), les âges des nœuds sont ainsi déterminés;
- (4) Remplacement des taxons terminaux par les aires d'endémismes correspondantes ;
- (5) Traitement de la paralogie géographique et des taxons à large répartition via LisBeth ;
- (6) Analyse de congruence via LisBeth ;
- (7) Interprétation de l'aérogramme obtenu en regard de l'histoire géologique connue de la région d'étude.

2. Perspectives

Alors même qu'une solution s'ébauche quant à l'intégration du temps dans les analyses biogéographiques cladistiques ; malgré tout, il n'en reste pas moins que la principale difficulté se rapporte à la disponibilité des données. Effectivement, pour pouvoir mener des analyses de Biogéographie historique, il faut des cladogrammes de taxons et leurs distributions précises associées. S'ajoute à cela les datations si des fossiles sont intégrés dans un cladogramme de taxons actuels et/ou fossiles.

Par conséquent, dans le but de valider notre protocole, il est absolument essentiel de réaliser des tests de deux ordres.

D'un côté des tests théoriques seraient à effectuer sur un jeu de données connues et pour lesquels nous avons déjà connaissance des histoires évolutives. Et de l'autre, des analyses empiriques exigeant de recueillir un échantillonnage de taille suffisante, présentant les critères exposés précédemment, qui renseigneraient sur la pertinence et la véracité de notre approche. Ce que nous ne manquerons pas d'explorer très prochainement.

Malgré tout, avons-nous répondu à la question introduite au début de ce chapitre ? Nonobstant la proposition d'une procédure, il faut reconnaître que nous ne savons toujours pas comment procéder à une analyse biogéographique lorsque, au sein d'un même cladogramme de taxons sont présents des taxons actuels et des taxons fossiles. *Quid* de l'aire du fossile qui n'existe plus maintenant ? Par quoi pouvons-nous la remplacer ? Cette question, fondamentale, en Biogéographie historique reste en suspens, mais nous sommes convaincus que la première étape de la prise en compte du temps repose sur l'implémentation de la paralogie taxique temporelle et de la représentation hiérarchique des durées temporelles.

Conclusion générale

Conclusion générale

Depuis les premiers travaux de Linné du XVIIème siècle, la Biogéographie et en particulier la Biogéographie historique n'a cessé d'évoluer : on dénombre maintenant une profusion de méthodes destinée à mettre au jour l'histoire évolutive des aires (Crisci et al. 2003, Morrone 2009) et par là même une (des) explication (s) concernant la distribution des taxons sur le globe terrestre.

Nous avons introduit au tout début de ce manuscrit trois questions fondamentales auxquelles la Biogéographie tente de répondre : *pourquoi, comment* et *quand* ?

Nous avons également présenté notre démarche en proposant de répondre aux questions *pourquoi* et *comment* dans les chapitres I et II ; de répondre au *quand* dans le chapitre III. Si les chapitres I et II apportent bien une réponse quant au *comment*, le chapitre III lui, répond bien aussi au *quand*.

Qu'en est-il du *pourquoi* ?

L'observation des différences de distributions taxiques est le préalable à toute analyse biogéographique. La question du *pourquoi* est donc tout à fait pertinente, seulement, il n'est pas possible d'y répondre directement. En réalité, on en arrive même à la conclusion que répondre à cette question, finalement, n'a pas de sens. On observe effectivement une distribution spécifique des taxons avec des taxons endémiques, des taxons à large répartition, des taxons à répartition plus restreinte mais non endémiques etc... Cette différence de distribution constitue les objets de base d'une analyse (nous parlons d'OTUs dans le chapitre I) biogéographique, nous partons des cladogrammes de taxons et de leur distribution. Et c'est précisément *parce qu'il y a* une telle distribution des taxons que nous pouvons en déduire une histoire évolutive des aires dans lesquelles ils se trouvent. D'ailleurs, le modèle hiérarchique que nous proposons, basé sur la conception sémantique (voir chapitre I, Ung et Grand, en préparation) met bien à jour ce point de détail : ce qui est intéressant c'est de définir une aire d'endémisme et ensuite de déterminer quelles sont les relations de parenté existant entre-elles par le biais de méthodes d'analyses (chapitre I, Zaragütea-Bagils et al. 2012) et d'en inférer une histoire évolutive des aires. Il n'est donc jamais possible de répondre à la question *pourquoi*, ni même souhaitable, *parce que* des centaines d'explications pourraient être évoquées pour y répondre. D'ailleurs, si l'on cherche du côté de la Biogéographie évolutive *sensu* Parenti et Ebach, on remarque que souvent, les discours sont d'ordre narratif comprenant des explications sur des patterns de dispersion à partir d'un centre d'origine (par exemple Richardson et al 2012, Baker et Couvreur 2012 ou encore Webb et Ree 2012 pour l'Asie du Sud-Est, Ung 2013). Là encore, le modèle ontologique nous rappelle que nous ne répondons pas à la bonne question car une analyse de Biogéographie comparative recherche des relations de parenté entre aires et non pas entre taxons, ceux-ci et plus particulièrement leur cladogrammes nous informent sur l'histoire des aires mais ne dit rien sur leur histoire propre, certains iraient même jusqu'à dire que ce sont des explications *ad hoc*.

En définitive, pour pouvoir un tant soit peu, valider le cladogramme d'aires obtenu, il ne reste que la géologie et les reconstructions paléogéologiques (chapitre II) qui, soit dit en passant, ne sont que des interprétations et donc inhérentes au (paléo) géologue qui le propose (Hall 1998). Dans notre cas d'étude, le Sud-Est asiatique et l'Ouest du Pacifique, sont bien documentés en ce qui concerne leurs reconstructions géologiques, nous pouvons donc nous y référer pour nous aider dans l'interprétation du résultat que nous obtenons (Hoagstrom et al. 2013 en révision, Ung et al 2013 en révision).

1. Les apports de la 3ia à la Biogéographie

Il était indispensable de dresser le cadre méthodologique de nos recherches, et clairement, nous posons la 3ia comme une méthode dorénavant incontournable pour la recherche de patterns généraux. Nous irons même plus loin en disant que plus que la 3ia qui finalement n'est que la méthode d'analyse que nous utilisons, nous insistons sur le fait que c'est la représentation des caractères sous forme hiérarchique qui revêt ici l'intérêt majeur de notre démonstration. Représentation qui est conforme au modèle ontologique que nous proposons. Cette représentation a été mise en application dans le logiciel LisBeth, le seul à implémenter la représentation hiérarchique des caractères et la méthode 3ia dans son intégralité. Dans ce cadre, il manquait une interface utilisateur graphique afin de rendre son utilisation simplifiée par les utilisateurs. Ces améliorations sont en cours d'intégration actuellement suite à nos propositions.

Nous rappelons que dans le cadre de la recherche de patterns généraux, nous testons des hypothèses d'homologies d'aires d'endémisme et que nous partons du principe que des évènements communs (globaux ou généraux) sont dus à des causes communes, ici des évènements de vicariance. En aucun cas, tout autre cause de la distribution taxique n'est nié (dispersion et/ou extinction) seulement ce sont des évènements que nous ne pouvons tester. C'est pourquoi notre propos reste dans le cadre d'une approche de Biogéographie historique et cladistique. Pour ce qui concerne le Sud-Est asiatique et l'Ouest du Pacifique, nous avons clairement mis en évidence un signal biogéographique qui nous indique une séparation nette entre le Sud-Est asiatique à proprement parlé (finalement presque tout le Sundaland) et toutes les îles du Pacifique (plus la Nouvelle-Guinée) et qui correspond à la ligne de Lydekker (ligne tracée en suivant le pourtour du continent australien). En définitive, qu'en est-il de toutes ces lignes tracées ? Est-il si pertinent d'en tracer ?

Nos résultats semblent indiquer que la question n'est pas si simple que cela. Les diverses lignes suivent des distributions faunistiques différentes et aucune ne prend en compte la flore locale. Il est évident que l'on observe des différences de distribution taxique (tant faunistique que florale) sur toute la région, seulement il faudrait un échantillonnage beaucoup plus exhaustif pour pouvoir avoir une réponse tranchée. Nous avons défini un échantillonnage composé de 35 taxons non-apparentés mélangeant des plantes et des animaux (insectes, lézards et oiseaux) mais nous ne n'avons pas intégré par exemple, de cladogrammes de petits et/ou moyens-grands mammifères, qui sont eux beaucoup plus assujettis à des évènements récents (et aussi en partie liés à une activité anthropomorphique). La question de « l'existence » des lignes comme barrières biogéographiques ne

semblent pas répondre à un processus commun mais est plutôt taxon-dépendant. Il serait d'ailleurs intéressant de reprendre des cladogrammes de taxons ayant servis à tracer ces différentes lignes (s'ils existent) et voir ce que la 3ia nous donne comme résultat, si ces lignes sont retrouvées ou non. Certes, nos résultats sont pour le moment préliminaires et gagneraient à être étoffer d'autres cladogrammes de taxons, néanmoins ils sont prometteurs car un véritable signal biogéographique en ressort, identifiant plusieurs « clades monophylétiques » (voir chapitre 2) que nous avons nommés selon leur localisation géographique actuelle. Pour autant, il ne faut pas les considérer comme des aires géographiques mais bien comme des aires biogéographiques c'est à dire correspondant à la présence d'un ensemble de « taxons synapomorphiques ». Cependant, il faut souligner que l'utilisation ici des termes « clade monophylétique » et « synapomorphie » n'est fait que selon un pur parallèle avec la Systématique. Ce ne sont pas des « synapomorphies vraies » *sensu* Brundin (Brundin 1966 dans Williams 2004), replacées sur l'arbre une fois l'analyse de congruence terminée, c'est pour cela que nous parlons de « taxons synapomorphiques » et de « clades » avec des guillemets. Il est nécessaire que nous améliorions encore LisBeth de manière à ce que le mode « interprétation » (Zaragüeta Bagils et al. 2012) qui remplace les synapomorphies aux nœuds soit disponible pour la Biogéographie. Pour le moment seule la traçabilité des caractères est possible et permet de replacer les taxons dont sont issus les 3is ayant servis à reconstruire l'arbre minimal aux nœuds (Zaragüeta-Bagils et al. 2012, Ung et al. 2013, en révision). Or le fait de replacer les « taxons synapomorphiques » aux nœuds est à l'heure actuelle uniquement réalisé par LisBeth et c'est une discussion totalement nouvelle dans la Biogéographie comparative d'aujourd'hui. Aucune analyse de Biogéographie historique actuelle ne propose de discuter des « synapomorphies » soutenant les nœuds des aréogrammes. Or il s'agit bien de cela, lorsque nous réalisons une analyse de Biogéographie historique, nous testons des hypothèses d'homologie d'aires biogéographiques, nous recherchons donc les synapomorphies qui permettent de soutenir tel ou tel regroupement d'aires. Et de ce fait, nous pouvons inférer de nouvelles relations de parenté entre aires d'endémismes. Pour la première fois, nous sommes en mesure d'éclairer le champ de la Biogéographie comparative d'un apport novateur qui s'intéresse au cœur même de la problématique. Ce qui nous permet ensuite de discuter de l'existence d'une classification naturelle des aires étudiées. Pour ce qui concerne le Sud-Est asiatique, nous envisageons de poursuivre cette recherche en explorant les différents points de perspectives énoncés au chapitre II.

2. Profession biogéographe ?

Malgré une existence des plus anciennes, la discipline ne compte que très peu de spécialistes. Il n'existe pas beaucoup de chercheurs qui se revendiquent biogéographes, et pourtant la plupart des grands noms de la biologie ont eu leur interrogations et apports sur la discipline tels de Candolle, Haeckel, Darwin ou encore bien sûr Wallace. Mais aucun ne se présentaient comme biogéographes véritablement. Or, il est indispensable que la Biogéographie trouve ses lettres de noblesse et devienne une discipline à part entière dans la recherche

moderne, comme le font très justement remarquer Parenti et Ebach dans leur livre « Comparative Biogeography ».

Maintenant que nous avons enfin pu mettre en évidence la pertinence de la 3ia, entre autre pour la Biogéographie, que nous disposons d'un logiciel, LisBeth, nous espérons que de plus en plus de biogéographes s'affirmeront et revendiqueront cette profession.

3. Vers une Biogéographie intégrative

Ainsi que nous l'avons déjà mentionné précédemment dans le manuscrit, outre la nécessité d'exister en tant que telle, la Biogéographie se doit d'être, aujourd'hui, intégrative. Intégrative car elle doit tenir compte des échelles temporelles (chapitre III), intégrative parce qu'elle est modélisée par une ontologie formelle de représentation des connaissances (chapitre I) et surtout intégrative car elle fait le lien entre la Systématique et la distribution des taxons, en ce sens, elle ouvre la voie à une nouvelle ère. Nous ne manquerons pas de poursuivre dans cette voie.

Dans un milieu scientifique où la compétition fait rage, la Biogéographie comparative intégrative pourrait bien sortir son épingle du jeu en offrant une nouvelle dimension aux analyses biogéographiques standard. En plus de la recherche d'une classification naturelle des aires d'endémismes, la Biogéographie comparative intégrative pourrait être un complément non négligeable dans l'aide à la délimitation des aires protégées telle que nous l'avons présenté au chapitre I et II (Ung et Grand 2013 en préparation, Ung et al 2013 en révision).

À l'heure où la conservation de la diversité biologique est sur toutes les lèvres, que les regards sont tournés vers les organismes mondiaux de protection de cette même diversité, une nouvelle approche, conceptuelle, formelle, fondée sur des ontologies rigoureuses pourrait bien apparaître comme annonciateur de bonnes perspectives pour le futur.

Epilogue

Epilogue

Nous voici parvenu à la fin d'un long périple qui nous a transporté dans les méandres de la méthodologie des plus théoriques aux portes de l'Asie. J'espère que chaque lecteur aura apprécié ce voyage.

Contrairement à tout le corps du texte, je souhaite ici, reprendre la parole et m'exprimer à la première personne. L'expérience de la réalisation d'une thèse est un exercice très difficile et pourtant si riche en enseignements. Tout particulièrement lorsque l'on touche à la Biogéographie, cette discipline à la croisée des chemins d'autant d'autres. J'espère avoir apporté ma pierre à l'édifice et montré que la Biogéographie était une science à part entière, d'ailleurs si riche et complexe.

Et pour ne pas quitter l'Asie et mon Cambodge natal, je voudrais aussi finir par une note un peu moins scientifique et un peu plus légère en commentant et illustrant un texte d'Alfred Russell Wallace sur ce merveilleux fruit qu'est le Durian !

"The Durian grows on a large and lofty forest-tree, something resembling an Elm in character, but with a more smooth and scaly bark. The fruit is round or slightly oval, about the size of a small melon, of a green colour, and covered with strong spines, the bases of which touch each other, and are consequently somewhat hexagonal, while the points are very strong and sharp. It is so completely armed that if the stalk is broken off it is a difficult matter to lift one from the ground. The outer rind is so thick and tough that from whatever height it may fall it is never broken. From the base to the apex five very faint lines may be traced, over which the spines somewhat curve and approximate; these are the sutures of the carpels, and show where the fruit may be opened with a heavy knife and a strong hand. The five cells are silky-white within, and are filled with a mass of firm, cream-coloured pulp, containing about three seeds each. This pulp is the eatable part, and its consistence and flavour are indescribable. A rich custard highly flavoured with almonds gives the best general idea of it, but there are occasional wafts of flavour that call to mind cream-cheese, onion-sauce, sherry-wine, and other incongruous dishes. Then there is a rich glutinous smoothness in the pulp which nothing else possesses, but which adds to its delicacy. It is neither acid nor sweet nor juicy; yet it wants neither of these qualities, for it is in itself perfect. It produces no nausea or other bad effect, and the more you eat of it the less you feel inclined to stop. In fact, to eat Durians is a new sensation worth a voyage to the East to experience. . . . as producing a food of the most exquisite flavour it is unsurpassed. If I had to fix on two only as representing . . . perfection . . . I should certainly choose the Durian and the Orange as the king and queen of fruits." **A.R. Wallace, On the Bamboo and Durian of Borneo, 1856, Hooker's Journal of Botany 8: 225-230, on pages 228-229.**

La description de Wallace est très proche de la réalité et les amateurs de ce fruit si différent ont du mal à décrire les sensations que procure sa dégustation. En voici ma version cambodgienne, un dessert typique qui exhale les parfums de ce délicieux fruit !

Crème de Durian au lait de coco sur son lit de riz gluant (pour 4 personnes)



Pour le riz gluant au lait de coco:

Il vous faut : 150 gr de riz gluant, un morceau de sucre de palme, 100 ml (ou plus) de lait de coco.

Ensuite, voici les différentes étapes :

1. Faire macérer 150 gr de riz gluant dans un récipient contenant de l'eau pendant une nuit (ou à défaut au moins une heure) ;
2. Le faire cuire à vapeur 40 minutes jusqu'à ce qu'il soit bien tendre et le laisser refroidir ;
3. Pendant ce temps, faire chauffer le lait de coco et y ajouter le sucre de palme, remuer jusqu'à ce que le sucre soit bien dissous ;
4. Une fois le riz refroidi, ajouter le lait de coco sucré petit à petit, le riz doit avoir une texture moelleuse mais ne doit pas être liquide.

Pour la crème de Durian :

Il vous faut : le contenu de 2 carpelles de Durian, 250 ml de lait de coco, deux morceaux de sucre de palme.

1. Séparer la chair des différents morceaux du Durian et ôter les noyaux ;
2. Faire chauffer le lait de coco avec le sucre de palme, remuer jusqu'à totale dissolution ;
3. Y rajoutez les morceaux de fruits et laisser cuire à feu doux pendant une quinzaine de minutes. Attention, le durian est un fruit fragile et ne supporte pas d'être porté à trop forte température et à une cuisson trop longue. Il faut impérativement bien le surveiller !

Il ne vous reste plus qu'à le servir !

Ce dessert typiquement cambodgien peut se décliner avec d'autres fruits : banane, longane, châtaigne d'eau ou encore fruit du jacquier selon votre inspiration et les fruits de saison !

Autre variante, si vous avez une belle mangue mûre, vous pouvez aussi la déguster avec la préparation de riz gluant à laquelle vous rajouterez des grains de soja légèrement torréfiés.

Bon appétit !

Références bibliographiques

Références bibliographiques

- Avise J.C. 2000. Phylogeography : the history and formation of species. Harvard University Press, Cambridge, Massachussets.
- Axelius B. 1991. Areas of distribution and areas of endemism. *Cladistics*, **7**, 197-199.
- Baker W.J. and Couvreur T.L.P. 2012. Biogeography and distribution patterns of Southeast Asian palms. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 164-190.
- Bellwood D.R., Renema W. and Rosen B.R. 2012. Biodiversity hotspots, evolution and coral reef biogeography : a review. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 16-32.
- Bourdon E. 2006. L'avifaune du Paléogène des phosphates du Maroc et du Togo. Diversité, systématique et apports à la connaissance de la diversification des oiseaux modernes (Neornithes). Thèse de doctorat (MNHN, Paris France).
- Bremer K. 1992. Ancestral areas : a cladistic reinterpretation of the center of origin concept. *Systematic Biology*, **41**, 436-445.
- Bremer K. 1995. Ancestral areas : optimisation and probability. *Systematic Biology*, **44**, 255-259.
- Briggs J.C. 1974. Operation of marine zoogeographic barriers. *Systematic Zoology*, **23**, 248-256.
- Brooks D.R. 1981. Hennig's parasitological method – a proposed solution. *Systematic Zoology*, **30**, 229-249.
- Brooks D.R. 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Zoology*, **39**, 14-30.
- Brundin L. 1966. Transantarctic relationships and their significance, as evidenced by chironomid midges, with a monograph of the subfamilies Podonominae and Aphroteniinae and the austral Heptagylae, K. *Sven. Vetenskapsakad. Handl.*, **4**(11), 1-472.
- Cannon C.H. 2012. Quaternary dynamics of Sundaland forests. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 115-137.
- Cao N. 2008. Analyse à trois éléments et anatomie du bois des Fagales Engl. Thèse de doctorat (MNHN, Paris, France).
- Cao N., Zaragüeta Bagils R., Vignes-Lebbe R. 2007. A hierarchical representation of the hypothesis of homology. *Geodiversitas*, **29** (1), 5-15.
- Cao N., Bourdon E., El Azawi M., and Zaragüeta-Bagils R. 2009. Three-item analysis and parsimony, intersection tree and strict consensus: a biogeographical example. *Bulletin de la Société Géologique de France*, **180** (1) : 13-15

- Cecca F., Morrone J.J. and Ebach M. C. 2011. Biogeographical convergence and time-slicing: concepts and methods in comparative biogeography. In *P. Upchurch; A. McGowan and C. Slater (eds.), Palaeogeography and Palaeobiogeography: Biodiversity in Space and Time*. Taylor and Francis (CRC Press), Boca Raton, Florida, 1-12.
- Conservation International 2005. Biodiversity hotspots, *CI facts*, 2p.
- Corvez A. 2012 L'origine de la mégaphylle chez les Monilophytes. Thèse de doctorat (MNHN, Paris, France).
- Craw R.C., Grehan J.R. and Heads M.J. 1999. Panbiogeography: tracking the history of life. Oxford University press, New York.
- Crisci J.V., Katinas L. and Posadas P. 2003. Historical biogeography: an introduction. Harvard University Press, Cambridge, Massachusetts.
- Croizat L. 1964. *Space, Time, Form: the biological synthesis*. Published by the author, Caracas.
- Cunningham C.W. and Collins T.M. 1994. Developing model systems for molecular biogeography: Vicariance and interchange in marine invertebrates. In *B. Schiewater, B. Streit, G.P. Wagner, and R. DeSalle (eds), Molecular Ecology and Evolution: approaches and applications*, 405-433. Birkhauser Verlag, Basel, Switzerland.
- de Boer A.J. (1995c) Islands and cicadas adrift in the West-Pacific. Biogeographic patterns related to plate tectonics. *Tijdschrift Voor Entomologie*, **138**: 169-244.
- de Boer A.J. and Duffels J.P. (1996) Historical biogeography of the cicadas of Wallacea, New Guinea and the West Pacific: a geotectonic explanation. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **124**(1-2), 153-177.
- de Bruyn M., von Rintelen T., von Rintelen K., Mather P.B. and Carvalho G.R. 2012. Molecular biogeography and phylogeography of the freshwater fauna of the Indo-Australian Archipelago. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 316-347.
- Donoghue M.J. and Moore B.R. 2003. Toward an integrative historical biogeography. *Integr.Comp.Biol.*, **43**:261-270.
- Donoghue M.J., Bell C.D. and Li J. 2001. Phylogenetic patterns in Northern Hemisphere plant biogeography. *Int. J. Plant. Sci.* **162**: S41-S52.
- Duffels J.P. (1986) Biogeography of indo-pacific cicadoidea: a tentative recognition of areas of endemism. *Cladistics*, **2**, 318-336.
- Estabrook G.F., Johnson C.S., Mc Morris F.R., 1976. A mathematical foundation for the analysis of cladistic character compatibility. *Math. Biosci.* **29** (1-2), 181-187.
- Ebach M.C. and Humphries C.J. 2002. Cladistic biogeography and the art of discovery. *Journal of Biogeography*, **29**, 427-444.
- Ebach M.C., Humphries C.J., Newman R.A., Williams D.W., Walsh S.A., 2005. Assumption 2: opaque to intuition? *Journal of Biogeography*, **32**, 781-787.
- Ebach M. C., Morrone J. J., Parenti L. R and Vilorio Á. L. 2008. International Code of Area Nomenclature. *Journal of Biogeography*, **35**, 1153-1157.

- Fox D.L.; Fisher D.C and Leighton L.R. 1999. Reconstructing phylogenies with and without temporal data. *Science*, **284**, 1816-1819.
- Frodin David G.; Lowry II, Porter P., Plunkett Gregory M. 2012 Schefflera (Araliaceae): taxonomic history, overview and progress. *Plant Diversity and Evolution*, **128** (3-4), 561-595.
- Ghiselin M. T. 1974. A radical solution to the species problem. *Systematic Zoology*, **23**, 536-544.
- Goetsh L., Eckert A.J. and Hall B.D. 2005. The molecular systematics of *Rhododendron* (Ericace): a phylogeny based upon RPB2 gene sequence. *Systematic Botany*, **30** (3), 616-626.
- Gower D., Johnson K., Richardson J., Rosen B., Rüber L. and Williams S. Biotic Evolution and Environmental Change in Southeast Asia. —. Cambridge: Cambridge University Press, 2012. xv+475 pp ISBN 978-1-107-00130-5 £70 \$US115 (hardback).
- Grand A., Corvez A, Duque Velez L.M., Laurin M. 2013 Phylogenetic inference using discrete characters: performance of ordered and unordered parsimony and of three-item statements. *Biological Journal of the Linnean Society* (in press).
- Grande L. 1985. The use of paleontology in systematics and biogeography, and a time control refinement for historical biogeography. *Paleobiology*, **11**(2), 234-243.
- Hall R. 1988. Basement and cover rock history in western Tethys: HT-LP metamorphism associated with extensional rifting of Gondwanaland. In : *Gondwana and Tethys*. Audley-Charles, M. G. and Hallam, A. (eds.). Vol. 37, p. 41-50. (Special Publication of the Geological Society London).
- Hall R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, **20**, 353-434.
- Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea: Journal of Plant Taxonomy and Plant Geography*, **54**, (1-3), 148-161.
- Hall R. 2011. Australia-SE Asia collision: plate tectonics and crustal flow. The SE Asian gateway: history and tectonics of Australia-Asia collision (eds R. Hall; M. A. Cottam; M. E. J. Wilson), Geological Society, London, Special publications, **355**, 75-109.
- Hall R. 2012. Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 32-78.
- Hall R. and Wilson M.E.J. 2000. Neogene sutures in eastern Indonesia. *Journal of Asian Earth Sciences*, **18**, 787-814.
- Hall R., Clements B. and Smyth H. R. 2009a. Sundaland: basement character, structure and plate tectonic development. In: Proceedings Indonesian Petroleum Association, 33rd Annual Convention, IPA09-G-134, 1-27.
- Hall R., Cottam M.A. and Wilson M. E. J. (2011) The SE Asian gateway: history and tectonics of the Australian-Asia collision. (eds R. Hall; M.A. Cottam; M.E.J. Wilson), Geological Society, London, Special publications, **355**, 1-6.

- Harold A. and Mooi R. D. 1994. Areas of endemism: definition and recognition criteria. *Systematic Biology*, **43** (2), 261-266.
- Hennig W., 1966. Phylogenetic systematics. University of Illinois Press, Chicago, Illinois, 264 p.
- Hermesen E. J. and J. R. Hendricks. 2006. Spotlight: The hierarchy of time. *Palaios*, **21**, 403-405.
- Hill K.D. and Stevenson D.W. 1998. The Cycad Pages Website: <http://plantnet.rbgsyd.nsw.gov.au/PlantNet/cycad/>
- Holloway J. D. 1979. A survey of the Lepidoptera, Biogeography and Ecology of New Caledonia. *Series Entomologica* **15**. The Hague: W. Junk.
- Hovenkamp P. 1997 Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics*, **13**, 67–79.
- Hull D. L. 1978. A matter of individuality. *Philosophy of Science*, **45**, 335-360.
- Humphries C.J. 2004. From dispersal to geographic congruence: comments on cladistic biogeography in the twentieth century. Chapter 10 In *Milestones in Systematics* (eds. Williams, D. M. and Forey, P. L.), special volume series 67. Pp. 225-260. CRC Press Boca Raton, USA. .
- Humphries C.J. and Parenti L.R. 1999. Cladistic biogeography. Interpreting patterns of plant and animal distributions. Oxford Biogeography Series No. 12. Oxford University Press, Oxford.
- Hunn C.A. and P. Hupchurch. 2001. The importance of time/space in diagnosing the causality of phylogenetic events: towards a “chronobiogeographical” paradigm? *Systematic Biology*. **50**(3), 391-407.
- Hutchison C.S. 2005. Geology of North-West Borneo. Elsevier, Amsterdam.
- Hutchison C.S., Bergman S.C., Swauger D.A. and Graves J.E. 2000. A Miocene collisional belt in north Borneo: uplift mechanism and isostatic adjustment quantified by thermochronology. *Journal of the Geological Society*, **157**, 783–793.
- Huxley T.H. 1868. On the classification and distribution of the Alectoromorphae and Heteromorphae. *Proceedings of the Zoological Society of London*, 294-319.
- Kluge A.G. 1988 Parsimony in vicariance biogeography – a quantitative method and a Greater Antillean example. *Systematic Zoology*, **37**, 315-328.
- Lamarck J.B.P.A.D.M.D. and Candolle A.P.D. 1805. Flore française, ou descriptions succinctes de toutes les plantes qui croissent naturellement en France : Disposées selon une nouvelle méthode d’analyse, et précédées par un exposé des principes élémentaires de la botanique. Desray, Paris.
- Lelièvre H., Zaragüeta-Bagils, R. and Rouget I. 2008. Temporal information, fossil record and phylogeny. *Comptes-Rendus de Palevol*, **7**, 27-36.
- Lieberman B.S. 2000. *Paleobiogeography: Using Fossils to Study Global Change, Plate Tectonics, and Evolution*, Kluwer Academic/Plenum, New York.
- Linder H.P. and Crisp M.D. 1995. *Nothofagus* and Pacific biogeography. *Cladistics*, **11**,5-32.
- Lohman D.J., de Bruyn M., Page T., von Rintelen K., Hall R., Ng P.K.L., Shih H., Carvalho, G.R. and von Rintelen T. 2011. Biogeography of the Indo-Australian Archipelago. *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–26.

- Lourie S.A. and Vincent A.C.J. (2004a) A marine fish follows Wallace's line: the phylogeography of the three-spot seahorse (*Hippocampus trimaculatus*, Syngnathidae, Teleostei) in Southeast Asia. *Journal of Biogeography*, **31**, 1975–1985.
- Makovicky P.J. 2008. Telling time from fossils: a phylogeny-based approach to chronological ordering of paleobiotas. *Cladistics*, **24** (3), 350-371.
- McCarthy D. 2009. Here Be Dragons: How the Study of Animal and Plant Distributions Revolutionized Our Views of Life and Earth.—. Oxford: Oxford University Press. xxiii+214 pp. ISBN 978-0-19- 954246-8. £16.99, \$29.95 (hardback). ISBN 978-0-19- 959566-2. £8.99, \$18.95 (paperback, due June 2011).
- Metcalfe I. 1998. Palaeozoic and Mesozoic geological evolution of the SE Asian region: multidisciplinary constraints and implications for biogeography. In: *Biogeography and Geological Evolution of SE Asia* (eds Hall, R. and Holloway, J.D.), 25-41, Backhuys Publishers, Amsterdam, The Netherlands.
- Metcalfe I. 2001. Tectonic History of the SE Asian-Australian region. In: *Bridging Wallace's Line: The Environmental and Cultural History and Dynamics of the SE-Asian-Australian Region* (eds Kershaw, P., David, B., Tapper, N., Penny, D. and Brown, J.). *Advances in GeoEcology* 34, 29-48, Catena Verlag. Reiskirchen. ISBN 3-923381-47-6.
- Metcalfe I. 2002. Permian tectonic framework and palaeogeography of SE Asia. *Journal of Asian Earth Sciences*, **20**, 551-566.
- Metcalfe I. 2008. Gondwana Dispersion and Asian Accretion: An update. *Proceedings of the International Symposia on Geoscience Resources and Environments of Asian Terranes (GREAT 2008), 4th IGCP 516, and 5th APSEG*, 23-25.
- Metcalfe I. 2011. Paleozoic-Mesozoic history of SE Asia. The SE Asian gateway: history and tectonics of Australia-Asia collision (eds R. Hall; M.A. Cottam; M.E.J. Wilson), Geological Society, London, Special publications, **355**, 7-35.
- Michaux B. 2010. Biogeology of Wallacea: geotectonic models, areas of endemism, and natural biogeographical units. *Biological Journal of the Linnean Society*, **101**, 193-212.
- Milne R.I. 2004. Phylogeny and biogeography of *Rhododendron* subsection *Pontica*, a group with a Tertiary relict distribution. *Molecular Phylogenetics and Evolution*, **33**, 389-401.
- Mooi R.D. and Gill A.C. 2010. Phylogenetics without synapomorphies – A crisis in Fish Systematics: Time to show some character. *Zootaxa*, **2450**:26-40.
- Morrone J.J., 2004a. Homología biogeográfica : Las coordenadas espaciales de la vida. Mexico, D.F. : Cuadernos del Instituto de Biología 37, Instituto de Biología, UNAM.
- Morrone J.J., 2009. Evolutionary biogeography. An integrative approach with case studies. Columbia University Press.
- Muellner A. N., Pannell C. M., Coleman A. and Chase M. W. 2008. The origin and evolution of Indomalayan, Australasian and Pacific island biotas: insights from Aglaieae (Meliaceae, Sapindales), *Journal of Biogeography*, **35**(10), 1365-2699.

- Myers N., Mittermeier R. A., Mittermeier C. G., da Fonseca G. A. B. and Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853-858.
- Nelson G.J. 1994. Homology and Systematics. In *Homology: the hierarchical basis of comparative biology*, ed. B.K. Hall, 110-149. San Diego: Academic Press.
- Nelson G. J. and Ladiges P. Y. 1991a. Standard assumptions for biogeographic analysis, *Australian Systematic Botany*, **4**, 41-58.
- Nelson G. J. and Ladiges P. Y. 1991b. Three-area statements: Standard assumptions for biogeographic analysis, *Systematic Zoology*, **40**, 470-485.
- Nelson G.J. and Ladiges P.Y. 1991c. TAX: MSDOS Program for Cladistic Systematics, published by the authors, New York.
- Nelson G.J., Ladiges P.Y. 1992. Information content and fractional weight of three-item statements. *Systematic Biology*, **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.J. and Ladiges P.Y. 2009. Biogeography and the molecular dating game: a futile revival of phenetics? *Bull. Soc. Geol. Fr.*, **180**, (1), 39-43.
- Nelson G.J. and Platnick N.I. 1978. The perils of plesiomorphy: widespread taxa, dispersal, and phenetic biogeography. *Systematic Biology*, **27** (4), 474-477.
- Nelson G.J. and Platnick N.I. 1980. A vicariance approach to historical biogeography. *BioScience*, **30** (5), 339-343.
- Nelson G.J. and Platnick N.I. 1981. *Systematics and Biogeography: Cladistics and vicariance*. Columbia University Press, New York.
- Nelson G.J. and Platnick N.I., 1991. Three-taxon statements: a more precise use of parsimony? *Cladistics*, **7**, 351-366.
- Norstog, K.J. and Nicholls, T.J. 1997. *The Biology of Cycads*. Cornell University Press, Ithaca, USA: 363 pp
- Page R.D. 1989. Comments on component-compatibility in historical biogeography. *Cladistics*, **5** (2), 167-182.
- Page R.D. 1990. Temporal congruence and cladistic analysis of Biogeography and cospeciation. *Systematic Zoology*, **39** (3), 205-226.
- Page R.D. 1993. COMPONENT: Tree comparison software for Microsoft windows, version 2. Natural History Museum, London.
- Page R.D.M. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biological Sciences*, **12**, 357-358.
- Parenti L.R. and Ebach M.C. 2009. *Comparative Biogeography: Discovering and Classifying Biogeographical Patterns of a Dynamic Earth*. University of California Press, Berkeley.
- Parenti L.R. and Ebach M.C. 2010. Wallacea deconstructed. Beyond cladistics: the branching of a paradigm (ed. by D. M. Williams and S. Knapp), 303-318. University of California Press, Berkeley, CA.

- Parenti L.R. and Ebach M.C. 2013. Evidence and hypothesis in biogeography. *Journal of Biogeography*, **40** (5), 813-820.
- Pelletier J.D. 1999. Species-area relation and self-similarity in a biogeographical model of speciation and extinction. *Physical Review Letters*, **82**, 1983-1986.
- Peterson A.J., Soberón and Sánchez-Cordero 1999. Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265-1267.
- de Pinna M.G.G. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics*, **7** (4), 367-394.
- Platnick N.I. 1991. Advances in spider taxonomy 1988-1991, with synonymies and transfers 1940-1980. New York: New York Entomological Society, v + 846 pp.
- Platnick N.I. and Nelson G.J. 1978. A method of analysis for historical biogeography. *Systematic Zoology*, **27**, 1-16.
- Powell C.M., Roots S.R., Veevers J.J. 1988. Pre-breakup continental extension in East Gondwanaland and the early opening of the eastern Indian Ocean. *Tectonophysics* **155**, 261-283.
- Rasnitsyn AP, Quicke DLJ, editors 2002. History of Insects. Kluwer Academic Publishers, Dordrecht, xii+517 pp. ISBN 1-4020-0026-X.
- Richardson J.E., Costion C.M. and Muellner A.N. 2012. The Malesian floristic interchange: plant migration patterns across Wallace's line. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (ed. Gower D. J. et al), 138-163.
- Rieppel O. 2004. The language of systematics, and the philosophy of 'total evidence'. *Systematics and Biodiversity*, **2**, 1-21.
- Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to quantification of historical biogeography. *Systematic Biology*, **46**, 195-203.
- Rosen B. R. 1985. Long-term geographical controls on regional diversity. *The Open University Geological Society Journal*, **6**, 25-30.
- Rosen B. R. 1988. From fossils to Earth history: applied historical biogeography. In: *Analytical biogeography: an integrated approach to the study of animal and plant distributions* (eds. A.A. Myers and P.A. Giller), 437-481. Chapman and Hall, London.
- Rosen B.R. and A.B. Smith 1988. Tectonics from fossils? Analysis of reef-coral and sea-urchin distributions from Late Cretaceous to Recent, using a new method. In: *Gondwana and Tethys* (eds. M.G. Audley-Charles and A. Hallam), 275-306. Geological Society of London Special Publication.
- Rosen D.E. 1975. A vicariance model of Caribbean biogeography. *Systematic Zoology*, **24** (4), 431-464.
- Rosen D.E. 1978. Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*, **27**, 159-188.
- Santini F. and Winterbottom R. 2002. Historical biogeography of Indo-western Pacific coral reef biota: is the Indonesian region a centre of origin? *Journal of Biogeography*, **29**, 189-205.
- Shih H.T., Yeo D.C.J. and Ng P.K.L. 2009. The collision of the Indian plate with Asia: molecular evidence for its impact on the phylogeny of freshwater crabs (Brachyura, Potamidae). *Journal of Biogeography*, **36**, 703-19.

- Teske PR, Cherry MI, Matthee CA (2004) The evolutionary history of seahorses (Syngnathidae: Hippocampus): molecular data suggest a West Pacific origin and two invasions of the Atlantic Ocean. *Mol. Phylogenet. Evol.*, **30**, 273–286.
- Thomas D.C., Hughes M., Phutthai T., Ardi W.H., Rajbhandary S., Rubite R., Twyford A.D., Richardson J.E. 2001. West to east dispersal and subsequent rapid diversification of the mega-diverse genus *Begonia* (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography*, **39** (1), 98-113.
- Upchurch P. and Hunn C.A. 2002. "Time": the neglected dimension in cladistic biogeography? *Geobios. Mémoire special n°24*.
- Upchurch P., Hunn C.A. and Norman D.B., 2002. An analysis of dinosaurian biogeography: evidence for the existence of vicariance and dispersal patterns caused by geological events. *Proceedings of the Royal Society of London, Serie B* **269**, 613-622.
- Ung V. Biotic Evolution and Environmental Change in Southeast Asia. — Edited by David Gower, Kenneth Johnson, James Richardson, Brian Rosen, Lukas Rüber and Suzanne Williams. Cambridge: Cambridge University Press, 2012. xv+475 pp ISBN 978-1-107-00130-5 £70 \$US115 (hardback). Book review, *Systematic Biology* (sous presse).
- van Oosterzee P. 1997. Where worlds collide: the Wallace Line. Cornell University Press, Ithaca and London.
- van Tol J. and Gasmann D. 2007. Zoogeography of freshwater invertebrates of Southeast Asia, with special reference to Odonata. In: *Biogeography, Time, and Space: Distributions, Barriers, and Islands* (eds W. Renema), 45-91. Springer.
- van Wyhe J. 2012. Wallace, Darwin and Southeast Asia : the real field site of evolution. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (eds. Gower D. J. et al), 16-32.
- Veevers J. J. 1988. Morphotectonics of Australia's north- western margin – a review. In: *The North West Shelf. Proceedings of Petroleum Exploration Society of Australia Symposium* (eds. Purcell, P. G. and Purcell, R. R.), PESA, Perth, WA, 19–27.
- Veevers J.J. (Ed.), 2000. Billion-year earth history of Australia and neighbours in Gondwanaland GEMOC Press, Sydney. 400 pp.
- Wagner P.J. 1998. Phylogenetic analysis and the quality of the fossil record. Pages 165-187 in *The adequacy of the fossil record* (S.K. Donovan and C.R.C. Paul, eds) Wiley Intersciences, New York.
- Wallace A.R. 1853a. Palm trees of the Amazon and their uses. Van Voorst, London.
- Wallace A.R. 1855. On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History*, (2nd ser.) **16**, 184-196.
- Wallace A.R. « On the zoological geography of the Malay Archipelago, (1859) ». 2009. Alfred Russel Wallace Classic Writings. Paper 4.
- Wallace A.R. 1869. The Malay Archipelago : the land of the Orang-Utan and the bird of paradise. Macmillan, London.

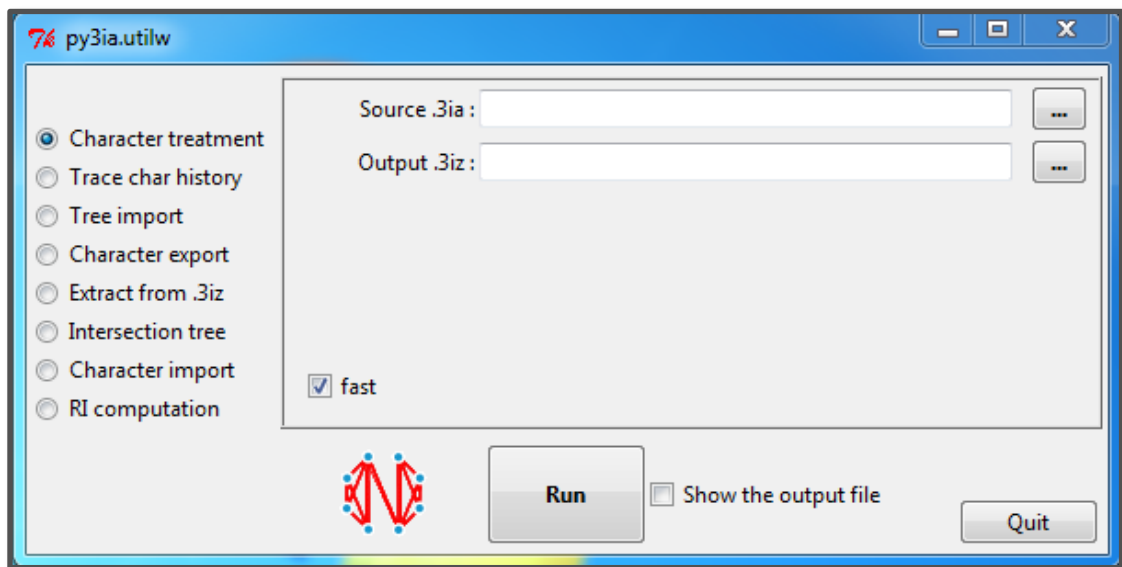
- Wallace A.R. 1881. *Island life*. Macmillan, London.
- Webb C.O. and Ree R. 2012. Historical biogeography inference in Malesia. In *Biotic Evolution and Environmental Change in Southeast Asia*, special volume series of the Systematic Association, Cambridge University press (eds. Gower D. J. *et al*), 191-215.
- Wiley E.O. 1987. Methods in vicariance biogeography. In : *Systematics and Evolution : a matter of diversity*, (P. Hovenkamp), 283-306. University of Utrecht, Utrecht.
- Wiley E.O. 1988a. Parsimony analysis and vicariance biogeography. *Systematic Zoology*, **37**, 271-290.
- Wiley E.O. 1988b. Vicariance biogeography. *Annu. Rev. Ecol. Syst.* **19**, 513-542.
- Williams DM. 2004. Homology and Homologues, Cladistics and Phenetics: 150 Years of Progress. In : *Milestones in Systematics* (eds. Williams, D.M. and Forey, P.L.), 191-224. London: Taylor and Francis.
- Wojcicki M. and Brooks D.R. 2005. PACT : An efficient and powerful algorithm for generating area cladograms. *Journal of Biogeography*, **32**, 755-774.
- Williams P.R., Johnston C.R., Almond R.A. and Simamora W.H. 1988. Late Cretaceous to Early Tertiary structural elements of West Kalimantan. *Tectonophysics*, **148**, 279-298.
- Williams S.T. and Duda T.F.Jr. 2008. Did tectonic activity stimulate oligo-miocene speciation in the Indo-West Pacific ? *Evolution*, **62**, 1618-1634.
- Woodruff D.S. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and future of the remaining refugial-phase biodiversity. *Biodiversity Conservation*, **19**, 919-941.
- Zandee, M., and Roos, M. C. (1987). Component-compatibility in historical biogeography. *Cladistics*, **3**, 305-332.
- Zaragüeta-Bagils, R. Lelièvre H. and Tassy P. 2004. Temporal paralogy, cladograms and the quality of the fossil record. *Geodiversitas* **26** (3), 381-389.
- Zaragüeta Bagils, R., Bourdon, E., Ung, V., Vignes-Lebbe, R. and Malécot, V. 2009. On the International Code of Nomenclature (ICAN). *Journal of Biogeography*, **36**, 1617-1619.
- Zimmermann E.A.W.V. 1777. *Specimen zoologiae geographicae, quadrupedum domocilia et migrationes systems*. Theodorum Haak, Leiden/Batavia.

Annexes

Annexes

- I. ToolBox de LisBeth ;
- II. Test de la classification de Silder et Silder : fichier d'entrée pour l'analyse biogéographique avec LisBeth de l'échantillonnage ;
- III. **Ung V.** (2013) Biotic evolution and environmental change in Southeast Asia. Book review, *Systematic Biology*, 62 (6) : 315-315 ;
- IV. Grand A., Zaragüeta-Bagils R., Duque-Velez L. M. & **Ung V.** (2013) A cladistic re-analysis of Gadiformes (Teleostei, Paracanthopterygii) using three-item analysis. (*Zootaxa*, révisé, **soumis**) ;
- V. Hoagstrom C. W., **Ung V.**, Taylor K. (2013) Comparative biogeography and Geographic Speciation of North American highland fishes from the Mesa Central to the Canadian Shield (*Journal of Biogeography*, **en révision**) ;
- VI. Zaragüeta-Bagils R., Bourdon E., **Ung V.**, Vignes-Lebbe R. et Malécot V. (2009): Comments on the International Code of Area Nomenclature. *Journal of Biogeography*, **38** : 1617-1619 ;
- VII. **Ung V.**, Dubus G., Zaragüeta-Bagils R. & Vignes-Lebbe R. (2010) Xper²: introducing e-Taxonomy. *Bioinformatics*, 26(5): 703-704 ;
- VIII. Burguiere T., Causse F., **Ung V.** et Vignes-Lebbe R. (2013) IKey+: A New Single-Access Key Generation Web Service. *Systematic Biology*, 62(1): 157-161 ;
- IX. Burguiere T., Bergamaschi A., Causse F., **Ung V.**, Paulhe N., Lison F. et Vignes-Lebbe R. (2013) MKey+ : Interactive Identification webservice (**en préparation**) ;

Annexe I : ToolBox de LisBeth



Nous avons choisi de fournir une « boîte à outils » intégrant tout un ensemble de petites fonctionnalités pouvant être utilisées de manière indépendantes les unes des autres.

- Décomposition des caractères en 3is (« Character treatment »)

Génération de la liste des 3is.

- Traçabilité des caractères (« Trace char history »)

Ce module permet de garder la trace les caractères utilisés en Biogéographie, *id est* le nom des taxons sources ayant produits les TAC libre de paralogie et de MAST qui ont servis à reconstruire l'arbre d'intersection (Zaraguëta-Bagils et al. 2012).

- Facilités d'import/export

Cette fonctionnalité permet d'importer et d'exporter des fichiers de/vers PAUP et TNT.

- Import d'arbres (« Tree import »)

Permet d'importer un fichier .tre contenant juste des arbres obtenus avec PAUP ou TNT.

- Import de matrices (« Character import »)

LisBeth peut désormais importer des matrices morphologiques mais aussi moléculaires sous différentes formes : ADN, ARN et acides-aminés avec en plus le contrôle du polymorphisme.

- Export de caractères (« Character export »)

LisBeth peut exporter facilement des fichiers au format .3iz (son propre fichier de sortie) vers les formats TNT et PAUP.

- Extraction de différentes parties indépendantes contenues dans le fichier .3iz

(format du fichier de sortie de LisBeth) (« Extract from .3iz »)

- Arbre d'intersection (« Intersection tree »)

LisBeth peut reconstruire un arbre d'intersection à partir de plusieurs sources (listes de taxons et d'arbres, listes d'arbres seulement).

- Calcul des RI (« RI Computation »)

LisBeth peut recalculer les RI à partir d'une liste externe de caractères et/ou d'arbres

Annexe II : Test de la classification de Silder et Silder

Fichier d'entrée pour l'analyse biogéographique avec LisBeth de l'échantillonnage suivant :

Taxons :

Lepidopetatum (van Welzen et al. 1992)

Rhysotoechia (Etman 1994)

Haloveloides (Andersen 1989a, 1991, 1992)

Halobates regalis (Andersen 1991, 1998; Andersen & Foster 1992; Andersen & Weir 1994)

Halobates princeps (Andersen 1991, 1998; Andersen & Foster 1992; Andersen & Weir 1994)

Halovelia (Andersen 1989b, 1991, 1998)

Megarthus (Cuccodoro 1998)

Erismanthus (van Welzen 1995)

Neoscortechinia (van Welzen 1994)

Fordia (Buijsen 1988; Dasuki & Schot 1991; Schot (1991)

Parkia (Hopkins 1994; Luckow & Hopkins 1995)

Hydrophilidae (Bameul 1997a, b)

Crocidura (Corbet & Hill 1992; Hutterer 1993; Ruedi 1995, 1996)

Aires :

MO Moluccans region

SU Sumatran region

JS Java Sea region

SU Sulu sea region

JR Japonese region

MR Melanesian region

SM Samoan region

DA Damperian region

MI Micronesian region

IN Indian region

AF African region

Références

- Andersen, N.M. (1989a) The coral bugs, genus *Halovelgia* Bergroth (Hemiptera, Veliidae). I. History, classification, and taxonomy of species except the *H. malaya*-group. *Ent. Scand.*, 20, 75±120.
- Andersen, N.M. (1989b) The coral bugs, genus *Halovelgia* Bergroth (Hemiptera, Veliidae). II. Taxonomy of the *H. malaya*-group, cladistics, ecology, biology, and biogeography. *Ent. Scand.*, 20, 179±227.
- Andersen, N.M. (1991) Cladistic biogeography of marine water striders (Insecta, Hemiptera) in the Indo-Pacific. *Austr. Syst. Bot.*, 4, 151±163.
- Andersen, N.M. (1992) A new genus of marine water striders (Hemiptera, Veliidae) with five new species from Malaysia. *Ent. Scand.*, 22, 389±404.
- Andersen, N.M. (1998) Marine water striders (Heteroptera, Gerromorpha) of the Indo-Pacific: cladistic biogeography and Cenozoic paleogeography. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 341±354. Backhuys Publishers, Leiden.
- Andersen, N.M. & Foster, W.A. (1992) Sea skaters of India Sri Lanka, and the Maldives, with a new species and a revised key to Indian Ocean species of *Halobates* and *Asclepios* (Hemiptera, Gerridae). *J. Nat. Hist.*, 26, 533±553.
- Andersen, N.M. & Weir, T.A. (1994) The sea skaters, genus *Halobates* Eschscholtz (Hemiptera, Gerridae), of Australia: taxonomy, phylogeny, and zoogeography. *Invert Tax*, 8, 861±909.
- Bameul, F. (1997a) A revision of *Protosternum* Sharp (Coleoptera, Hydrophilidae). *Nouv. Rev. Ent. (N.S.)*, 14, 17±41.
- Bameul, F. (1997b) A revision of *Mucetum* D'Orchymont and *Rhombosternum* Balfour-Browne (Coleoptera: Hydrophilidae) with a phylogenetic analysis of Protosternini. *Ann. Soc. Entomol. Fr. (N.S.)*, 33, 375±403.
- Buijsen, J.R.M. (1988) Revision of the genus *Fordia* (Papilionaceae: Millettieae). *Blumea*, 33, 239±261.
- Corbet, G.B. & Hill, J.E. (1992) The mammals of the Indomalayan region: a systematic review, pp. 39±46. Natural History Museum Publications, Oxford University Press, Oxford.
- Cuccodoro, G. (1998) Revision and phylogeny of *Megarthrus* Curtis 1829 from New Guinea, New Caledonia and Fiji (Coleoptera, Staphylinidae, Proteininae). *Trop. Zool.*, 11, 103±137.
- Dasuki, U.A. & Schot, A.M. (1991) Taxonomy of *Fordia* Hemsley (Papilionaceae: Millettieae). *Blumea*, 36, 191±204.
- Etman, B. (1994) A taxonomic and phylogenetic analysis of *Rhysotoechia* (Sapindaceae). *Blumea*, 39, 41±71.
- Hutterer, R. (1993) Order Insectivora (Family Soricidae, sub-family Crocidurinae, genus *Crocidura*). *Mammal species of the world. A taxonomic and geographic reference*, 2nd edn, pp. 81±98 (eds D.E. Wilson and D.M. Reeder). Smithsonian Institution Press, Washington.
- Ruedi, M. (1995) Taxonomic revision of shrews of the genus *Crocidura* from the Sunda Shelf and Sulawesi with description of two new species (Mammalia: Soricidae). *Zool. J. Linn. Soc.*, 115, 211±265.
- Ruedi, M. (1996) Phylogenetic evolution and biogeography of Southeast Asian shrews (genus *Crocidura*: Soricidae). *Biol. J. Linn. Soc.*, 58, 197±219.
- van Welzen, P.C. (1994) Taxonomy, phylogeny, and geography of *Neoscortechinia* Hook.f. ex Pax (Euphorbiaceae). *Blumea*, 39, 301±318.
- van Welzen, P.C. (1995) Taxonomy and phylogeny of the

Euphorbiaceae tribe Erismantheae G.L. Webster (Erismanthus, Moultonianthus, and Syndyophyllum). *Blumea*, 40, 375±396.

van Welzen, P.C., Piskaut, P. & Windadri, F.I. (1992) *Lepidopetalum* Blume (Sapindaceae): taxonomy, phylogeny, and historical biogeography. *Blumea*, 37, 439±465.

```
;
Dimensions
nchar=11 ntax=14
;
```

Taxa

```
100 = L.xylocarpum
101 = L.subdichotomum
102 = L.fructoglabrum
103 = L.micans
104 = L.montana
105 = L.perrottetii

200 = Rh.bifoliolata
201 = Rh.nitida
202 = Rh.mortoniana
203 = Rh.florulenta
204 = Rh.robertsonii
205 = Rh.grandifolia
206 = Rh.ramiflora
207 = Rh.koordersii
208 = Rh.congesta
209 = Rh.multiscarpa
210 = Rh.applanata
211 = Rh.obtusa
212 = Rh.bilocularis
213 = Rh.gracilipes
214 = Rh.elongata
215 = Rh.flavescens

400 = H.papuensis
401 = H.brevicornis
402 = H.browni
403 = H.sundaensis
404 = H.danpolhemi
405 = H.cornuta

500 = H.zephyrus
501 = H.regalis
502 = H.whiteleggel
503 = H.sexualis
505 = H.peronis
506 = H.herringi
507 = H.darwini
508 = H.acherontis

600 = H.princeps
601 = H.kelleni
602 = H.mariannarum
603 = H.katherinae
604 = H.fijiensis
605 = H.salotae

700 = H.annemariae
701 = H.sulawesi
702 = H.abdominalis
703 = H.nicobarensis
704 = H.malaya
705 = H.n.sp.Goa

800 = M.montanus
801 = M.depressus
802 = M.zulu
```

803 = *M.conspirator*
804 = *M.parallrius*
805 = *M.rufomarginatus*
806 = *M.sumatrenis*
807 = *M.vastus*
808 = *M.incubifer*
809 = *M.gressitti*
810 = *M.kuscheli*
811 = *M.alienu*
812 = *M.fijianus*
813 = *M.peckorum*
814 = *M.mukankundiyorum*
815 = *M.inaequalis*
816 = *M.auricola*
817 = *M.primus*
818 = *M.anggiensis*
819 = *M.baliemensis*
820 = *M.riedeli*

900 = *G.laurifolius*
901 = *G.bicoronatus*
902 = *G.aurantiacus*
903 = *G.siamicus*
904 = *G.hastatus*
905 = *G.ellipticus*
906 = *G.valvatus*
907 = *G.rectinervus*
908 = *G.rufo-velutinus*
909 = *G.crassifolius*
910 = *G.macrophyllus*
911 = *G.maingayi*
912 = *G.nicobarensis*
913 = *G.ridleyi*
914 = *G.subpeltatus*

1000 = *M.leembruggianus*
1001 = *E.obliquus*
1002 = *E.sinensis*
1003 = *S.excelsus*
1004 = *S.occidentalis*

1100 = *N.angustifolia*
1101 = *N.sumatrensis*
1102 = *N.nicobarica*
1103 = *N.philippinensis*
1104 = *N.forbesi*
1105 = *N.kingii*

1200 = *F.bracteolata*
1201 = *F.albiflora*
1202 = *F.unifoliolata*
1203 = *F.incredibilis*
1204 = *F.spec.a*
1205 = *F.leptobotrys*
1206 = *F.ngii*
1207 = *F.nivea*
1208 = *F.johorensis*
1209 = *F.pauciflora*
1210 = *F.lanceolata*
1211 = *F.ophirensis*
1212 = *F.cauliflora*
1213 = *F.brachybotrys*
1214 = *F.splendidissima*
1215 = *F.stipularis*
1217 = *F.rheophyta*

1300 = P.leiophylla
1301 = P.speciosa
1302 = P.timoriana
1303 = P.singularis
1304 = P.sumatrana
1305 = P.korom
1306 = P.versteeghii
1307 = P.sumastrep
1308 = P.biglandulosa
1309 = P.parvifoliola

1500 = Mu.carinatum
1501 = Rh.birmanense
1502 = Rh.intervallum
1503 = Rh.minutum
1504 = Rh.consobrinum
1505 = Rh.index
1506 = Rh.wagneri
1507 = Pr.punctatum
1508 = Pr.atomarium
1509 = Pr.newtoni
1510 = Pr.hanseni
1511 = Pr.longicarinatum
1512 = Pr.obscurum
1513 = Pr.abnormale
1514 = Pr.hammondi

1600 = Cr.fuliginosa
1601 = Cr.monticola
1602 = Cr.paradoxura
1603 = Cr.malayana
1604 = Cr.negligens
1605 = Cr.mindorus
1606 = Cr.foetida
1607 = Cr.nigripes
1608 = Cr.grayi
1609 = Cr.beatus
1610 = Cr.beccarii
1611 = Cr.hutani
1612 = Cr.brunnea
1613 = Cr.orientalis
1614 = Cr.lepidura
1615 = Cr.rhoditis
1616 = Cr.lea
1617 = Cr.levicula
1618 = Cr.elongata
1619 = Cr.musseri

;

Areas

MO: 100 101 102 103 204 205 208 209 210 211 212 213 214 400 402 401 501 503 505 600 700
816 817 818 819 820 1003 1104 1302 1306
MR: 100 101 102 103 204 208 209 210 211 212 213 214 400 402 501 503 505 600 603 700 809
810 811 813 816 817 818 819 820 1003 1104 1302 1306
DA: 100 200 201 202 203 204 215 500 501 502 506 507 508 600
SM: 601 604 605 812
MI: 602 1305 1309
JS: 205 206 207 401 403 503 600 701 702 704 900 901 902 903 904 909 905 906 911 910 908
913 914 1000 1001 1004 1002 1100 1101 1102 1103 1105 1200 1201 1202 1203 1204 1206 1208
1209 1210 1211 1205 1212 1213 1214 1217 1301 1302 1303 1307 1304 1500 1501 1503 1502 1505
1506 1507 1514 1504 1511 1600 1601 1603 1604 1612 1613 1606 1607 1615 1616 1617 1618 1619
SR: 104 105 403 503 702 703 806 905 909 910 912 1000 1001 1004 1101 1102 1105 1207 1215
1301 1302 1303 1304 1513 1601 1602 1610 1611 1614 1612 1613 1504 1510
SU: 105 205 206 207 401 404 405 403 505 600 701 807 905 906 909 911 907 1000 1001 1004
1100 1101 1102 1103 1105 1213 1214 1217 1301 1302 1303 1304 1502 1505 1506 1507 1514 1504

1511 1601 1606 1605 1608 1609 1607 1615 1616 1617 1618 1619
IN: 705 805 900 902 909 1300 1302 1307 1308 1501 1503 1508 1509 1512
JR: 800 801 803 804 808 1513
AF: 801 802 814
;

Characters

[2] (100,(101,(102,(103,(104,105))))
[3] (((200,201),(202,203)),(204,(((205,206),207),208)),(209,210),(211,(212,(213,(214,215))))))
[5] (((402,403),400,401),(405,404))
[6] (500,(501,(502,(503,(505,(506,(507,508))))))
[7] (600,(601,(602,(603,(604,605))))
[8] (700,(701,702,(703,(704,705))))
[9] (800,((801,802),(803,(804,(805,806,807))),(808,((809,(810,(811,812,813))),(814,(816,817,(818,819,820))))))
[10] (((906,907),905),((900,901),(902,903)),904),(((910,911),908,909),912),913),914))
[11] (((1001,1002),(1003,1004)),1000)
[12] ((1100,1101),(1102,(1103,(1104,1105))))
[13] (((1213,1214),1215),(((((((1201,1200),1202),(1203,1204)),1205),(1206,1207)),1208),1209),1210),1211),1212),1217)
[14] (1300,1301,((1302,1303),(1304,(1305,1306)),(1307,(1308,1309))))
[16] (1500,(1501,1502,(1503,(1504,(1505,1506))),1507,((1509,1508),(1510,1511),(1512,(1513,1514))))))
[17] (((1602,(((1603,1604),1605,1607,1606),(1608,1609),1610),(1613,1612),1611))),1614,(1615,(1616,((1619,1618),1617))))),1600,1601)
;

Annexe III : Biotic Evolution and Environmental Change in Southeast Asia.

Book review

Ung Visotheary (2013)

Systematic Biology Systematic Biology, 62 (6) : 315-315

Book Review

Visotheary Ung

Centre de Recherche en Paléobiodiversité et Paléoenvironnements

UMR 7207: CNRS MNHN UPMC

57 rue Cuvier CP 48

75005 Paris

France

Email: visotheary.riviere-ung@snv.jussieu.fr

Phone: +33 1.40.79.80.67

Biotic Evolution and Environmental Change in Southeast Asia. — Edited by David Gower, Kenneth Johnson, James Richardson, Brian Rosen, Lukas Rüber and Suzanne Williams. Cambridge: Cambridge University Press, 2012. xv+475 pp ISBN 978-1-107-00130-5 £70 \$US115 (hardback).

Southeast Asia is a region of a great biological interest because of its dynamics and species richness, and the region has long been noted as a centre of biodiversity for both marine and terrestrial taxa (Michaux 2010; Woodruff 2010). Thus, it is crucial to analyse and understand the underlying processes that have led to this diversity, not just for systematics but for evolution in general, including biogeography.

Although varied literature currently exists dealing with broad-scale topics such as geological evolution, biogeography (including phylogeography and zoogeography) and conservation issues, no book has been available that gathers all of the subjects together. This impediment is overcome with the publication of *Biotic Evolution and Environmental Change in Southeast Asia*. The book is intended for “students and researchers as well as environmental NGOs”, aiming to summarize recent work in various fields like population genetics, historical biogeography, paleogeography, paleoclimatology, and conservation biology based on case studies.

As a PhD student working on Southeast Asia and the West Pacific region, I was anticipating this book with a lot of interest, and I haven't been disappointed. The book is a large survey of present-day key questions regarding the region and its underlying biogeography. This is more than just a proceedings volume of the Southeast Asia Gateway Evolution 2009 meeting — it is a collection of selected contributions from the meeting that surveys the very latest developments regarding the region.

The book contains 17 chapters (most about 20 to 30 pages long, only chapters 3 and 14 are a little longer) and covers a broad range of fields, from geology to conservation concerns, which makes it very complete. The book suffers the usual “proceedings” problem of being a collection of independent chapters, but Chapter 1 is an introduction to the volume, and it presents a brief but valuable summary of the remaining 16 chapters, so that readers get a good overview of the book, and can then start anywhere while still maintaining their perspective.

Chapter 2 is a retrospective on Darwin and Wallace's work, but after that the chapters are arranged in themes. The following three chapters deal with the paleo-evolution of the region, discussing its geological reconstruction, a review of the paleoclimate, and the impact of that climate on the Sundaland forests. The next three chapters focus on plant biogeography, covering Malesian floristic interchange, distribution patterns of palms, and the historical

biogeography of Malesian rhododendrons. This is followed by chapters presenting studies of aquatic biotas, both marine and freshwater: biogeography of coral reefs, Tsunami impacts in the marine environment, biodiversity hotspots in Wallacea, the freshwater fauna of the Indo-Australian Archipelago, and the history of amphibian species discoveries. The final three chapters cover the very concrete problem of conservation rules and how to protect this megadiverse region, covering the impediment of the wildlife trade, human impacts on the tropical peat swamps, and the biodiversity crisis within Southeast Asia.

As the book is quite substantial it is impossible to review it in all, but rather I will focus on some of the subjects that I found to be essential reading.

Personally, I have been particularly interested in the third, fourth and fifth chapters, dealing with the paleo-evolution of the region, since my own research concerns the biogeography of Southeast Asia (and extending to the West Pacific region) from a historical perspective. Cladistic biogeography involves looking for general pattern of area relationships in the combined analysis of cladograms of unrelated groups of taxa, which are then compared to the geological history in order to reveal congruence or incongruence. This highlights the possibility of testing for area homologies, which makes historical biogeography a comparative science. Hence, the better the geological evolution is known the better it is for such biological analyses.

These three chapters of the book are an important source of relevant information on the paleogeography and paleoclimate, particularly of Sundaland and Wallacea; and this will greatly improve the quality of cladistic biogeographic analyses of the region. In my case, comparison with geological reconstructions might give clues regarding evolutionary processes that have lead to the distributions of my study taxa. However, one does have to keep in mind that geological reconstructions are geologist's 'interpretation' (Hall 1998). Consequently, any reconstructions should be carefully used and integrated with full knowledge of the facts.

However, I do think that these chapters are likely to be the most interesting ones to most biogeographers, because the historical dynamics developed there clearly shows how it has shaped the Southeast Asian region. Furthermore, dispersalist evolutionary scenarios can be investigated in the light of the life-history traits of taxa inhabiting the region.

The next three chapters (6, 7 and 8) are specifically about plant biogeography, illustrated by the use of the Phylogenetic Niche Conservatism model to explore the reality of Wallace's Line for flora, in Chapter 6; and, in Chapter

7, the biogeography and distribution of palms through extensive new data regarding their distribution, phylogenies, paleogeography and paleoclimate.

Chapter 8, one of my favourites, is about 'Historical inference in Malesia'. The use of three different analysis programs (LAGRANGE, SHIBA and DIVA) is detailed to reconstruct ancestral areas of the rhododendron genus *Vireya*. These programs differ in their inherent methodologies, and there is an apparent antagonism between these software and the one I'm using (i.e. cladistic biogeography through three-item analysis). Nevertheless, it is crucial to know what kind of information different tools can provide; and it is much better to see them as complementary rather than as competitors.

Aquatic biotas and their species richness are presented in the next five chapters (9-14) which I will briefly mention here. All of them present new results, and Chapters 9 and 10 reach conclusions regarding how the species richness can be preserved. Chapters 11 and 12 focus on Sulawesi, an intriguing island known to be a composite of different terranes (Hall 2009, 2011). Alfred Russel Wallace (1880) himself referred to it as an 'anomalous island', and this particular nature makes Sulawesi very interesting for biogeographic inferences, which explains the large number of publications regarding it (e.g. van Welzen et al. 2011; van Welzen and Raes 2011; Stelbrink et al. 2012).

The final chapters of the book concern animal welfare (Chapter 15), the fragility of peatland ecosystems (Chapter 16) and the dramatic biodiversity losses mainly due to human activities (Chapter 17). This section particularly attracts attention since it is obvious that we humans have a very deep impact on our environment. Needless to say, a very important social responsibility relates to the wildlife trade (e.g. reptiles are hunted for their skins, as food and traditional medicines, and even as pets), which is at the moment an impediment to conservation. Human responsibilities are emphasized in Chapter 17, and the authors provide solutions, starting with economic priorities to help reduce poverty, which is a big cause of deforestation and hunting for bushmeat. Education and environmental awareness are the next challenges for local communities, to teach them how to protect their own biodiversity richness. Since protecting areas of high value like Southeast Asia is a global issue, and the authors enjoin for collaborative international incentives to help reduce deforestation, for example. Peatland ecosystems are used as an example in Chapter 16, as they are clearly threatened and need new protection policies to maintain these incredible and particular ecological niches. Here, again, the environment is totally dependant on human activities, and it thus relies on human-protection initiatives to avoid deforestation and enhance rehabilitation of peatland ecosystems. These concepts can be extended and generalized to any fragile ecosystems (e.g. coral reefs, mentioned in Chapter 10).

This book is very well written. To me, all of the chapters are of equal quality, although I really appreciated that the geological part is very understandable, and simplified compared to what is available in the literature (Hall 2009, 2011), giving a very nice and concise introduction to paleogeology.

Nevertheless, I have one regret, in that I would have liked to read about the historical biogeography of the region. Only one chapter deals with this (Chapter 8), and it does not use cladistic methodology. This geologically dynamic and highly complex region is a choice one for retrieving general patterns. The use of cladistic methods increases the potential to discover new patterns (Parenti and Ebach, 2009, 2013), and the incorporation of hierarchical, phylogenetic hypotheses introduces evolutionary hypotheses into the analysis area relationships are revealed based on the combined presence of non-related taxa, not just shared traits (Williams and Ebach 2006; Cao et al. 2007). This allow more direct reciprocal lightening between biological and geological evidence. These additional inputs can bring new insights regarding delimitation of conservation areas, for example, based on the taxa assemblage of each area. This sort of analysis is currently lacking for Southeast Asia.

In summary, this book is an almost complete overview of current knowledge concerning the Southeast Asian region, with both pure scientific studies such as the first ten chapters, and some more public-oriented papers, which point out human damage to the environment, and urge for conservation and biodiversity protection via sustainability policies. It perfectly suits its aims (being “an ideal resource for graduate students and researchers as well as environmental NGOs”), and I enjoyed reading this book very much. It is quite straightforward to read, and I would recommend it to any person interested in Southeast Asia, whatever their background.

REFERENCES

- Cao N., Zaragüeta Bagils R. & Vignes-Lebbe R. 2007. Hierarchical representation of hypotheses of homology. *Geodiversitas* 29:5-15.
- Hall R. 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In: R. Hall, J.D. Holloway, editors. *Biogeography and geological evolution of SE Asia*. Leiden, Backhuys, 99-131.
- Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea* 54:148-161.
- Hall R. 2011. Australia-SE Asia collision: plate tectonics and crustal flow. In: R. Hall, M.A. Cottam, M.E.J. Wilson, editors. *The SE Asian gateway: history and tectonics of Australia-Asia collision*. London: Geological Society, 75-109.

- Michaux B. 2010. Biogeology of Wallacea: geotectonic models, areas of endemism, and natural biogeographical units. *Biological Journal of the Linnean Society* 101:193–212.
- Parenti L.R., Ebach M.C. 2009. *Comparative biogeography*. Berkeley (CA): University of California Press.xw
- Parenti L.R., Ebach M.C. 2013. Evidence and hypothesis in biogeography. *Journal of Biogeography* 40:8013-820.
- Stelbrink B., Albrecht C., Hall R., von Rintelen T. 2012. The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace’s “anomalous island”? *Evolution* 66:2252–2271.
- Van Welzen P.C., Parnell J. A. N. and Slik J. W. F. 2011. Wallace’s Line and plant distributions: two or three phytogeographical areas and where to group Java? *Biological Journal of the Linnean Society*, 2011, 103, 531–545.
- Van Welzen P.C., Raes N. 2011. The floristic position of Java. *Gardens’ Bulletin Singapore* 63:329–339.
- Wallace A.R. 1880. *Island life, or the phenomena and causes of insular faunas and floras, including a revision and attempted solution of the problem of geological climates*. London: Macmillan.
- Williams D.M., Ebach M.C. 2006. The data matrix. *Geodiversitas* 28:409-420.
- Woodruff D.S. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today’s patterns and future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation* 19:919-941.

Visotheary Ung, Centre de Recherche en Paléobiodiversité et Paléoenvironnements (UMR 7207: CNRS MNHN UPMC), 57 rue Cuvier CP 48, 75005 Paris, France; E-mail: visotheary.riviere-ung@snv.jussieu.fr

Annexe IV : Three-item analysis and Gadiformes (Teleostei, Paracanthopterygii): how to conduct a cladistic analysis using LisBeth.

Anaïs Grand, Rene Zaragüeta Bagils, Lina Maria Duque Velez et **Visotheary Ung** (2013)

Zootaxa (révisé, soumis)

Three-item analysis and Gadiformes (Teleostei, Paracanthopterygii): how to conduct a cladistic analysis using LisBeth.

ANAÏS GRAND^{1,3,4}, RENE ZARAGÜETA BAGILS^{1,2,3}, LINA MARIA DUQUE VELEZ¹ & VISOTHEARY UNG^{1,2,3,4}

¹ Muséum national d'histoire naturelle, CR2P - UMR 7207 CNRS, MNHN, Univ Paris 06, 57 rue Cuvier, CP48, F-75005, Paris, France.

² UPMC Univ Paris 06, CR2P - UMR 7207 CNRS, MNHN, Univ Paris 06, 4 Place Jussieu, Tour 46-56, 5ème étage, F-75005, Paris, France.

³ Laboratoire Informatique et Systématique (LIS) ;
UPMC Paris 06, 4 Place Jussieu, F-75005, Paris, France

⁴ Corresponding author, grand@mnhn.fr, visotheary.riviere-ung@snv.jussieu.fr

Abstract

In the present paper, LisBeth, a newly published phylogenetic program is described. LisBeth implements the three-item analysis (3ia) cladistic methodology for systematics and biogeography. We show how LisBeth handles character representation, character analysis, exact search, reconstruction of the intersection tree and other features e.g. retention index, completeness index, character history and interoperability with other programs. The phylogeny based on morphological characters of the order Gadiformes (Teleostei, Paracanthopterygii) published by Endo (2002) is reanalysed using the 3ia rationale. The generation of hypotheses of homology following Endo's guidelines and 3ia representation schemes is discussed. We compare the topologies of all results and show the implications of the interpretation of character history.

Key words

Cladistics, Gadiformes, Intersection tree, LisBeth program, Systematics, Phylogenetic characters, Phylogeny, Three-item analysis.

Introduction

In November 2012, LisBeth, a new software program for the analysis of cladistic phylogenetics and biogeography data, was published (Zaragüeta Bagils *et al.* 2012). Since that time, an updated version (v.1.3) with some new features has been offered early 2013. In this paper we present a step by step account of how to conduct a systematic analysis using LisBeth. Our account is demonstrated using the data from a morphological study of fishes from the order Gadiformes (Teleostei, Paracanthopterygii, *sensu* Nelson 1994). Gadiforms constitute over one quarter of the world's marine fish catch (Nelson 2006). Their distribution ranges from Arctic to Antarctic waters in all oceans, from the deep-sea benthos to the shore, with both estuarine and fresh waters habitats (Cohen *et al.* 1990). The order displays a great morphological diversity. Despite the general agreement of the monophyly of Gadiforms with its current composition, its classification is still controversial (Endo 2002; Nelson 2006). This demonstration is based on the classification presented by Endo (2002). We have re-examined the relationships among taxa using the 23 morphological characters he published (Endo 2002) and evaluated the effect of theoretical assumptions and methodological choices. His cladistic approach led him to document in detail the 23 characters he used and provided all the evolutionary assumptions required for their representation in three-item analysis (3ia hereafter) using LisBeth.

LisBeth is the only software to date that implements a hierarchical representation of characters for three-item analysis (3ia hereafter). This method was introduced in the 1990's (Nelson & Platnick 1991; Nelson & Ladiges 1992, 1993; Wilkinson 1994a) with further developments emerging later (Kitching *et al.* 1998; Cao *et al.* 2007; Zaragüeta Bagils & Bourdon 2007; Williams & Ebach 2008; Zaragüeta Bagils *et al.* 2012). The foundations of 3ia are still discussed (Harvey 1992; Farris & Kluge 1998; Farris 2012) nonetheless the method has been applied to several studies in systematics (Patterson & Johnson 1995; Udovicic *et al.* 1995; Ladiges *et al.* 1999; Corvez 2012) and biogeography (Ladiges *et al.* 1992; Morrone & Carpenter 1994; Escalante *et al.* 2007; Cao *et al.* 2009; Parenti & Ebach 2009). We decided to specifically implement a representation of characters as rooted trees directly, thus making the usual taxon x character matrix unnecessary. In the current version of LisBeth, a compatibility analysis is applied using an exact algorithm to compute a summary tree (= the intersection tree) from the relationships common to all optimal trees, with character history being traced by synapomorphies at each node of the final tree. LisBeth is also interoperable with other systematic programs, e.g. parsimony programs or knowledgebase programs, namely Xper² (Ung *et al.* 2010).

Material and Methods

Material

The phylogenetic analysis involves gadiform families, subfamilies and genera: Bathygadinae *Jordan & Evermann 1898*; Gadinae *Rafinesque 1810a*; Gaidropsarinae (Actually synonyme of Phycinae); Lotinae; Macrourinae *Gilbert and Hubbs 1916*; Moridae *Moreau 1881*; Phycinae; *Bregmaceros* *Thompson 1840*; *Euclichthys* *McCulloch 1926*; *Macruronus* *Günther 1873*; *Merluccius* *Rafinesque 1810b*; *Melanonus* *Günther 1878*; *Muraenolepis* *Günther 1880*; *Raniceps* *Oken 1817*; *Squalogadus* *Gilbert & Hubbs 1916*; *Steindachneria* *Goode & Bean 1896* and *Trachyrincus* *Giorna 1809*.

All scientific names (nomina) were carefully confirmed using freely accessible online web sites: *The Nomenclatur Zoologicus* (<http://uio.mbl.edu/NomenclaturZoologicus/>) for the genus-group nomina; the *Index of Organism Names* (<http://www.organismnames.com/>); the *Integrated Taxonomic Information System* (<http://www.itis.gov/>) and the *Catalog of fishes*: (<http://research.calacademy.org/redirect?url=http://researcharchive.calacademy.org/research/Ichthyology/catalog/fishcatmain.asp>).

Method: Three-item analysis

This study is based on the data and data matrix published by Endo (2002:95), which includes 17 gadiform taxa, and 1 outgroup taxon (representing a generalization of all other paracanthopterygians). Endo detailed 23 morphological characters all of which were represented as unordered transformation series. All 23 characters are variable (11 presenting more than 2 states) and 22 are parsimony-informative. Here, we begin with a matrix-based approach (Table 1) to demonstrate how to convert conventional matrix characters into hierarchical characters so as to conduct three-item analyses using Lisbeth (Zaragüeta Bagils *et al.* 2012).

character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1st analysis	1	1	2	3	4	1	1	1	5	4	1	6	1	1	3	7	1	6	1	1	4	8	1
2nd analysis	1	1	9	4	10	1	1	1	11	4	1	12	1	1	1	13	14	15	1	14	16	8	1
Gadinae	0	0	2	0	1	0	1	1	2	1	0	5/6	1	1	0	1	0	4	0	0	2	2	0
Lotinae	0	0	2	0	1	0	1	1	2	1	0	5	1	1	0	1	0	4	0	0	2	2	0
Gaidropsarinae	0	0	2	2	1	1	1	1	2	1	0	5	1	1	0	2	0	3	0	0	1	2	0
Phycinae	0	0	2	?	1	1	1	1	1	1	0	5	1	1	0	2	0	5	0	0	1	2	0
<i>Bregmaceros</i>	0	0	3	0	2	0	0	1	1	2	0	5	1	1	2/0	3	1	3	0	0	1	3	0
<i>Muranolepis</i>	0	0	1	1	2	0	1	1	1	1	0	5	0	1	1	4	1	3	0	0	1	2	0
<i>Raniceps</i>	0	0	2	0	1	0	0	1	1	1	1	3	0	1	0	1	0	3	0	0	1	1	0

<i>Merluccius</i>	0	0	2	0	1	0	0	1	1	1	1	4	0	1	0	1	0	3	0	0	1	2	0
<i>Macruronus</i>	0	0	2	0	0	0	0	0	1	0	0	1/5	0	1	0	1	0	2	0	0	1	2	0
<i>Euclichthys</i>	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	2	0	0	1	0	0
<i>Steindachmeria</i>	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	3	0	1	0	0	0
Moridae	0	0	0	0	0	0	0	1	1	0	1	1	?	1	0	1	0	2	0	0	1	0	0
Bathygadinae	0	0	0	0	0	0	0	0	1	0	1	1	0	1	2/0	0	0	2	0	1	0	0	0
<i>Trachyrincus</i>	1	1	0	0	0	0	0	0	0	0	0	2	0	1	1	1	0	1	1	0	0	2	1
Macrourinae	1	1	0	0	0	0	0	0	1	0	0	5	?	1	2/0	0	0	2	0	1	0	0	1
<i>Squalogadus</i>	1	1	0	0	0	0	0	0	0	1	0	3	0	0	1	1	1	4	1	1	0	0	1
<i>Melanonus</i>	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0

TABLE 1. Taxon x character matrix published by Endo (2002) with 17 taxa and 23 characters. The first and second lines provide the hierarchy of states for each character for the 1st and 2nd analysis respectively. Hierarchies of states that differ between both analyses are shown with a bold markup. Hierarchies are enumerated as follow:

1: (0(1)); 2: (0 (1)(2)(3)); 3: (0(1)(2)); 4: (0(1(2))); 5: (1(2)(0)); 6: (0 (1)(2)(3)(4)(5)); 7: (1(0)(2)(3)(4)); 8: (0(1(2(3))))); 9: (3(0(1(2))))); 10: (2(0(1))); 11: (2(1(0))); 12: (5(2((0)(6,3,1))))); 13: (4(0,3(1(2))))); 14: (1(0)); 15: (0(1(2(3(4(5)))))); 16: (0,2(1)).

For the second analysis, the coding for the characters 12 and 15 differs from the matrix for some states which are indicated after / in the same matrix cell.

3ia applies the outgroup rationale (i.e. identification of plesiomorphic states) for each character individually, rendering the concept of an “outgroup taxon” useless. The hierarchical character representation always starts with a general class containing every taxon (except taxa with missing data), i.e. the root of the hierarchy. The 3ia approach consists of choosing nested classes. Only putative synapomorphies, that is the traits used as evidence to group terminal taxa, are considered.

Hierarchies of states are provided in Table 1 for two analyses. The first one strictly observes Endo’s assumptions for all of the 23 variable characters 22 of which are informative in 3ia and parsimony. The second analysis differs from the first one for 11 characters due to a different representation of “losses”, “states present in a single taxon” and “inapplicabilities”. All 23 characters are informative for 3ia. 3ia analyses were conducted using an exact algorithm (branch and bound). For both analyses, the intersection tree (Cao *et al.* 2009) was calculated. The characterization of clades by synapomorphies is calculated according to the 3ia method of character history traceability. The technical discussion providing a mini-manual of LisBeth is available in the appendix.

Results

Our first analysis yields a single tree (Fig. 1A). The overall RI = 0.767 and Compl. = 96.3% (please, refer to Zaragüeta Bagils *et al.* 2012 for the meaning of the completeness index Compl.) The per-character RI values are given in Table 2. Our second analysis also yields a single tree (Fig. 1B). The overall RI = 0.657 and Compl. = 97.9%. The per-character RI values are given in Table 2.

Synapomorphies for some clades are provided by Figure 2 for both analyses (Fig. 2A, B) and for Endo's phylogeny (Fig. 2C). Figure 2D provides Endo's optimization of characters.

characters	First analysi	Second analysis
1	1	1
2	1	1
3	0.728	0.866
4	NA	0.714
5	1	0.719
6	1	1
7	0.967	0.933
8	0.944	0.896
9	0.969	0.318
10	0.806	0.733
11	0.339	0.515
12	0.546	0.211
13	0.960	0.800
14	0.767	0.871
15	0.345	0.381
16	0.954	0.505
17	0.381	0.407
18	0.469	0.698
19	1	1
20	0.936	0.705
21	1	0.771
22	0.792	0.776
23	1	1

TABLE 2. per-character RI values for the 23 characters and for both analyses. The RI value cannot be calculated for the single uninformative character of the first analysis (as indicated by “NA” in the Table).

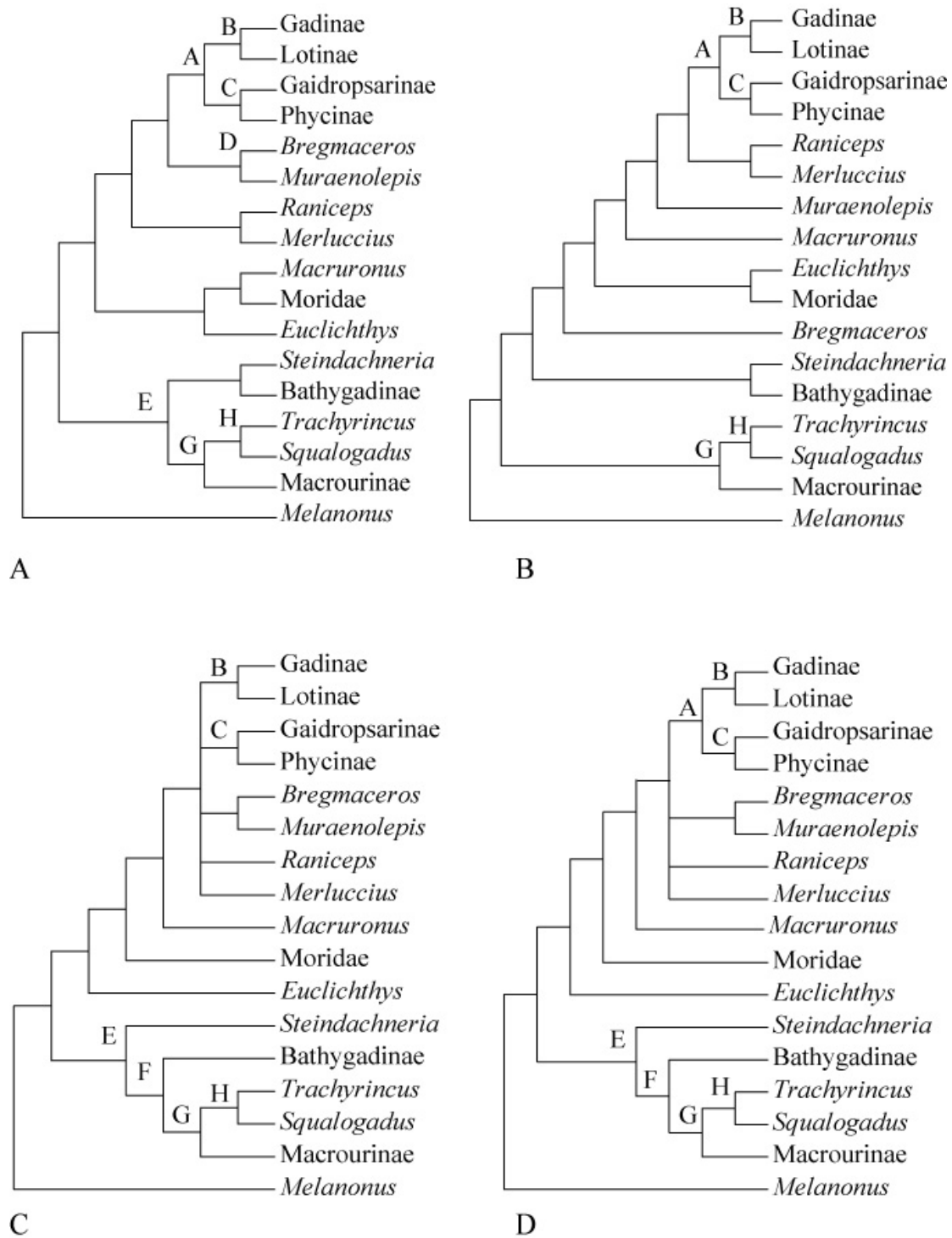


FIGURE 1. A: intersection tree from our first 3ia analysis. B: intersection tree from our second 3ia analysis. C: Strict consensus of 11 most-parsimonious trees found by Endo (2002). D: intersection tree of the 11 most-parsimonious trees reconstructed by Endo (2002).

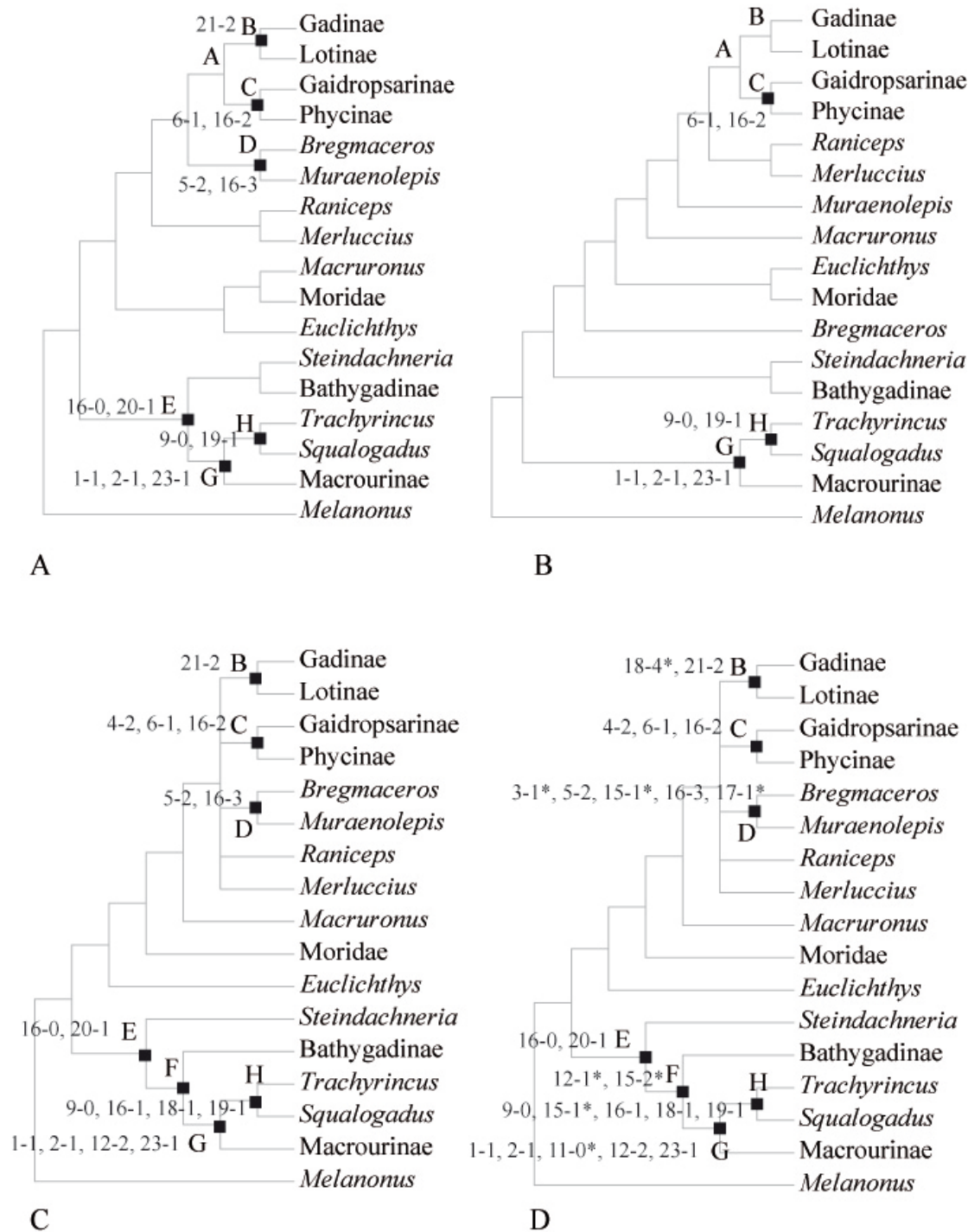


FIGURE 2. Character history traced by synapomorphies for some clades. A: on the intersection tree from our first 3ia analysis, using the 3ia method of traceability of character history. B: on the intersection tree from our second 3ia analysis, using the 3ia method of traceability of character history. C: on the strict consensus from Endo's analysis (2002), using the 3ia method of traceability of character history. D: on the strict consensus from Endo's analysis (2002), using the parsimonious optimization of character states. * indicates the states which appear at least twice in the tree.

Discussion

Character representation. In spite of past (and recent) controversies regarding the merits of 3ia, we consider it to offer great benefits especially in its hierarchy-based approach. LisBeth is the only program that does not rely on a conventional matrix-representation of systematic characters. Instead, the characters are represented and treated as rooted trees. This representation manages the “inapplicable” versus “missing” data distinction and thus removes that difficulty when choosing an appropriate coding strategy. Our choice of conducting two analyses allows for illustrating some of the 3ia characteristics, such as the dependencies among states, the inapplicability conditions, the constraint concerning states present in a single taxon, and the representation of secondary losses.

The first analysis follows all of Endo’s assumptions for the 23 characters. For instance Endo described the character 2 “Presence of anterolateral processes on lachrymal” as follows (2002:84):

“In most paracanthopterygians, such a process on the lachrymal is absent (“state 0”). In macrourines, *Squalogadus* and *Trachyrincus*, one or two processes are present on the anterolateral region of lachrymal (“state 1”).

By specifying the coding of the outgroup (most paracanthopterygians, “state 0”), Endo clearly claimed that “state 1” is the only feature which is supposed to group some Gadiforms i.e. is a putative synapomorphy. The hierarchical representation of this statement leads to the creation of one nested class: taxa coded “1” instantiate this less inclusive class, taxa coded “0” instantiate the most inclusive class (Fig.3). The root does not provide any hypothesis of synapomorphy; thus the binary matrix character corresponds to a character with a single state in 3ia. Only putative synapomorphies are regarded as states in 3is.

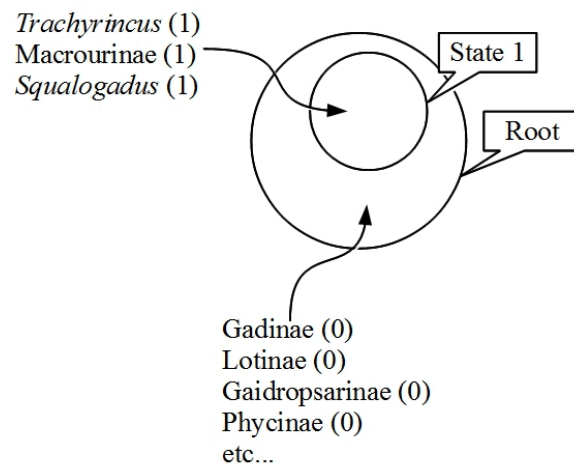


FIGURE 3. Hierarchical representation of the homology hypothesis regarding the “Presence of anterolateral processes on lachrymal” and placement of taxa so as to build a character. Taxa displaying a state which is not regarded as a putative synapomorphy are placed at the root of the hierarchy, i.e. the most inclusive class.

For the second analysis, 12 characters out of 23 follow Endo’s assumptions and 11 characters differ from the first analysis following the 3ia theoretical foundations. For instance, the constraint concerning states present in a single taxon is exemplified by the character 4 “*Adductor mandibular* A2 covering *levator arcus palatine* (lap)” (Endo 2002:85), which describes the three possible shapes of A2 muscle regarding the lap. The lateral position (state 0) is considered by Endo as ancestral (as specified by the coding of the outgroup): in 3ia, this condition cannot lead to any hypothesis of a grouping of taxa. State 0 is then represented as the root of the hierarchical character i.e. it is not a proper state in 3ia. Two variations are regarded as independently derived: state 1 and state 2. Both states could have been represented as nested classes, i.e. (0(1)(2)), however they are each instantiated by a single terminal taxon (*Muranolepis* has state 1; Gaidropsarinae has state 2). As in 3ia the states must be putative synapomorphies (grouping at least 2 taxa), a single taxon cannot be represented by a class in itself in the hierarchy, e.g. neither *Muranolepis* nor Gaidropsarinae can be represented as an internal class in itself in the hierarchy. *Muranolepis* and Gaidropsarinae can be placed either at the root (but the character would be uninformative) or in a single class of the hierarchy (Fig.4) characterized by “lap partly or entirely covered by A2”, a property that they both share. Moreover, Endo’s matrix indicates a missing data (?) for Phycinae. In 3ia, taxa for which data are missing are absent from the character, they are not placed in any class of the hierarchy. Missing data are treated as truly missing.

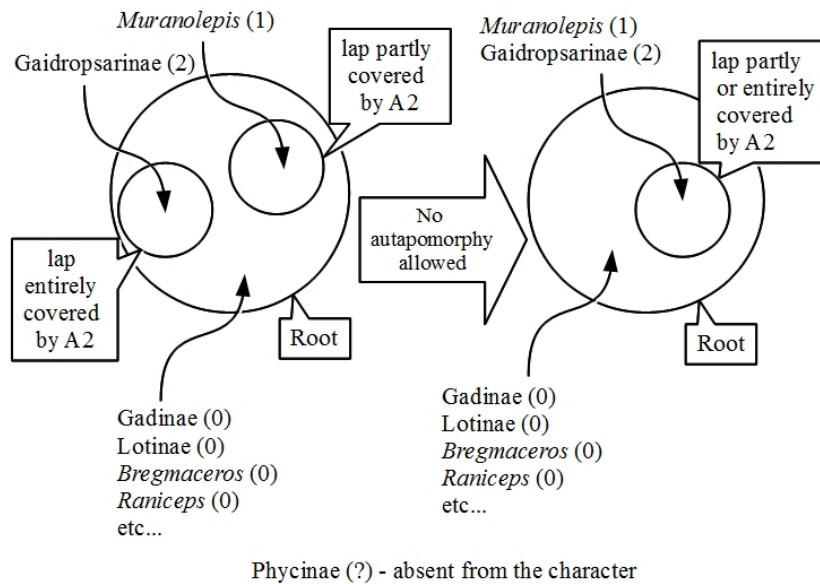


FIGURE 4. Two competing homology hypotheses regarding the “*Adductor mandibular* A2 covering *levator arcus palatine* (lap)”. In 3ia, the retained hypothesis must be the one without autapomorphies as classes represent putative synapomorphies. Taxa instantiate the classes so as to build a character. Those which do not display a state which is regarded as a putative synapomorphy are placed at the root of the hierarchy, i.e. the most inclusive class. Taxa with missing data are absent from the character, e.g. Phycinae here.

The consideration of inapplicability conditions is exemplified by the character 12 “dorsal gill-arch formation” (Endo 2002:89). Endo suggests “a transformation series for the dorsal gill-arch is proposed here, based on the condition of pharyngobranchial 1 (= Pb 1) and the interarcual ligament and cartilage.” This character has 6 states. In 3ia, it could have been represented as the inclusion of classes from the root of the hierarchy. However, since the outgroup taxon is coded as “?”, Endo did not specify which state represents the ancestral condition. The character representation provided in the supplementary data complies with the most parsimonious path for the transformation series, according to the optimization of character on the phylogenetic tree. This option has been chosen not to introduce other hypotheses than Endo’s in our characters, given that we do not aim at revisiting Endo’s work. However, the character “dorsal gill-arch formation” is particularly interesting for the representation of logical relationships in 3ia due to inapplicable data and we emphasize here an alternative representation. The cartilaginous or bony nature of Pb1 characterizes some Gadiforms and allows their grouping i.e. 2 classes in the hierarchy. According to Endo’s expertise, this character is meaningful when considering the combination of Pb1 with interarcual ligament (ial) and interarcual cartilage (iac). Ial and iac exist under several states, and it would be meaningless to consider the apparition of these states independently according to the nature of Pb1, i.e. it would be meaningless to represent twice the states of ial and iac as classes included in both “cartilaginous Pb1” and “bony Pb1” classes (Fig.5). The condition of ial and iac is logically more general than the tissular nature of Pb1. We chose creating a class labelled “presence of ial and iac” itself containing a single class “Presence of Pb1”. Taxa that present no elements of the dorsal gill-arch are placed in the most inclusive class: other states are inapplicable for them. Taxa presenting ial and iac but no Pb1 are placed in the “Presence of ial and iac” class: the state “Presence of Pb1” is inapplicable for them. Taxa presenting Pb1 are placed in one of the two classes included in the class “Presence of Pb1”, either “bony Pb1” or “cartilaginous Pb1”. All taxa placed in less inclusive classes inherit characteristics from the parent class i.e. taxa presenting Pb1 automatically inherit the presence of ial and iac. However, *Raniceps* for instance does not show iac. With our logical representation of the character, this lack of iac may be interpreted as a secondarily loss, contradicting the claim that the existence of secondary losses is denied in 3ia (Kluge 1994; Farris 1997; Farris 2010).

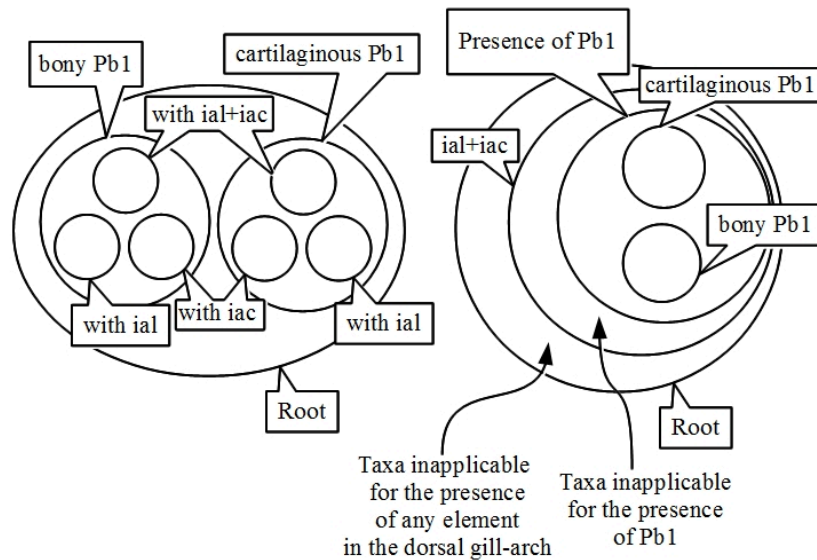


FIGURE 5. Two competing homology hypotheses regarding the “dorsal gill-arch formation”. The tissular nature of Pb1 is regarded more (left) or less (right) general than the presence of ial and iac. Taxa with inapplicable data are placed in the general class in order not to inherit some characteristics from the parent classes.

Phylogeny topologies and synapomorphies. Both of our analyses (Fig. 1A, B) recovered the clade B (grouping Gadinae and Lotinae) and the clade C (grouping Gaidropsarinae and Phycinae) found by Endo (Fig. 1C). Endo’s optimization of characters provided two synapomorphies characterizing the clade B (Fig. 2D): the dorsal supernumerary fin-ray represented by a single soft-ray (18-4) and the secondary loss of the X and Y bones (21-2). The 3ia method of traceability of character history retained the secondary loss of X and Y bones (21-2) as a unique synapomorphy (Fig. 2C). Since the presence of a single soft-ray on the first dorsal radial (18-4) appeared twice in the phylogeny, it would have been interpreted as a homoplasy with the 3ia method. Our first analysis provided 21-2 as a synapomorphy for the clade B (Fig. 2A) but not the second analysis (Fig. 2B) in which the secondary loss is not considered as a putative synapomorphy (Table 1). Endo’s optimization provided three synapomorphies to corroborate the monophyly of the clade C (Fig. 2D): *adductor mandibular* A2 almost covering *levator arcus palatini* (4-2), a unique palatopterygoid formation (6-1) and a long median process of pelvic girdle directed posteromedially (16-2). The 3ia method of traceability of character history (Fig. 2C) provided the same support. Both of our analyses did not recover the *adductor mandibular* A2 almost covering *levator arcus palatini* (4-2) as a synapomorphy of the clade C because the character 4 is uninformative for 3ia (Table 1). Both analyses recovered the two synapomorphies 6-1 and 16-2 (Fig. 2A, B).

Contrary to Endo’s strict consensus (Fig. 1C), both of our analyses (Fig. 1A, B) recovered the clade A (Gadidae) as also found by Roa-Varón & Ortí’s (2009) analysis of molecular data. In 3ia, relationships from the characters 3 (*Adductor mandibular* A1 formation), 7 (Presence of lower process of hyomandibular) and 13 (*Hyohyoideus abductores* 2

formation) supported the clade A, but none of these character states were retained as synapomorphies because they also appeared at other nodes of the trees.

Our first analysis (Fig. 1A) recovered the clade D (grouping *Bregmaceros* and *Muranolepis*) found by Endo (Fig. 1C). Endo's optimization (Fig. 2D) provided 5 synapomorphies for this clade: posterior Al divided into lateral Ala and medial Alb

(3-1), the absence of endopterygoid (5-2), the presence of many actinosts (15-1), a specialized pelvic girdle (16-3) and the absence of epipleurals (17-1). The 3ia method of traceability of character history did not retain 3-1, 15-1 and 17-1 as synapomorphies (Fig. 2C). On Endo's strict consensus (Fig. 2C) as well as on the tree from our first analysis (Fig. 2A), *Bregmaceros* and *Muranolepis* shared the absence of endopterygoid (5-2) and a specialized pelvic girdle (16-3) according to the 3ia method. With our second analysis, neither the states "absence" nor the presence of a specialized pelvic girdle (16-3) which is present in a single taxon were coded as putative synapomorphies within the characters. As there is no support for the clade D, this clade was not recovered by our second analysis (Fig. 2B).

The clade E found by Endo (Fig. 1C) was recovered by our first analysis (Fig. 1A) but not our second (Fig. 1B) due to the position of *Steindachneria* and *Bathygadinae*. Endo's optimization of characters provided two synapomorphies for the clade E (Fig. 2D): a transverse median process on pelvic girdle (16-0) and the absence of caudal fin (20-1). The 3ia method of traceability of the character history also retained both states as synapomorphies for the clade E on Endo's strict consensus (Fig. 2C), as well as on the tree from our first analysis (Fig. 2A).

None of our analysis recovered the clade F (Fig. 1A, B) found by Endo (Fig. 1C). Endo's optimization of characters provided 2 synapomorphies for this clade (Fig. 2D): the presence of cartilaginous Pb 1 and both interarcual elements (12-1) and a reduced number of actinosts (15-2). However, the 3ia method of traceability of character history did not retain any of these states as synapomorphies. Thus, this is not surprising that the clade F was not recovered by our analyses.

Both of our analyses recovered the clade G (Fig. 1A, B) found by Endo (Fig. 1C). Endo's optimization of characters characterized this clade with 5 synapomorphies (Fig. 2D): the enlargement of nasals (1-1), the presence of an anterolateral processes on lachrymal (2-1), the normal condition of *rectus communis* (11-0), the absence of Pb 1 and presence of interarcual elements (12-2) and the presence of spinoid scales (23-1). The 3ia method of traceability of character history did not retained 11-0 as a synapomorphy (Fig. 2C). Both of our analyses did retain neither 11-0 nor 12-2 as synapomorphies for the clade F (Fig. 2A, B) because 12-2 is present in the single taxon *Trachyrincus* and could not be a putative synapomorphy within the character.

Both of our analyses recovered the clade H (Fig. 1A, B), grouping *Trachyrincus* and *Squalogadus* (*Trachyrincidae sensu* Roa-Varón & Ortí 2009), found by Endo (Fig. 1C). Endo's optimization of characters supported the clade H with 5 synapomorphies (Fig. 2D): two branchiostegal rays on epihyal (9-0), many actinosts (15-1), a posteromedial extension of the median process of the pelvic girdle (16-1), two dorsal supernumerary fin-rays with one pseudospine (18-1) and the reduction of pelvic fins (19-1). The 3ia method of traceability of characters did not retain 15-1 as a synapomorphy for this clade (Fig. 2C). Both of our analyses recovered two branchiostegal rays on epihyal (9-0) and the reduction of pelvic fins (19-1) as synapomorphies (Fig. 2A, B) for the clade H.

Intersection tree. The intersection tree (Cao *et al.* 2009) reconstructed by LisBeth is a summary tree of greater significance than a consensus tree: It summarizes all the information (in terms of relationships) that is common to the optimal trees, and only this information. The intersection tree has the desirable characteristics of consensus trees (Wilkinson 1994b). It is an accurate summary as exemplified by the intersection of Endo's 11 optimal trees (2002): the intersection tree topology is exactly congruent with the strict consensus tree (the clades B, C, D, E, F, G, H are recovered) and even more resolved since it yields the clade A (Fig. 1.D). Since Endo's strict consensus did not yield this clade, Endo (2002) discussed one of the shortest trees presenting it (2002), hence our choice of emphasizing it.

Other considerations. Another consequence of hierarchical character representation is allowing a clear distinction between orthology and paralogy (Zaragüeta Bagils *et al.* 2004) as the identification of orthologs remains an important problem in the field of comparative systematics. The 3ia method of traceability of character history relies on this distinction.

Numerous question marks within a taxon x character matrix are still problematic. This problem is inherently avoided with our hierarchical character representation where taxa with missing data are truly absent from characters.

LisBeth's representational model can be extended to problems in biogeography since the source of biogeographic characters are cladograms of taxa, these too are hierarchical trees.

Conclusion

There are fewer differences between our first 3ia analysis and Endo's strict consensus (2002) than between our second 3ia analysis and Endo's strict consensus. This result is not surprising since, for our first analysis, the representation of all of the 23 characters followed Endo's evolutionary assumptions. What is surprising is that the second analysis does not provide a meaningless result, even if 11 of the characters were represented differently (Table 1). For this second analysis, we observed all of the constraints promoted by the 3ia foundations, e.g. a "secondarily loss" is not a putative synapomorphy within the character, a state present in a single taxon does not lead to represent such taxon as an internal class by itself in the character hierarchy, taxa for which data are missing are truly absent from the character, taxa for which data are inapplicable are not placed in the less inclusive classes in the character hierarchy. Considering such constraints, some authors consider 3ia to be more a mathematical method than a biologic method (pers. comm.). Contrary to that claim, and with regard to our results, we consider 3ia to be relevant to deal with real biologic data. Researchers who are used to conduct parsimony analyses may have some hesitations to represent constraints as "secondarily losses are not putative synapomorphies", but we consider important to remain consistent with the theoretical foundations of each method. Such foundations indicate how to deal with hypotheses and constraints in a proper way, i.e. in order to recover the phylogenetic signal.

Not to take into account a secondarily loss in the character does not mean that 3ia forbids loss (or reversion): the traceability of character history can lead to interpret some patterns as secondarily loss events (among other interpretations). However, they are neither considered synapomorphies nor homoplasies, as in parsimony methods

because they are not tested as hypotheses of homologous features. For instance, with our second analysis, the state 4-1 (character 4 being “*Adductor mandibular A2* covering *levator arcus palatine*”) which is “lap partly or entirely covered by A2” characterizes a clade including *Muranolepis*, *Raniceps*, *Meluccius* and the Gadidae. However, Gadinae, Lotinae, *Raniceps* and *Merluccius* do not present such state. Here, this absence can be interpreted as a secondarily loss. LisBeth’s user-friendly interface allows for an easy handling of items thanks to a pointing device gesture, i.e. drag-and-drop (see the Appendix, Fig. 8-10). This intuitive interface makes it easy for both novice and expert users to start and conduct 3ia studies. LisBeth satisfies the long-standing need for a user-friendly and intuitive environment in 3ia.

Acknowledgements

We thank Malte C. Ebach and Anthony C. Gill for fruitful discussions. We warmly thank an anonymous reviewer for his advices and English improvements. We also thank all LisBeth users (from systematic and biogeographic community) for their feedbacks and encouragements in enhancing the software.

This work is supported by the French Ministry of Higher Education and Research (Ministère de l’Éducation Nationale et de la Recherche) and the Centre National de la Recherche Scientifique.

Appendix

LisBeth: Functions (All figures are screen-shots from LisBeth v.1.3.)

Taxa edition – Under the “Taxa edition” mode, taxon names may be entered, renamed, removed from the list and sorted by name or by code (Fig. 6). For instance, in Fig. 6, Gadinae, Lotinae and Gaidropsarinae have already been entered and Phycinae is being typed. LisBeth generates a text file (with the extension “.3ia”), which stores the list of taxa.

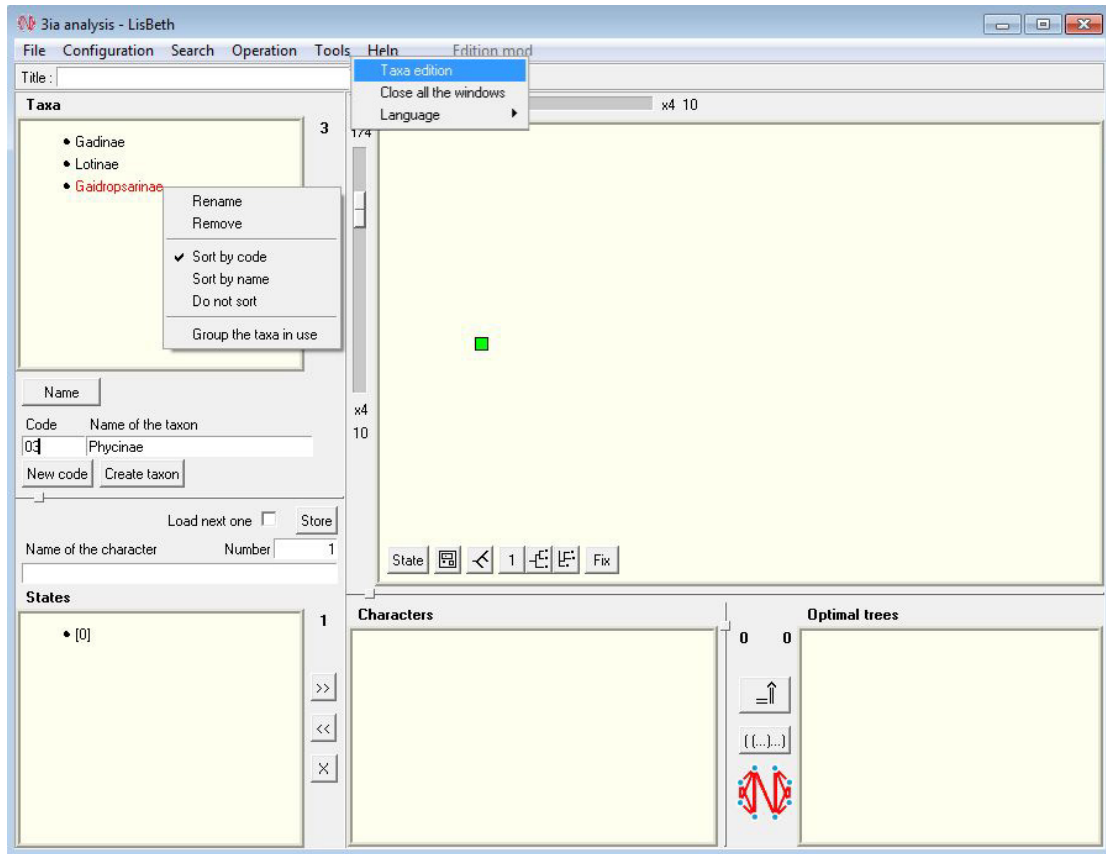


FIGURE 6. Tool menu: “Taxon Edition mode”. Taxa names are entered in the “Taxa” window.

Character edition – Hierarchical characters are named and can be represented as Venn diagrams or rooted trees, both representations being equivalent (use the Tree / Venn diagram button, Fig. 7). Under the Venn diagram representation, included states (i.e. classes of the Venn diagram) are built using <Ctrl-Click> and can be named using <Shift-Click>. Characters are stored in the “Characters” window. In Fig. 7, the “nasal morphology” character has already been stored and the “presence of anterolateral processes on lachrymal” character is being built, displaying one informative state (“present”). Taxa are then dragged in their corresponding class, e.g. *Trachyrincus*, *Macrourinae* and *Squalogadus* present the anterolateral process on the lachrymal (Fig. 8); all of the other taxa are dragged in the most general class, i.e. the root of the hierarchy

(Fig. 9). This hierarchical representation allows for distinguishing missing and inapplicable data (Zaragüeta Bagils & Bourdon 2007): when information is missing, the corresponding taxa are not dragged into the character; when an informative state is irrelevant for a taxon, this taxon is placed in the general class. The rooted tree representation is particularly useful for the modification of taxa connection. For instance, Fig. 10 shows how *Squalogadus* can be dragged from the included class to the inclusive one. LisBeth automatically writes characters in the text file.3ia under a given syntax (for more details on the syntax, see LisBeth documentation; the characters of our analysis are provided as supplementary data).

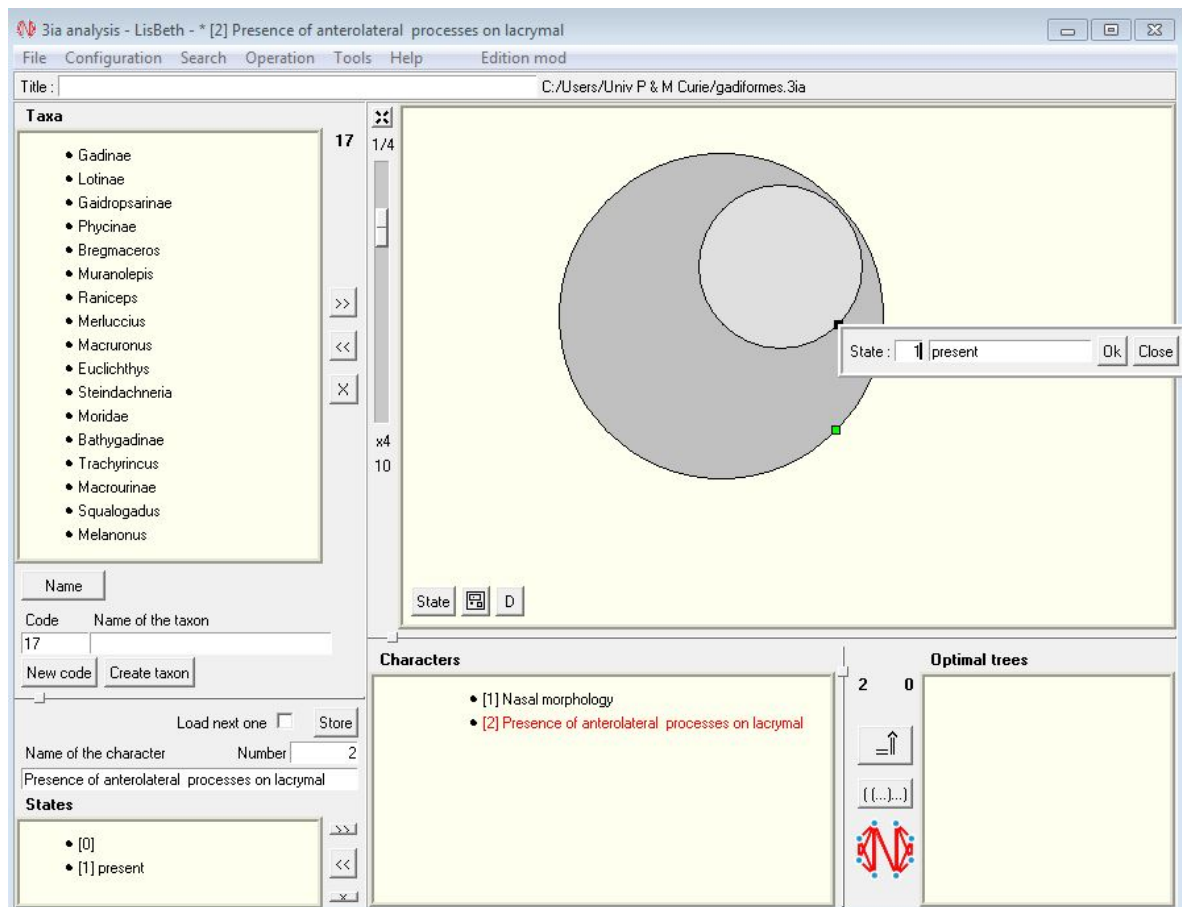


FIGURE 7. Venn diagram representation for the character “presence of anterolateral processes on lachrymal” displaying one informative state (“present”).

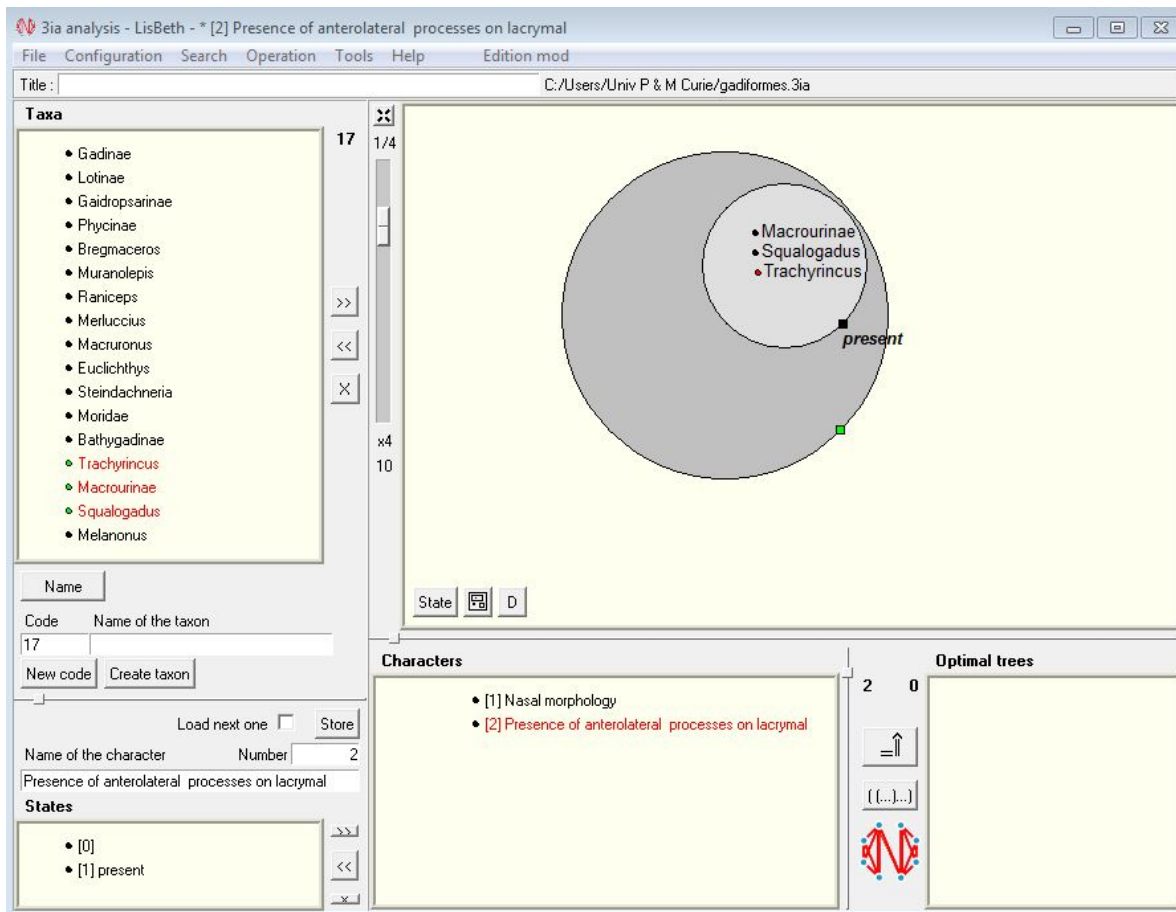


FIGURE 8. Venn diagram representation for the character “presence of anterolateral processes on lachrymal” for which Macrourinae, *Squalogadus* and *Trachyrincus* taxa instantiate the state “present”. Selected taxa (in red) are simply dragged from the taxa list and dropped in the main window.

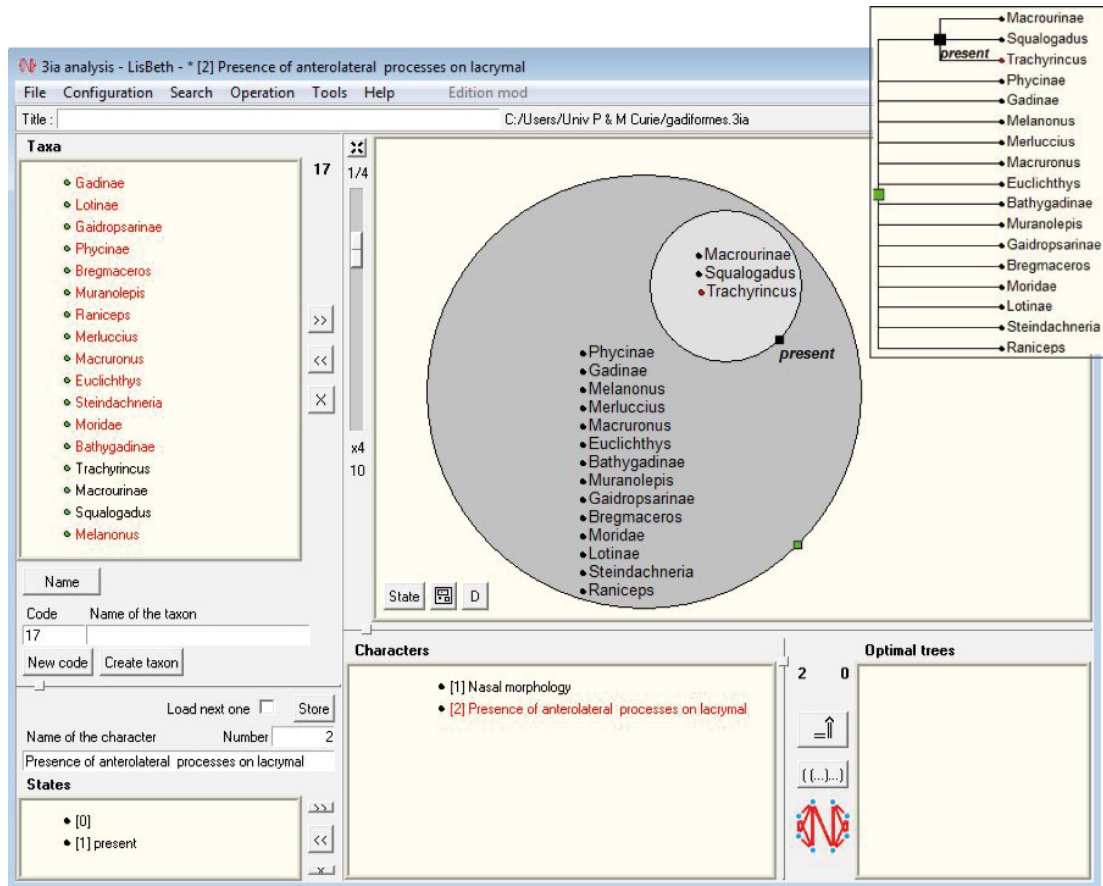


FIGURE 9. Venn diagram and tree representations for the character “presence of anterolateral processes on lacrymal”. Selected taxa (in red) are simply dragged from the taxa list and dropped in the main window.

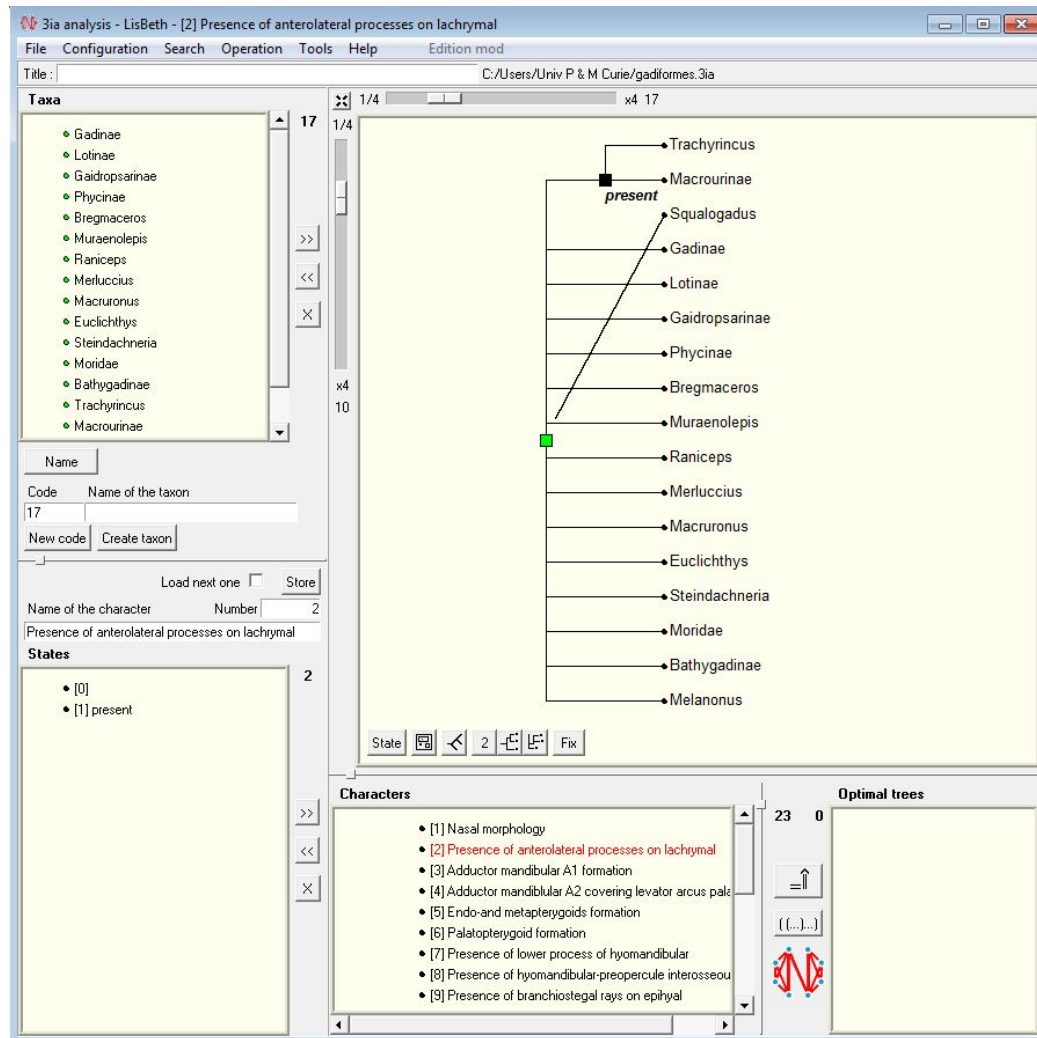


FIGURE 10. Modification of *Squalogadus* position within the tree, the branch is dragged from one node and dropped to another.

Character analysis – From the list of characters, LisBeth computes their three-item statements (3is) with their fractional weights. A text file (with the extension “.3iz”) is automatically generated.

Search – LisBeth implements branch and bound algorithms, applied with the “Search for the compatibility trees” command (Fig. 11). The “Execute 3ia tool” windows specify defaults for the search options (Fig 12), e.g. “branch and bound”, “load the target file when finished”, etc. The information processing is detailed in the “LisBeth window” (Fig. 13), e.g. number of retained trees, elapsed time (in % of the search), etc... The optimal trees are stored in the file.3iz. They are listed in the interface “Optimal trees” window (Fig.14).

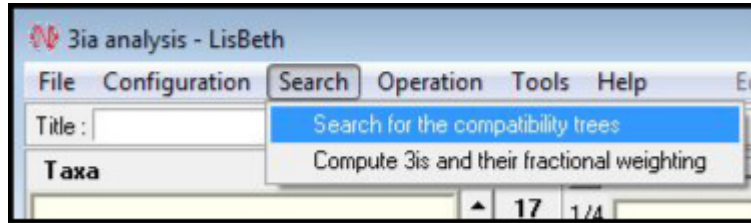


FIGURE 11. Search menu: search for compatible trees.

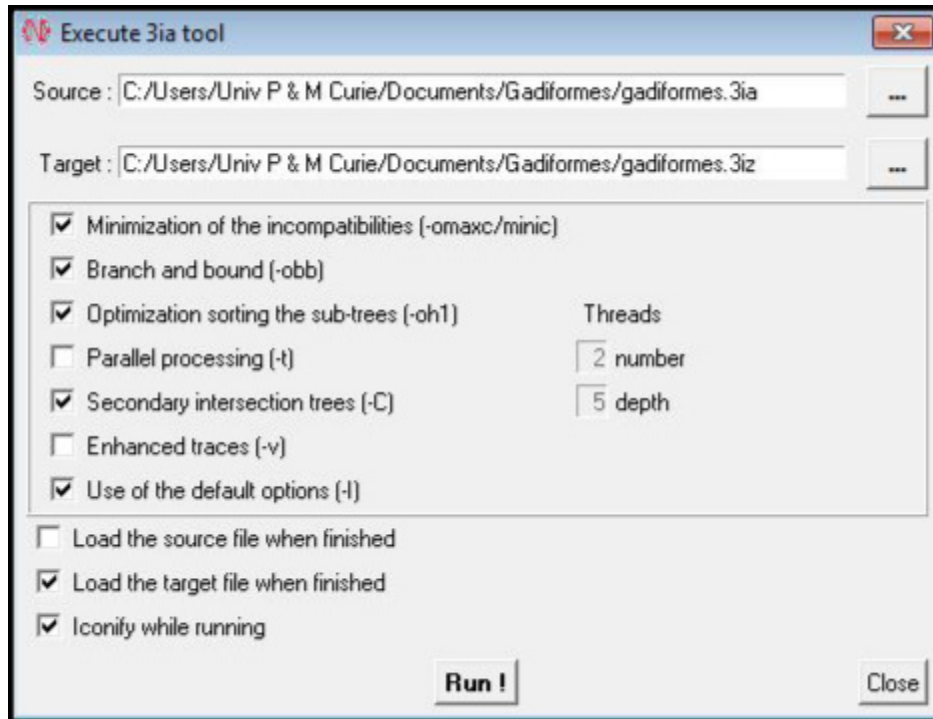
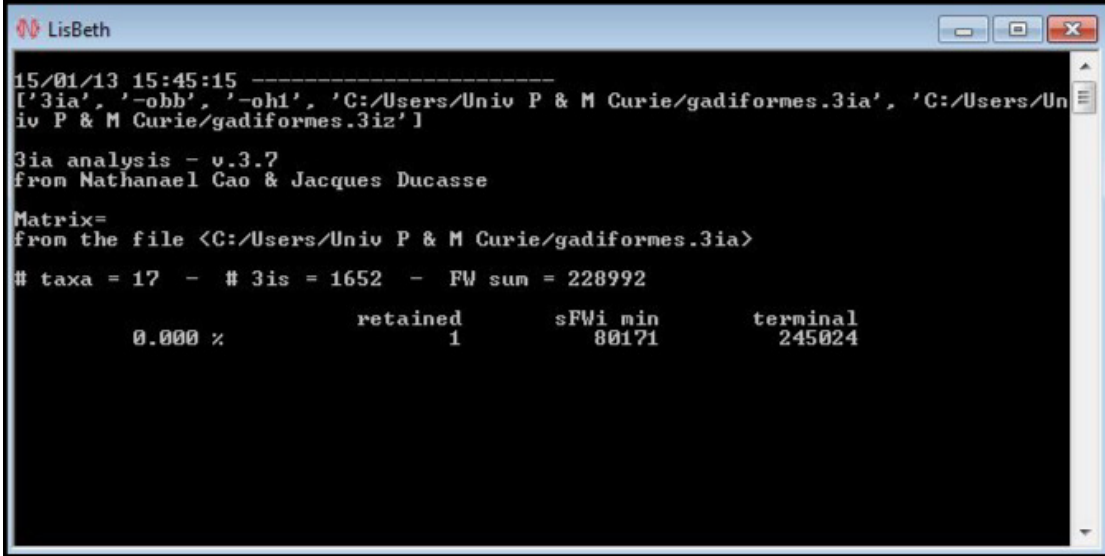


FIGURE 12. Details of the Executive 3ia tool, main options are selected by default.



```
LisBeth
15/01/13 15:45:15 -----
['3ia', '-obb', '-oh1', 'C:/Users/Univ P & M Curie/gadiformes.3ia', 'C:/Users/Univ P & M Curie/gadiformes.3iz']
3ia analysis - v.3.7
from Nathanael Cao & Jacques Ducasse

Matrix=
from the file <C:/Users/Univ P & M Curie/gadiformes.3ia>

# taxa = 17 - # 3is = 1652 - FW sum = 228992

      0.000 %      retained      sFWi min      terminal
                   1             80171          245024
```

FIGURE 13. Processing information details.

Summary tree – LisBeth summarizes the optimal trees reconstructing a tree from the list of their common 3is, hence the name “intersection tree” for the summary tree (Cao *et al.* 2009). This function is given by the default options (Fig.12) (“secondary intersection tree”). With the interface, the intersection tree is always labeled [0] and can be displayed in the main window (Fig. 14).

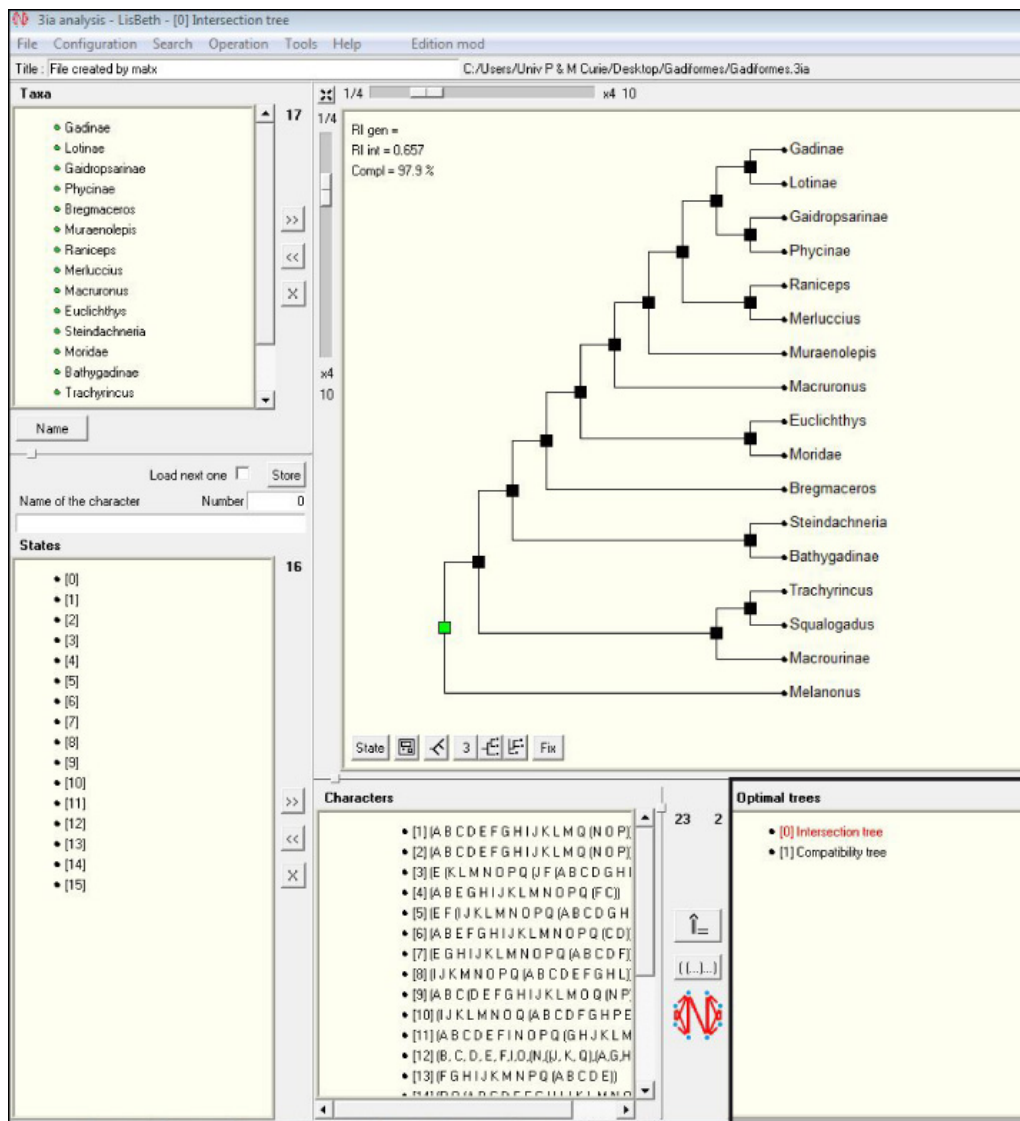


FIGURE 14. Optimal tree window with, in red and labeled [0], the Intersection tree also displayed in the main window.

Save trees – Each tree which is being displayed in the interface main window can be saved as a vector file format, using the “Save image” command from the “File” menu (Fig. 15). Several representations are available via the main window buttons, varying according to the branch relative length, the squared angle / oblique branches or the Venn diagram / rooted tree representation. Trees can be edited by dragging taxa to a given position and saved using the “Fix” button, Fig. 16.

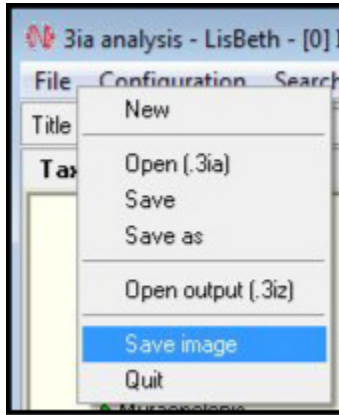


FIGURE 15. File menu: Save image.

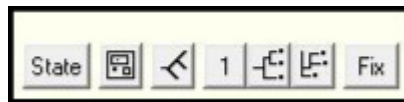


FIGURE 16. Interface buttons specifying representation options e.g. the branch relative length, the squared angle / oblique branches or the Venn diagram / rooted tree representation.

Support – LisBeth calculates various indexes: the per-character RI, the overall-RI and the completeness index (Compl., introduced in Zaragüeta Bagils *et al.* 2012). The overall-RI and Compl. are stored in the file.3iz and displayed in the interface main window (Fig. 14). The per-character-RI values are stored in a table (.3ri) and displayed in front of the listed characters under the “interpretation mode” (see below).

Character history – Character history can be inferred under the “interpretation mode” of the “Operation” menu (Fig. 17). It is displayed (Fig. 18) with coloured boxes listing the character states (e.g. 2:1 stands for state “1” of character “2”), “green” meaning unambiguous synapomorphy and “red” homoplasy. This information is available for the selected characters (e.g. the “Presence of anterolateral processes on lachrymal” character in Fig. 18; for all characters Fig. 19) or for a given node (Fig. 20). The relative placement of coloured boxes can be modified by dragging them to a given position. Under the “interpretation mode”, the “Display the tree” command (Fig. 21) allows for the character lookup in independent windows. Character states can be displayed as well using the “State” button.

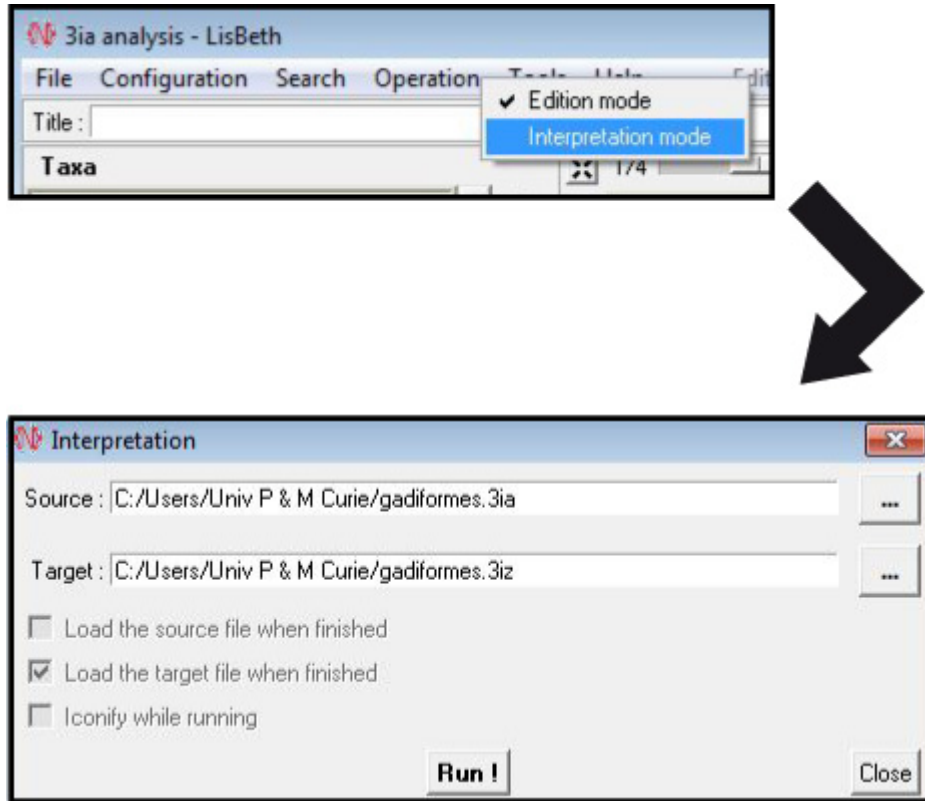


FIGURE 17. Operation menu: Interpretation mode; display of an independent window where the « Source » contains the characters and the « Target » the tree(s).

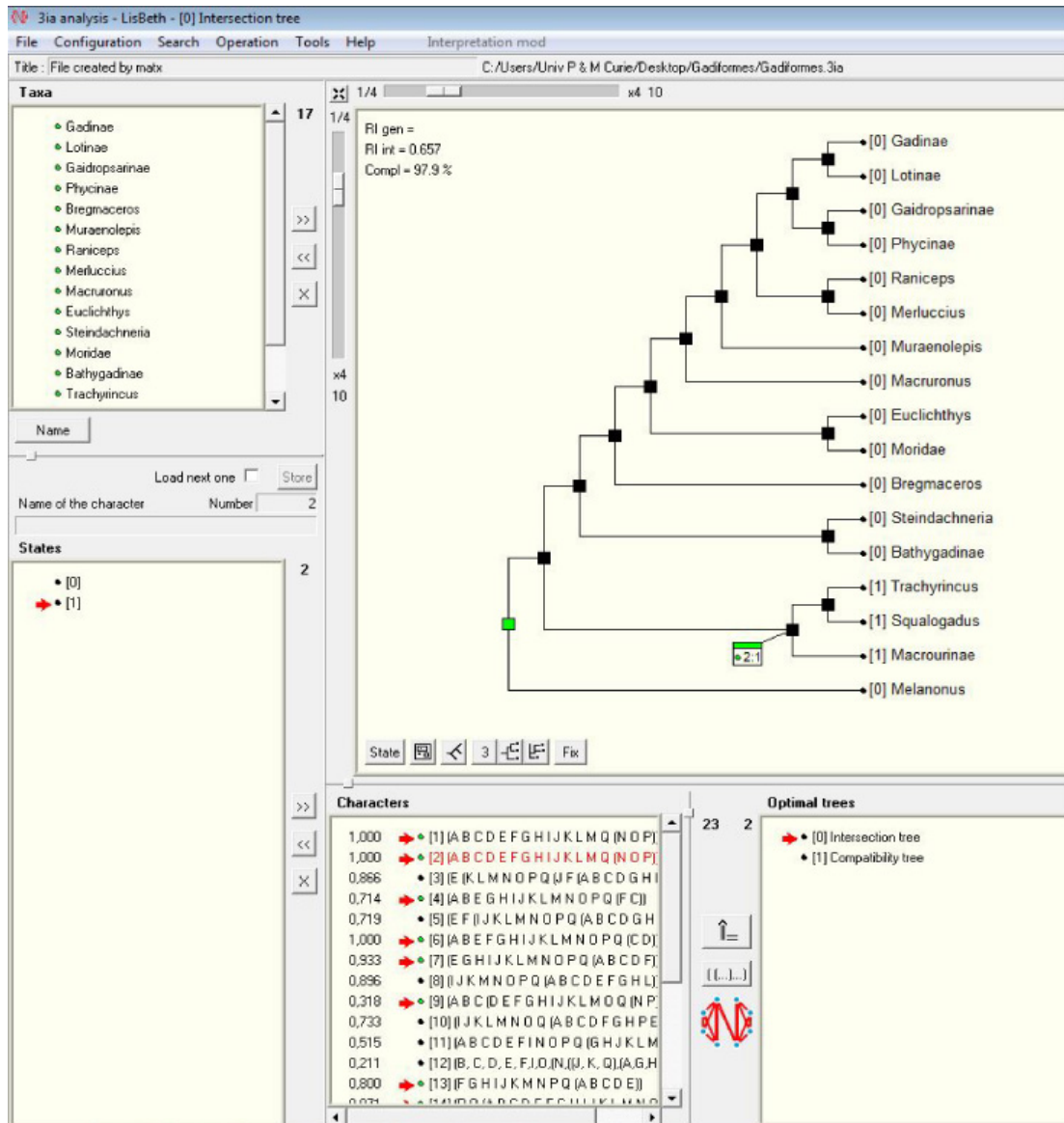


FIGURE 18. The evolution of selected character (in red in the Character window) is traced on the tree of the main window with a code-colour, here green for synapomorphy. For the sake of clarity, each box can be dragged in a given position without changing its link to the node.

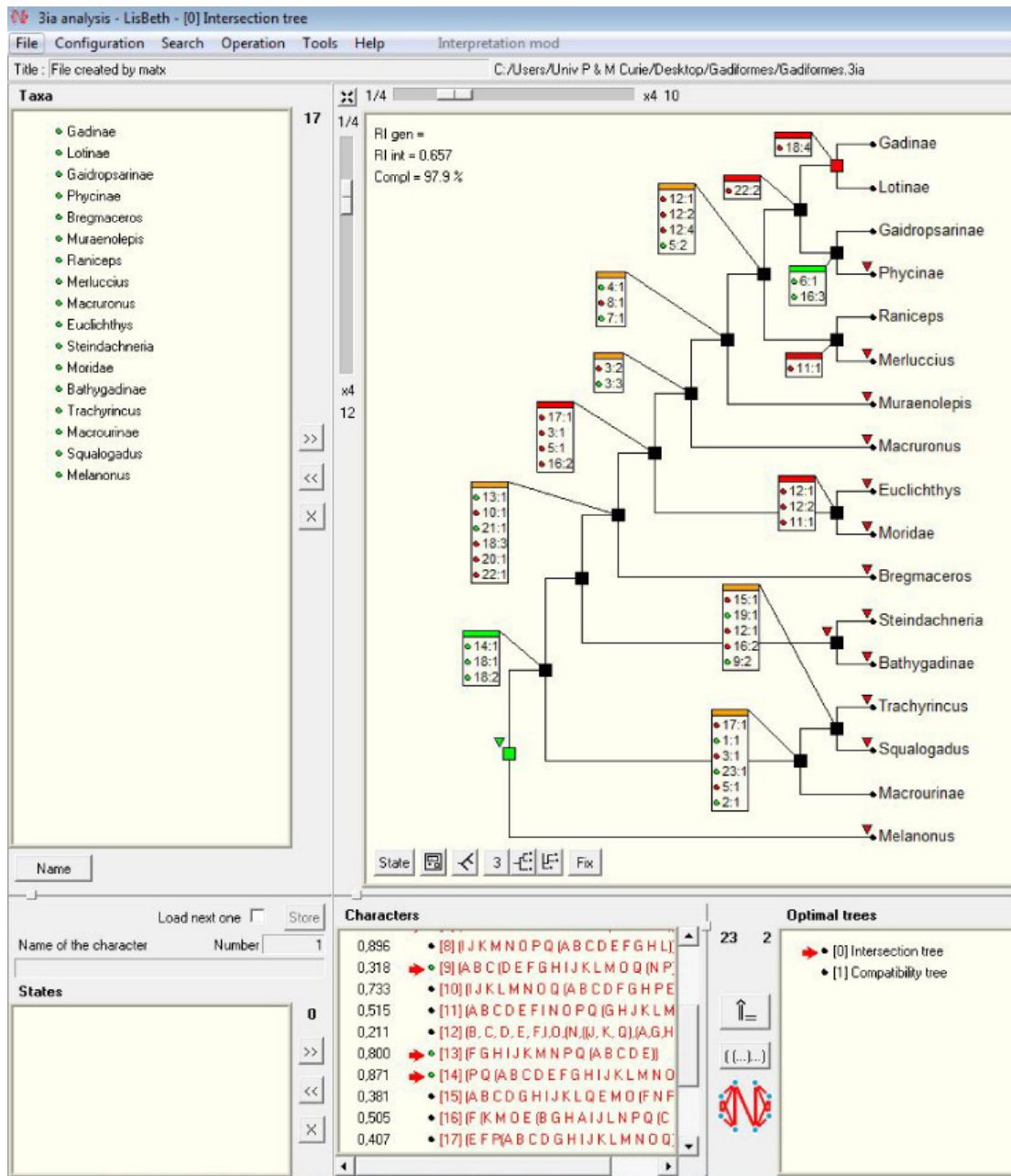


FIGURE 19. The evolution of all characters (in red in the Character window) is traced on the tree of the main window with a code-colour, here green for synapomorphy, red for homoplasy and orange if both synapomorphy and homoplasy are present for a single node. For the sake of clarity, each box can be dragged in a given position without changing its link to the node.

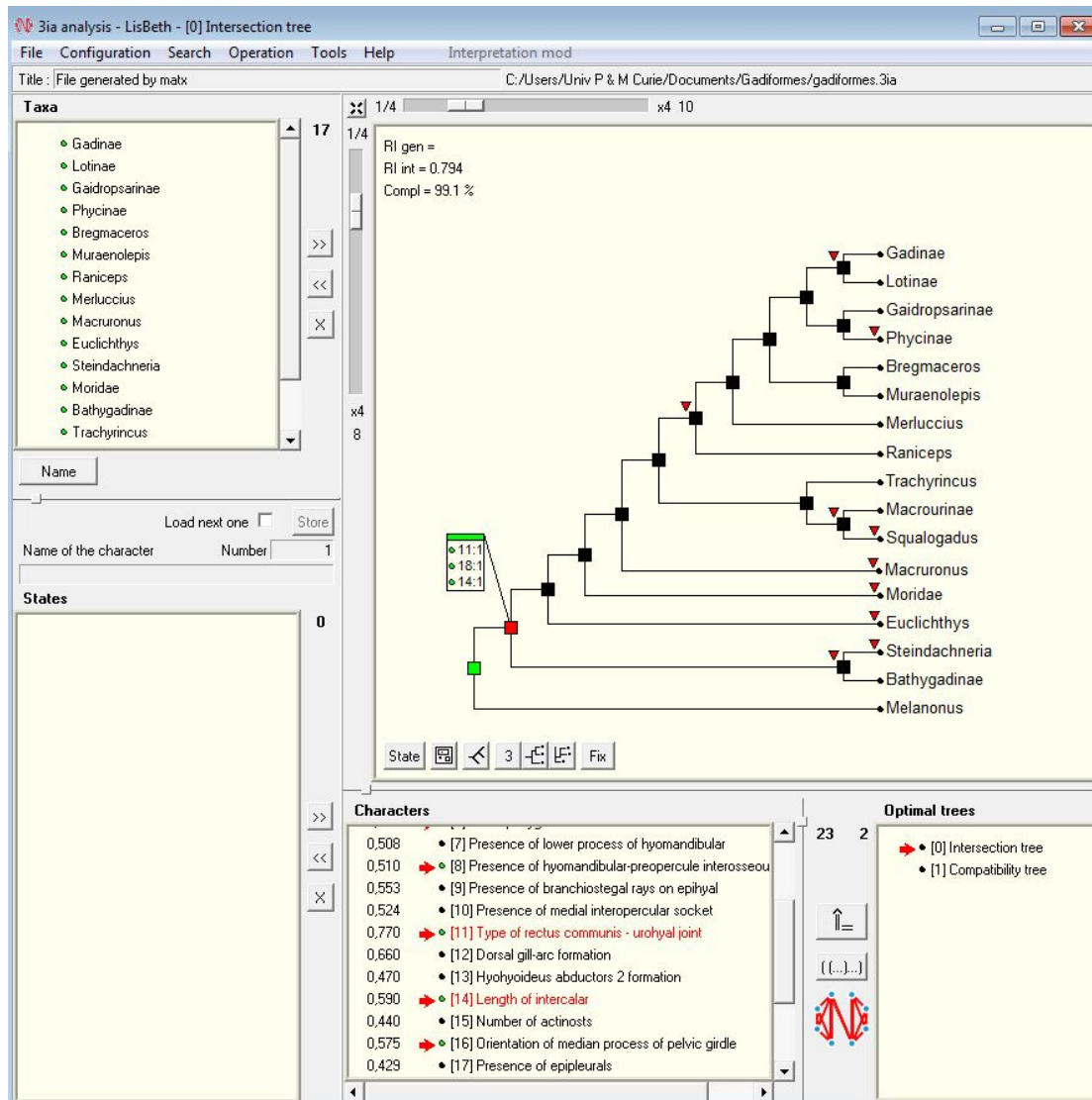


FIGURE 20. Display of supporting synapomorphies for a given node with a code-colour, here green for synapomorphy. Each box can be dragged into a given position without changing its link to the node.

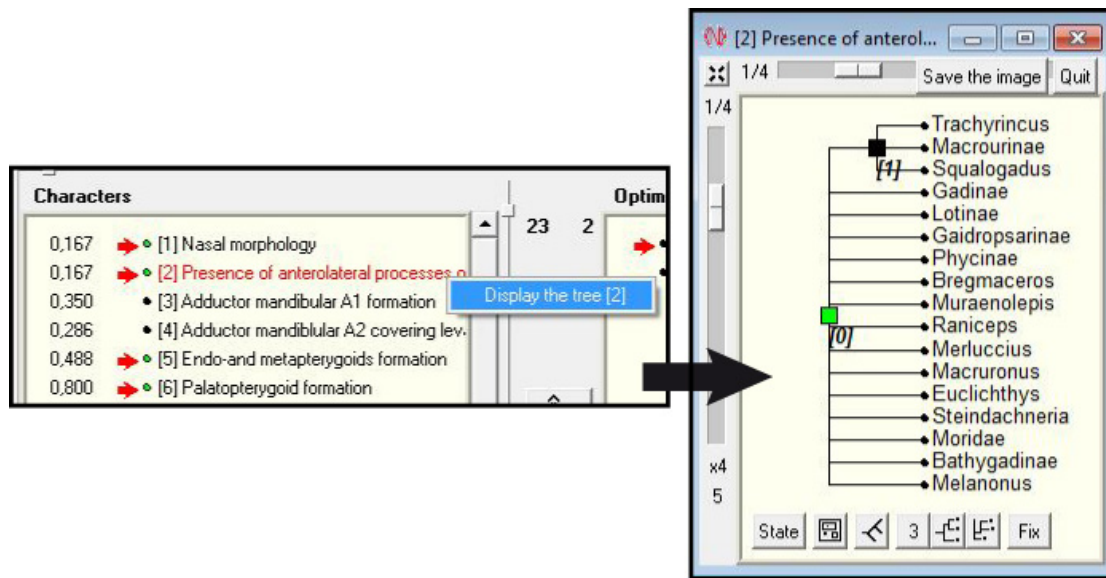


FIGURE 21. Display of any character in an independent window.

Interoperability with other programs and other functionalities

Character export – LisBeth exports 3is matrices (with 3is fractional weights) formatted for other phylogenetic programs (e.g. PAUP* 4b10, Swofford 2002; TNT, Goloboff *et al.* 2008). The “Export” command of LisBeth Toolbox allows for the inclusion of specific commands in the exported nexus or tnt file, e.g. name of the tree file, number of replicates (“replica”) etc...(Fig. 22).

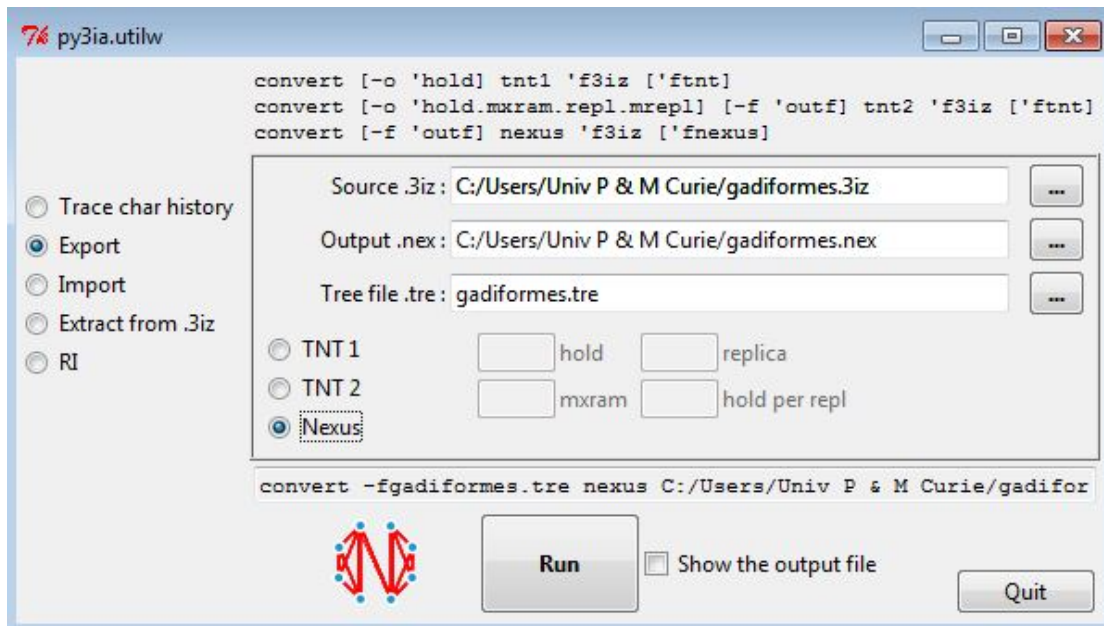


FIGURE 22. Export to Nexus format. The options “hold”, “replica” etc... are available when exporting to TNT format.

Character import – LisBeth imports characters from phylogenetic matrices and from Xper² (Ung *et al.* 2010) knowledge bases. Matrices must be enhanced with hierarchy of states for each character (Fig. 23) so that LisBeth can automatically reconstruct hierarchical characters and generate a text file (.3ia). Xper² program (Fig. 24) implements descriptions of taxa and distinguishes missing data (“unknown state”) from inapplicable data (thanks to parent-child relationships depending on the relevance of descriptors); hierarchical characters are automatically reconstructed from descriptors when specifying a hierarchy of states (Fig. 25).

	1	2
	(0(1))	(0(1))
Gadinae	0	0
Lotinae	0	0
Gaidropsarinae	0	0
Phycinae	0	0
Bregmaceros	0	0
Muraenolepis	0	0
Raniceps	0	0
Merluccius	0	0
Macruronus	0	0
Euclichthys	0	0
Steindachneria	0	0
Moridae	0	0
Bathygadinae	0	0
Trachyrincus	1	1
Macrourinae	1	1
Squalogadus	1	1
Melanonus	0	0

FIGURE 23. Character state hierarchy enhanced matrix.

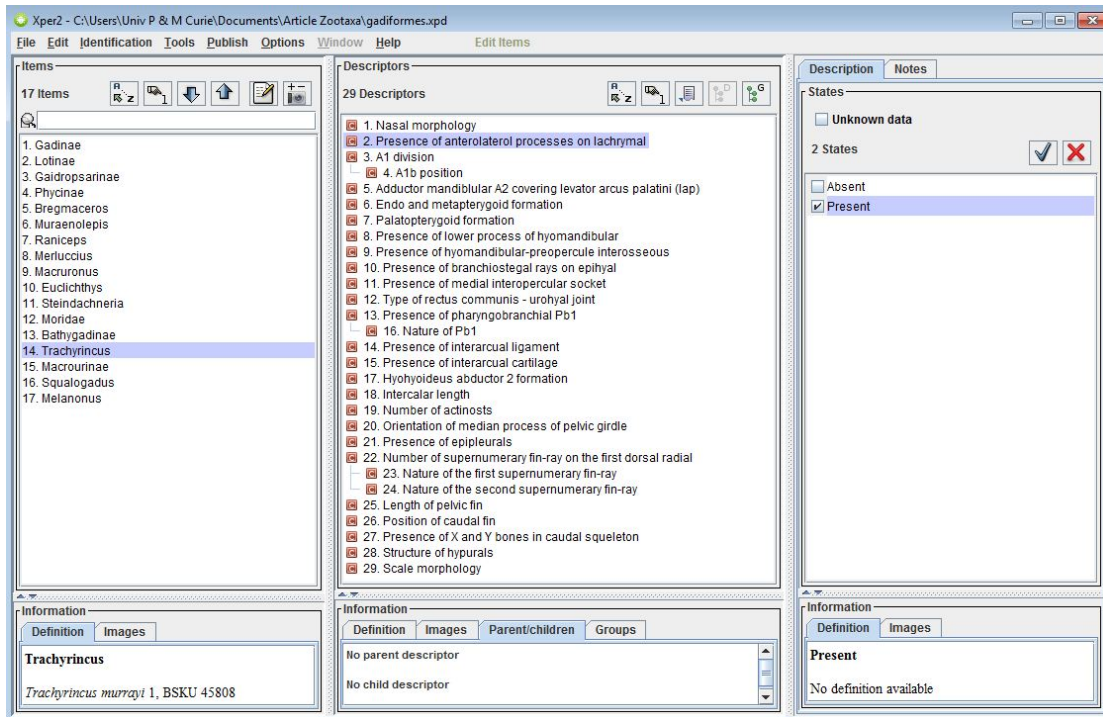


FIGURE 24. Xper² « Edit items mode » screen-shot displaying descriptions of taxa.

```

Echantillonage
1-17
;
Representation-3ia
[1] (1:1 (1:2))
[2] (2:1 (2:2))
[3] (3:1 ((4:1) (4:2)))
;

```

FIGURE 25. Syntax for the import of phylogenetic characters from Xper² program. [1] stands for the character label. 2:1 stands for the state “1” of the second Xper² descriptor. A character is represented as a hierarchy of states. A single character (e.g. [3]) can include states from several Xper2 descriptors, e.g. 3rd and 4th) since hierarchy allows for the representation of inherited relationships.

Tree import – The “Import” command of LisBeth Toolbox imports trees from other phylogenetic programs, e.g. PAUP* 4b10 and TNT (Fig. 26).

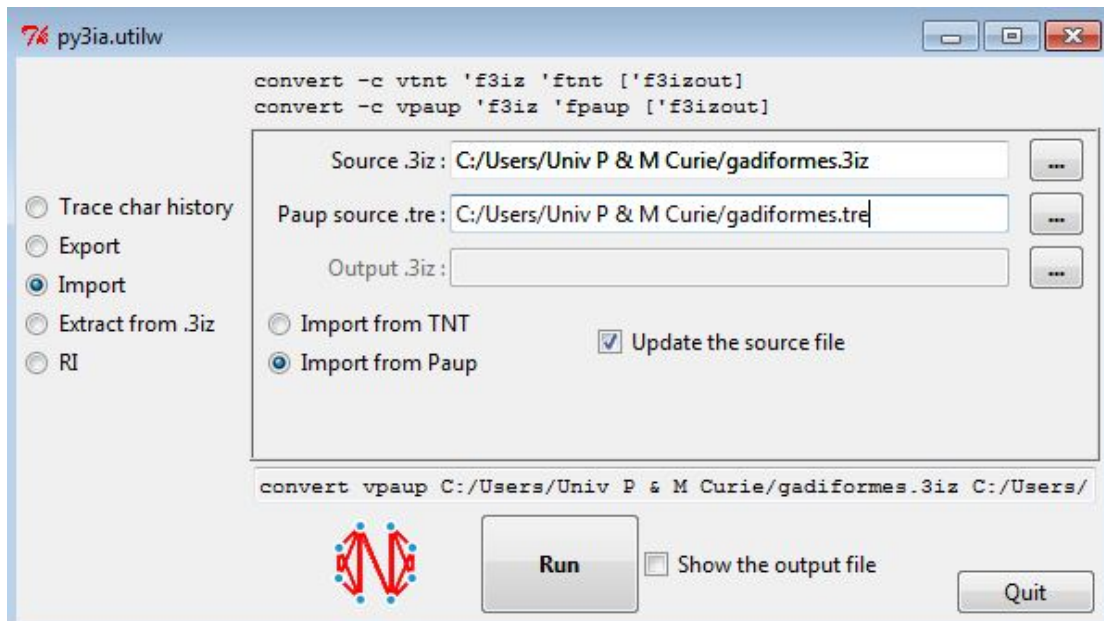


FIGURE 26. Import trees from PAUP* 4b10.

Extract specific information – The “Extract from .3iz” command of LisBeth Toolbox allows for the extraction of the selected information from a file (.3iz). The operation presented in Fig. 27 results in the extraction of the Intersection tree from the “gadiformes.3iz” file and its storage in a new file. The table with per-character RI values may also be updated in order to retain values for the intersection tree only.

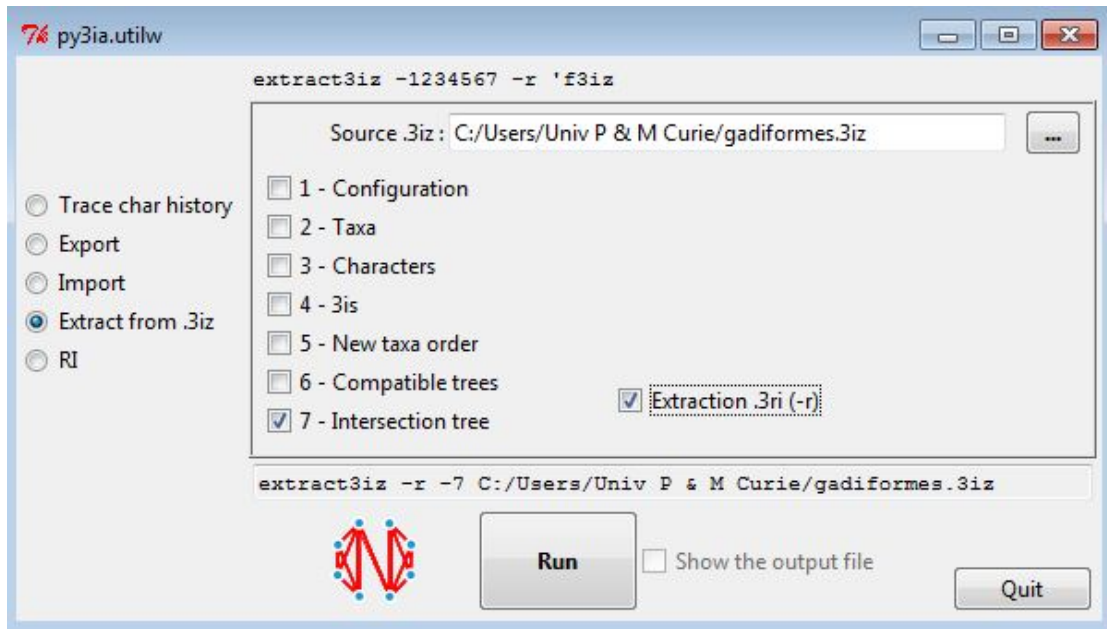


FIGURE 27. Extract the Intersection tree from the 3iz file. The “Extraction .3ri” option is also selected: the table with per-character RI values will present values for the intersection tree only.

RI values – The “RI” command of LisBeth Toolbox calculates the retention index values (Archie 1989; Farris 1989) for any pair of trees or {character, tree} (Fig. 28). This command is particularly helpful when trees have been imported from other programs.

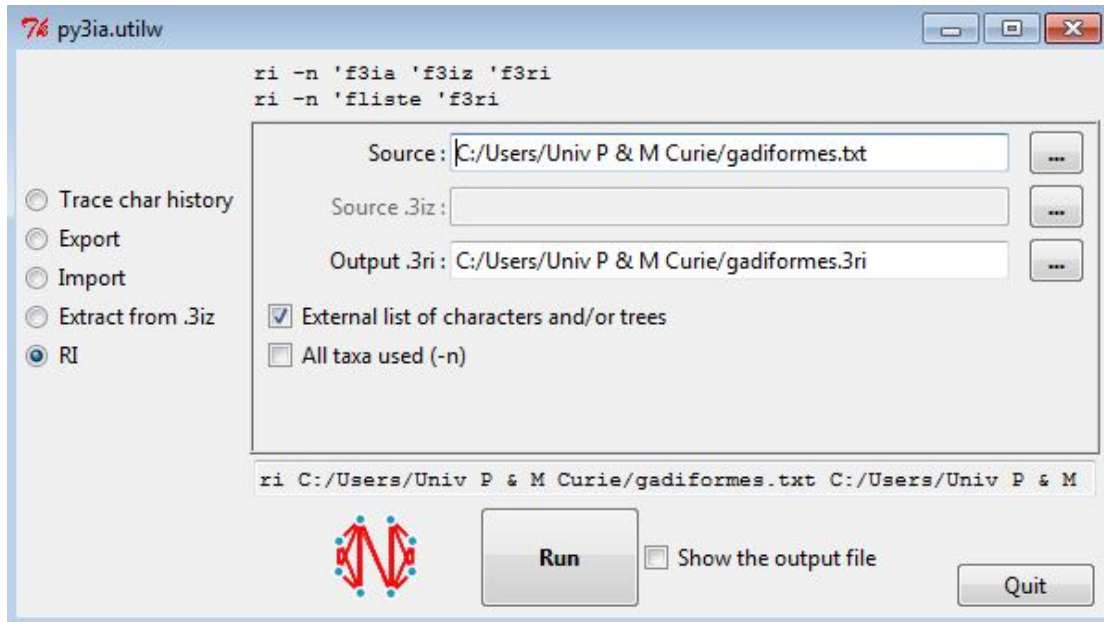


FIGURE 28. Calculation of RI values from external list of characters or trees.

LisBeth is the only software that implements hierarchical representation of characters for 3ia. It is a package of programs primarily written in C language for improved performance, including parallel computation when needed. LisBeth has a user-friendly interface written in Python language (www.python.org).

LisBeth offers a ToolBox with new features providing more flexibility and advanced functionalities. Based on the same ergonomics, the LisBeth ToolBox has a user-friendly interface written in Python language too.

Requirements and availability:

Software name: LisBeth

Version: 1.3 (2013.01.27)

Software homepage:

<http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/lisbeth>

Operating system(s): Windows, Linux version under test.

Programming language: C (executables) and Python (interface)

Other requirements: Python (from 2.7 version) interpreter

License: This program is licensed under a Attribution-NonCommercial- NoDerivative Works 3.0 Unported (CC BY-NC-ND 3.0).

Cost: free.

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. *Systematic Biology*, 38 (3), 253-269.
- Cao, N., Zaragüeta Bagils, R. & Vignes-Lebbe, R. (2007) Hierarchical representation of Hypotheses of Homology. *Geodiversitas*, 29 (1), 5-15.
- Cao, N., Bourdon, E., El Azawi, M. & Zaragüeta Bagils, R. (2009) Three-item analysis and parsimony, intersection tree and strict consensus a biogeographical example. *Bulletin de la Société géologique de France*, 180(1), 13-15.
- Cohen, D.M., Inada, T., Iwamoto T., & Scialabba., N. (1990) *FAO species catalogue. Vol. 10. Gadiform fishes of the world (Order Gadiformes). An annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date. FAO Fisheries Synopsis. No. 125, Vol. 10.* FAO (Eds) Rome, 442 pp.
- Corvez, A. (2012) *L'origine de la Mégaphylle chez les Monilophytes. [PhD dissertation].* Muséum national d'Histoire naturelle, Paris (Eds) France, 276 pp.
- Endo, H. (2002) Phylogeny of the Order Gadiformes (Teleostei, Paracanthopterygii). *Memoirs of the Graduate School of Fisheries Sciences*, 49(2), 75-149.
- Escalante, T., Rodríguez, G., Cao, N., Ebach, M.C. & Morrone J.J. (2007) *Naturwissenschaften*, 94, 561–565.
- Farris, J.S. (1989) The Retention Index and Homoplasy Excess. *Systematic Zoology*, 38(4), 406-407.
- Farris, J.S. (1997) Cycles. *Cladistics*, 13, 131-144.
- Farris, J.S. (2010) Systematic foundering. *Cladistics*, 26, 1-15.
- Farris, J.S. (2012) 3ta Sleeps with the fishes. *Cladistics*, 28(4), 422-436.
- Farris, J.S. & Kluge, A.G. (1998) The Brief History of Three-Taxon Analysis. *Cladistics*, 14, 349-362.
- Gilbert, C.H. & Hubbs, C.L. (1916) Report on the Japanese macrouroid fishes collected by the United States Fisheries steamer Albatross in 1906, with a synopsis of the genera. *Proceedings of the United States National Museum*, 51(2149), 135-214.
- Giorna, M.E. (1809) Mémoire sur des poissons d'espèces nouvelles et de genres nouveaux. Plus: Suite et conclusion du mémoire (pp. 177-180). *Mémoires de l'Académie Imperiale des Sciences, Littérature et Beaux-arts de Turin*, 9, 1-19, 177-180.
- Goloboff, P.A., Farris J.S. & Nixon K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, 24(5), 774-786.
- Goode, G.B. & Bean, P.W. (1896) Oceanic ichthyology. A treatise on the deep-sea and pelagic fishes of the world, based chiefly upon the collections made by the steamers BLADE, ALBATROSS and FISHHAWK in the northwestern Atlantic. *Memoirs of the Museum of Comparative Zoölogy at Harvard College*, 22, 1-553.
- Günther, A.C.L.G. (1873) 1873, Andrew Garrett's fische der Südsee, beschrieben und redigirt von A.C.L.G. Günther. *Journal des Museum Godeffroy*, 2(3), 128.
- Günther, A.C.L.G. (1878) Preliminary notices of deep-sea fishes collected during the voyage of H.M.S. 'Challenger'. *The Annals and Magazine of natural History*, (5)2(7), 17-28, 179-187, 248-251.

- Günther, A.C.L.G. (1880) Report on the shore fishes procured during the voyage of H.M.S. 'Challenger' in the years 1873-1876. *Challenger Report (Zoology)*, 1(6), 1-82.
- Harvey, A.W. (1992) Three-taxon statements: more precisely, an abuse of parsimony? *Cladistics*, 8, 345-354.
- Jordan, D.S. & Evermann, B.W. (1898) The fishes of north and middle America : a descriptive catalogue of the species of fish - like vertebrates found in the waters of North America, north of the Isthmus of Panama. *Bulletin of the United States National Museum*, 47, 2183-3136.
- Kitching, I.J., Foreyn P.L., Christophern J., Humphriesn J. & Williams, D.W. (1998) *Cladistics: the theory and practice of parsimony analysis*. The Oxford University Press (Eds) Oxford, 228 pp.
- Kluge, A.G. (1994) Moving targets and shell games. *Cladistics*, 10, 403-413.
- Ladiges, P.Y., Prober, S.M., & Nelson, G. (1992) Cladistic and biogeographic analysis of the 'blue ash' Eucalypts. *Cladistics*, 8, 103-124.
- Ladiges, P.Y., McFadden, G.I., Middleton, N., Orlovich, D.A., Treloar, N. & Udovicic, F. (1999) Phylogeny of *Melaleuca*, *Callistemon*, and related genera of the *Beaufortia* suballiance (Myrtaceae) based on 5S and ITS-1 spacer regions of nrDNA. *Cladistics*, 15, 151-172.
- Oken, L. (1817) V. Kl. Fische. *Isis (Oken)*, 8(148), 1779-1782.
- McCulloch, A.R. (1926) Report on some fishes obtained by the FIS 'Endeavour' on the coasts of Queensland, New South Wales, Victoria, Tasmania, South and South-western Australia. Part V. *In: Dannevig, H.C. (Eds), Biological results of the fishing experiments carried on by the FIS 'Endeavour'* pp. 157-216.
- Moreau, C.E (1881) Histoire naturelle des poissons de la France. Masson G. (Eds) Paris.
- Morrone, J.J. & Carpenter, J.M. (1994) In search of a method for cladistic biogeography: an empirical comparison of component analysis, Brooks parsimony analysis, and three-area statements. *Cladistics*, 10, 99-153.
- Nelson, G. J. & Platnick, N.I. (1991) Three-taxon statements: A more precise use of parsimony? *Cladistics*, 4:351-366.
- Nelson, G.J. & Ladiges, P.Y. (1992) Information Content and Fractional Weight of Three-Item Statements. *Systematic Biology*, 41(4), 490-494.
- Nelson, G.J. & Ladiges, P.Y. (1993) Missing data and three-item analysis. *Cladistics*, 9, 111-113.
- Nelson, J. (1994) *Fishes of the World. 3rd edition*. John Wiley & Sons (Eds), New York, 600 pp.
- Nelson, J. (2006) *Fishes of the World. 4th edition*. Jhon Wiley & Sons (Eds) New Jersey, 601 pp.
- Parenti, L.R. & **Ebach, M.C.** (2009). *Comparative Biogeography: Discovering and Classifying Biogeographical Patterns of a Dynamic Earth*. University of California Press (Eds) Berkeley, 312 pp.
- Patterson, C. & Johnson, G.D. (1995) The intermuscular bones and ligaments of teleostean fishes. *Smithsonian Contributions in Zoology*, 559, 1-78.
- Rafinesque, C.S. (1810a) *Indice d'ittiologia siciliana; ossia, catalogo metodico dei nomi latini, italiani e siciliani dei pesci, che si rinvencono in Sicilia disposti second un metodo natural e seguito da un'appendice che contiene la descrizione di alcuni nuovi pesci siciliani*. Presso Giovanni del Nobolo (Eds) Messina, 70 pp.
- Rafinesque, C.S. (1810b) *Caratteri di alcuni nuovi generi e nuove specie di animali e piante della Sicilia, con varie osservazioni sopra i medesimi*. Sanfilippo (Eds) Palermo. 1-105.

- Roa-Varón, A. & Ortí, G. (2009) Phylogenetic relationships among families of Gadiformes (Teleostei, Paracanthopterygii) based on nuclear and mitochondrial data. *Molecular Phylogenetics and Evolution*, 52, 688-704.
- Swofford, D.L. (2002) *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4*. Sunderland: Sinauer.
- Thompson, W. (1840) On a new genus of fishes from India. *Charlesworth's Magazine of Natural History*, 4, 184-187.
- Udovicic, F., McFadden, G. I. & Ladiges, P.Y. (1995) Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. *Molecular Phylogenetics and Evolution*, 4, 247-256.
- Ung, V., Dubus, G., Zaragüeta Bagils, R. & Vignes-Lebbe, R. (2010) Xper² : 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. *Systematic Biology*, 43(3), 343-368.
- Williams, D.M. & Ebach, M.C. (2008) *Foundations of Systematics and Biogeography*. Springer (Eds) New York, 300 pp.
- Zaragüeta Bagils, R. & Bourdon, E. (2007) Three-item analysis: Hierarchical representation and treatment of missing and inapplicable data. *Comptes Rendus Palevol*, 6, 527-534.
- Zaragüeta Bagils, R., Ung, V., Grand, A., Vignes-Lebbe, R., Cao, N. & Ducasse, J. (2012) LisBeth: New cladistics for phylogenetics and biogeography. *Comptes Rendus Palevol*, 11(8), 563-566.

Annexe V : Comparative biogeography and Geographic Speciation of North American highland fishes from the Mesa Central to the Canadian Shield

Hoagstrom C. W., **Ung V.**, Taylor K. (2013)

(Journal of Biogeography, accepté)

Article Type: Synthesis

Title: **Miocene Rivers and Tertiary Taxon Cycles Clarify the Comparative Biogeography of North American Highland Fishes**

Author names: Christopher W. Hoagstrom^{1*}, Visotheary Ung², Kathie Taylor³

Author research addresses: ¹*Department of Zoology, Weber State University, Ogden, UT 84408, USA*, ²*CNRS UMR 7207 CR2P, 75005 Paris, France*, ³*Argenta Ecological Consultants, Salt Lake City, UT, 84171, USA*.

*Correspondence: Chris Hoagstrom, Department of Zoology, Weber State University, 2505 University Circle, Ogden, UT, 84408-2505, USA.

E-mail: christopherhoagstrom@weber.edu

Running head: Comparative biogeography of highland fishes

Word count, Abstract (line 13) through References (line 829): 9762

ABSTRACT

Aim To (1) use recently published phylogenies of six widely distributed clades of highland-type fishes in a comparative analysis that investigates relations among North American highlands, (2) construct a map of relevant (pre-historic) river geography, and (3) apply ecological paradigms to interpret patterns of highland-fish cladogenesis.

Principal questions are: (1) does highland endemism correspond to pre-historic river drainages, and (2) what ecological concept explains patterns of speciation?

Location Twenty-two North American highlands, east to west from Appalachia to the Basin and Range, north to south from the Canadian Shield to the Mesa Central.

Methods We (1) used three-item analysis to find shared highland-area relations, (2) used dated phylogenies and geological literature to construct a map of relevant, pre-historic river drainages, and (3) applied the taxon-cycle concept to interpret results.

Results Three-item analysis identified 369 most-parsimonious trees with a retention index of 80.4%. An intersection tree reconstructed from common three-area statements had a completeness index of 80.3%. Nodes on the tree identified seven major branches, of one to eight highland areas each, that were congruent with Late Miocene river drainage patterns. Sister-area relations within nodes were congruent with Plio-Pleistocene events. Older taxa have

restricted distributions and some younger members of each clade are broadly distributed, consistent with predictions of the taxon cycle.

Main conclusions Progenitors of highland endemics were widespread across an aggraded, alluvial landscape by the Early Miocene. Middle Miocene drainage rearrangements facilitated further dispersal. Accelerated erosion in the Middle Miocene, caused by tectonic uplift and climate change, created highland-type (sediment-starved) habitats. Parallel colonization and adaptation to these habitats and geographical isolation in the Late Miocene and early Pliocene created highland endemics. Plio-Pleistocene drainage rearrangements allowed tolerant (derivative) taxa to disperse and again colonize highland areas, including emerging ones.

KEYWORDS

Endemism, inland aggradation, drainage rearrangement, Miocene, parallel evolution, preglacial rivers, reciprocal illumination, taxon cycle, tectonism, three-item analysis

INTRODUCTION

Methods of Comparative Biogeography provide a means to test geological and phylogenetic hypotheses and develop new testable hypotheses that bring about "reciprocal illumination" (Rosen, 1978; Parenti & Ebach, 2009). Recent studies of North American freshwater fish distributions (Miller, 2005; Page & Burr, 2011) and phylogenetics (Table 1) along with recent studies of North American geomorphological history (Galloway *et al.*, 2011) provide unprecedented opportunity for reciprocal illumination. Here, we incorporate this information into a comparative biogeographical study and combine biogeographical and geological evidence to construct the (to our knowledge) first map of Late Miocene rivers among North American highlands. The map and accompanying ecological interpretations are presented as tools for future hypothesis testing.

Major highland areas of North America (Fig. 1a) have long been recognized as centres of aquatic biodiversity. Highland areas share the geographic qualities of locally high elevation and topographic relief. Highland streams have locally high gradients, coarse substrates (i.e., gravel, cobble, bedrock), and cool waters. The insular nature of highland-type habitats corresponds to endemism of (for example) crayfishes (Crandall & Templeton, 1999), salamanders (Highton, 1995), and fishes (Mayden, 1988). Highland-fish endemism occurs throughout North America east of the Rocky Mountains (e.g., Table 1) but no comparative analysis has yet included the entire region. Here, we focus on six recently revised lineages that are distributed widely among North American highlands (Table 1) to (1) update and broaden hypotheses of relations among highlands and (2) present a hypothesis for cladogenesis.

MATERIALS AND METHODS

We selected six clades of North American freshwater fishes (Euteleostei) for study based on presence of endemics among highlands throughout the study area and availability of recent, comprehensive phylogenetic hypotheses (Table 1) and followed Rosen's (1978) method of: (1) identifying endemic areas with endemic distributions and (2) discovering hierarchical, three-area relations to reveal sister areas (Parenti & Ebach, 2009). Disjunct highlands with endemic taxa were treated separately (Fig. 1). Sister taxa in the same highland area were grouped into a single, geographical taxon (Parenti & Ebach, 2009).

A three-area statement (e.g., if areas A, B, and C have the relationship A(BC), areas B and C are most closely related) is the smallest hierarchical statement in biogeography. Three-item analysis is the only cladistic-biogeographical method to use exclusive hierarchical hypotheses when constructing hypotheses of area relations (Cao *et al.*, 2007). It also filters missing and inapplicable data (Zaragüeta Bagils & Bourdon, 2007). In three-item analysis, three-area statements are combined to create optimal area cladograms, then an intersection tree is built from the three-area statements common to all optimal area cladograms (Cao *et al.*, 2009).

We used LisBeth (Laboratoire d'Informatique et Systématique program for Biogeography and Evolution using Three-item analysis; Zaragüeta Bagils *et al.*, 2012) to conduct analyses. It applies fractional weighting (Nelson & Ladiges, 1992), paralogy-free subtree analysis (Nelson & Ladiges, 1996), and the transparent method (Ebach *et al.*, 2005) to discover distributional congruence and resolve redundant areas (i.e., taxonomic paralogy) and widespread taxa (i.e., MASTs). The large number of areas analyzed (22) was accommodated by exporting three-area statements into a NEXUS matrix for analysis in Phylogenetic Analysis Using Parsimony software (PAUP, Swofford, 2003; Zaragüeta Bagils *et al.*, 2012). PAUP was used to find optimal taxon-area cladograms, which were then imported into LisBeth to construct an intersection tree (Zaragüeta Bagils *et al.*, 2012). Support for the validity of findings was estimated with two measures: (1) the retention index (Archie, 1989; Farris, 1989) is the percentage of three-area statements included in optimal area cladograms and (2) the completeness index (Zaragüeta Bagils *et al.*, 2012) is the percentage of three-area statements present in optimal area cladograms that could be incorporated into the intersection tree. In each case, greater inclusion indicates greater congruence among taxa.

The intersection tree was interpreted using a comprehensive review of literature on geomorphic evolution of river drainages among highland areas. Only geological evidence exclusive of previous biogeographical findings was used so that the resulting hypothesis of highland-area relations would be independent from previous biogeographical hypotheses (Parenti & Ebach, 2009). A recent review of Gulf of Mexico drainage evolution by Galloway *et al.* (2011) was instrumental, but the literature was scoured for evidence relevant to the geomorphic evolution of river drainages and highland areas. A map representing pre-historic river drainages congruent with the intersection tree and compatible with previous estimates of lineage cladogenesis was constructed to provide an upgraded hypothesis of highland evolution and highland-fish cladogenesis.

RESULTS

Ninety-one geographical taxa were included in this analysis (Appendix S1). Not all taxa of each clade are confined to highland areas, but endemism is primarily associated with highlands. Twenty-two highland areas that harbour endemic geographical taxa (Fig. 1a; Appendix S1) are distributed among major Atlantic and Gulf Coast river drainages from the Hudson River (USA) to the Ríos Nautla and Amería (México). The three-item analysis could not produce a final intersection tree until three widespread taxa were removed from the analysis. *Cyprinella venusta* "north and southwest" (as defined by Schönhuth & Mayden, 2010) occupies the Devils, Llano, Loess, Mobile, Pine, and Sabine highlands. *Cyprinella whipplei* "Mississippi tributaries" (as defined by Schönhuth & Mayden, 2010), occupies the Loess, Mobile, Moraine, Ohio, Ozark, and Tennessee highland areas. The other widespread taxon, a polytypic clade within *Oligocephalus* (*Etheostoma*) including *E. collettei*, *E. ditrema*, *E. nuchale*, and *E. swaini* (as presented in Near *et al.*, 2011), occupies the Apalachicola, Loess, Mobile, Ocala, Ouachita, and Pine highlands. Explanations for wide distributions and implications for this study are discussed below.

Three-item analysis identified 369 most-parsimonious trees with a retention index of 80.4%, which indicates substantial congruence within and among the fish clades analyzed. An intersection tree reconstructed from the three-area statements common among optimal taxon-area cladograms (Fig. 2) had a completeness index of 80.3%, demonstrating considerable congruence among most-parsimonious trees. Six major nodes identified seven sub-branches composed of one to eight highland areas each. Six sister-area relations were revealed. The Lerma highlands were excluded because only one geographical taxon was present (Appendix S1).

Geological Interpretation and map explanation

The age of highland areas varies, but major highland uplifts (e.g., Appalachia, Ozark, Ouachita, Llano, Devils, Coahuila) were part of Palaeozoic Pangea (~330-260 ma) and remained tectonically active in the Mesozoic (~240-230 ma; Hatcher, 1989; Viele & Thomas, 1989; Harry & Londono, 2004). Thereafter, tectonism diminished except in the Basin and Range (Galloway *et al.*, 2011). River-drainage geography was dynamic because of subtle, in-situ tectonic events, active tectonism to the west, and climate change (Galloway *et al.*, 2011).

Fossil and molecular evidence indicate progenitors of at least some clades in this study arose in the Eocene-Oligocene period (Near *et al.*, 2005, 2011). The Oligocene was a period of >10 ma with geomorphic stasis (relatively speaking) during which river-drainages established in the Late Eocene persisted (Galloway *et al.*, 2011). Drainage divides were presumably barriers to fish dispersal and could have promoted initial cladogenesis. Thereafter, tectonic and climatic change in the Early and Middle Miocene caused major drainage rearrangement (Galloway *et al.*, 2011). Emerging river drainages fragmented long-standing Oligocene drainages to create a new configuration that persisted for ~12 my (Galloway *et al.*, 2000, 2011). Drainage rearrangement must have facilitated passive (e.g., stream capture) and active (e.g., corridor) geographic dispersal of ancestral highland fishes. Subsequent stability created new potential for allopatric speciation.

The basal node in the intersection tree (Fig. 2, node 1) corresponds to isolation of the Mobile highlands by formation of the Tennessee and Apalachicola rivers. The present course of the Tennessee River upstream of the Coastal Plain existed by Cretaceous time (Adams, 1928; Milici, 1968; Mills & Kaye, 2001) when erosional unroofing of the Appalachians began (Boettcher & Milliken, 1994). The Eocene Tennessee River discharged to the inundated Mississippi Embayment (Grim, 1936; Milici, 1968; Self, 2000). Middle Miocene uplift of the Blue Ridge and Cumberland Plateau, a cooler, more seasonal climate, and sea-level fall accelerated sedimentation (Boettcher & Milliken, 1994) and entrenched the upper Tennessee River. Retreating sea level and delta formation allowed the Tennessee River to extend southward onto the emerging Coastal Plain (Brown, 1967; Self, 2000, 2002). It reached the Continental Shelf in the Middle Miocene and continued in that course into the Pleistocene (Galloway *et al.*, 2000, 2011; Fig. 1b). The modern Apalachicola River formed in the Middle Miocene when the Chattahoochee River (the west branch of the Apalachicola) beheaded the Tallapoosa River (Staheli, 1976; Galloway, 2008; Fig 1b). The area between the Tennessee and Apalachicola rivers, including the Black Warrior, Cahaba, Coosa, and Tallapoosa rivers of the Mobile highlands, was thus isolated.

The Black Warrior, Cahaba, Coosa, and Tallapoosa rivers maintained individual courses across the Coastal Plain (Fig. 1b) and deposited a continuous band of sediments along the coastline during the Late Miocene and Pliocene (Isphording, 1983; Otvos, 1998, 2004), but coalesced in the late Pliocene to form the Alabama River (Adams, 1929). The timing of coalescence corresponds to cessation of deposition along the coastline due to uplift, sea-level drop, and

climatic cooling (Otvos, 1998, 2004). Entrenchment of the Alabama River formed a drainage divide within Mio-Pliocene alluvial deposits (Maxwell, 1971) that separated the Mobile-Pine sister areas (Fig 2, sister relation 1). Continued coastal uplift and increased, seasonal precipitation formed new coastal drainages (e.g., Escambia, Choctawhatchee) that dissected these deposits and created the Pine highlands (Price & Whetstone, 1977; Otvos, 1981, 1998). Sustained uplift precluded development of a coastal lowland (Price & Whetstone, 1977; Otvos, 2009), so river-drainage outlets remained proximate. By this time, the coastal band of Mio-Pliocene sediments extended west, to the Mississippi Embayment (Self 1986, 1993; Meylan & Li, 1995) and low sea-stands connected adjacent river drainages (Locker & Doyle, 1992), which presumably facilitated east-west fish dispersal. The Tennessee River was eventually captured by the Ohio River (Fig. 1a) and some remnants of the ancestral Lower Tennessee River joined the Mobile Bay drainage (Mobile highlands), whereas others were incorporated into the Pine highlands.

Emergence of Appalachian drainage divides accounts for two other main branches on the intersection tree (Fig. 2, nodes 3 and 4). Geographical relief was low across the Appalachians until uplift of the Blue Ridge and Cumberland Plateau, climatic cooling, and lowered sea levels of the Middle Miocene combined to amplify geographical relief (Cleaves, 1989; Stanford, 2010; Gallen *et al.*, 2013). In response, developing river routes were rapidly entrenched (Bartholomew & Mills, 1991; Cleaves, 1993; Stanford *et al.*, 2002). The energetic advantage of Atlantic Coast streams (i.e., shorter distance to sea level) allowed them extend onto the Allegheny Plateau (Thompson, 1939, 1949). The erosion-resistant Blue Ridge became a major drainage divide in the Southern Appalachians (Hack, 1982).

Nevertheless, Atlantic Coast rivers still beheaded some interior streams (Hack, 1982; Prince *et al.*, 2010, 2011) and the Savannah River captured the Chatooga River portion of the Chattahoochee-Apalachicola River (Staheli, 1976). Thereafter, Atlantic Coast rivers draining to the Salisbury and Southeast Georgia embayments (Chesapeake and Georgia highlands, respectively) were isolated from Gulf of Mexico and Gulf of St. Lawrence drainages. Tributaries to the St. Lawrence River (e.g., Kentucky, New-Kanawha, Monongahela, Allegheny) were isolated northeast of the Cumberland Plateau, Blue Ridge, and Allegheny Plateau and flowed through the Appalachian Basin (Gray, 1991; Jacobson *et al.*, 1988; Melhorn & Kempton, 1991; Kaktins & Delano, 1999; Fig. 1b).

As with the Mobile-Pine sister areas, the Chesapeake-Georgia and Apalachicola-Ocala sister areas (Fig. 2, sister relations 2 and 3) separated relatively recently. Uplift of the Cape Fear Arch in the Pliocene (Hack, 1982; Markewich,

1985) elevated the divide between the Chesapeake and Georgia highlands. To the south, land subsidence via dissolution began, in the Late Miocene, to uplift the Ocala Highlands (Opdyke *et al.*, 1984; Adams *et al.*, 2010). Subsequent formation, by dissolution, of the Dougherty Plain and Solution Escarpment (Herrick & LeGrande, 1964) created an Apalachicola-Ocala divide (Fig. 1b). Mio-Pliocene incision of the Aucilla, Ochlockonee, and Suwanee rivers created the Ocala highlands (Otvos, 1981; Torak *et al.*, 2010).

Association of the geographically disjunct Ohio and Apalachicola-Ocala highlands (Fig. 2) suggests similar ancestral traits were preserved in disjunct refugia. Glaciation explains the lack of a proximate sister area for the Ohio highlands. Most of the St. Lawrence River drainage was repeatedly glaciated during the late Pliocene and Pleistocene, which isolated the unglaciated Ohio highlands. Adjacent areas that developed during interglacial periods and may have been colonized by Ohio-highland fishes (as they are today), but were subsequently re-glaciated.

Notably, the Ohio(Apalachicola-Ocala) relation (Fig. 2) reinforces historically debated evidence that Ohio-highland rivers were part of the pre-glacial St. Lawrence River drainage (Fig. 1b). The Kentucky and New-Kanawha rivers arose in the Middle Miocene and incision during regional uplift and climate change superimposed their courses across underlying strata (Bartholomew & Mills, 1991; Teller & Goldthwait, 1991). The Cincinnati Arch, which divides the Illinois and Appalachian basins (Sheets & Simonson, 2006), was the western boundary, except the Kentucky River was superimposed across it (Coffey, 1958; Melhorn & Kempton, 1991). Repeated glaciation formed proglacial lakes (Frolking *et al.*, 2006; Fleege *et al.*, 2011) and diverted unglaciated headwaters into the Plio-Pleistocene Teays River (Gray, 1991; Melhorn & Kempton, 1991). Later (0.78 to 1.50 ma), the Ohio River breached the Madison Divide and rapidly incised its modern course (Pease & Gomez, 1997; Anthony & Granger, 2007).

The isolated position of the Ozark highlands (Fig. 2, node 2) corresponds to Late Tertiary isolation of the Upper Mississippi River drainage. Three pre-glacial rivers converged at Cap au Grès (Horberg, 1946; Rubey, 1952; Fig. 1b): (1) the Platte-Des Moines River arose in the Northern Rocky Mountains and crossed the Central Great Plains (Witzke & Ludvigson, 1990; Wayne *et al.*, 1991), (2) the Mississippi River arose in the Superior Uplands and crossed the Illinois Basin (Anderson, 1988), with the Mahomet River as a major tributary (Kempton *et al.*, 1991; Melhorn & Kempton, 1991), and (3) the Osage, Meramec, and Gasconade rivers arose in the High Plains and Boston Mountains and crossed the Ozark Plateaus (Aber, 1985, 1997; Elfrink & Siemens, 1998). Late Tertiary subsidence in the Northern Mississippi

Embayment (Ervin & McGinnis, 1975; Cox & Van Arsdale, 2002) and uplift of the Boston Mountains (McKeown *et al.*, 1988) promoted erosion of the Ozark Plateaus and Illinois Basin (Gharrabi & Velde, 1995; Elfrink & Siemens, 1998) with increased deposition in the northwestern Mississippi Embayment (Potter, 1955; Van Arsdale *et al.*, 2007). Ample sediment supply, aridity, and stable sea levels caused aggradation of a high alluvial surface that spanned much of the embayment (Guccione *et al.*, 1990; Van Arsdale *et al.*, 2007) and reduced output to the Gulf of Mexico (Galloway *et al.*, 2011).

Aggradation in the northern Mississippi Embayment isolated the Ozark highlands (Fig. 1b). The only Appalachian river connected to the northern Mississippi Embayment was the Cumberland River (Anthony & Granger, 2007), which contributed to aggradation in the Northern Mississippi Embayment (Potter, 1955), but maintained a separate, easterly course (Fig. 1b) for hundreds of kilometres downstream (Guccione *et al.*, 1990; Saucier, 1994; Van Arsdale *et al.*, 2007). Connection to the Red River drainage (i.e., Ouachita highlands) was even more distant (Mange & Otvos, 2005; Galloway *et al.*, 2011; Fig. 1b). The lack of an Ozark highland sister area is explained by Plio-Pleistocene glaciation, which subjected the Upper Mississippi River basin to glacial ice lobes (Anderson, 1988; Aber, 1991), periglacial permafrost (Péwé, 1983; Wayne *et al.*, 1991), inundation by proglacial lakes (Todd, 1917), and meltwater floods (Aber, 1999; Kehew *et al.*, 2009).

The Ouachita(Sabine-Loess) branch (Fig. 2, node 5) reflects Early Miocene isolation of the Ouachita highlands within the Red River drainage and proximity to the southern Mississippi Embayment and adjacent Gulf Coastal Plain (Galloway *et al.*, 2011; Fig. 1b). Sedimentary deposits of the Mississippi, Red, and Sabine-Neches rivers accumulated along the Gulf Coast and in the southern Mississippi Embayment throughout the late Tertiary (Isphording & Flowers, 1983; Saucier, 1994). River courses frequently changed and merged (Saucier, 1994; Blum *et al.*, 2000), creating passive and active dispersal opportunities. Dissection of Tertiary deposits commenced in response to regional uplift and Plio-Pleistocene climate change (Self, 1986; Saucier, 1994; Otvos, 2005). The Sabine-Loess sister areas (Fig 2, sister relation IV) likely separated during Pleistocene establishment of the Red-Mississippi river valley (Galloway *et al.*, 2011), which presumably created a dispersal barrier.

The Tennessee-Moraine branch of the intersection tree (Fig. 2, node 6a, sister relation V) corresponds to Mio-Pliocene isolation of the Cumberland and Tennessee rivers (Potter, 1955; Brown, 1967; Galloway *et al.*, 2011; Fig. 1b). Delta

deposits indicate the Tennessee River remained independent into the Pleistocene (Galloway *et al.*, 2011). The nascent Ohio River joined the Cumberland River ~1.5 ma (Granger *et al.*, 2001; Anthony & Granger, 2007; Galloway *et al.*, 2011) and presumably captured the Tennessee River thereafter, which provided a direct connection, via the Wabash and Upper Mississippi rivers, to the Moraine highlands. Later, the Mississippi River was diverted east of Crowley's Ridge and, eventually, into Thebes Gap (Blum *et al.*, 2000), where it even more closely connected to the Ohio-Cumberland-Tennessee River (Fig. 1a).

The remaining eight branches composed the crown of the intersection tree, which includes all southwestern highland areas (Fig. 2, node 6b). Middle Miocene dismemberment of ancient Brazos and Rio Grande drainages created potential east-west stream connections from the Mississippi Embayment to the Tampico-Misantla Basin, but tectonism and climate change thereafter caused river-drainage fragmentation and long-term isolation (Galloway *et al.*, 2011; Fig. 1b). Uplift of the Llano and Capitan highlands (Hill, 1996; Cather *et al.*, 2012) created the Miocene Guadalupe River (Galloway *et al.*, 2011) and isolated the peripheral Cabeza de Vaca, Parras, and Tampico highlands. These disjunct areas form an unresolved (i.e., polytomic) branch of the crown group. Remaining highlands became isolated in the Late Miocene and Pliocene as tectonism and climate change fragmented the Río Bravo, Rio Grande, and Guadalupe drainages (Galloway *et al.*, 2011). Early Miocene faulting and aridity isolated the Conchos highlands in the Chihuahua Trough (King & Adkins, 1946; Haenggi, 2002; Galloway *et al.*, 2011). The Capitan and Llano highlands (a polytomic pair) were part of the Guadalupe River until Late Miocene dissolution-subsidence diverted streams of the Capitan highlands into the Delaware Basin (Maley & Huffington, 1953; Hawley, 1993a, 1993b). Isolation of the Coahuila-Devils sister highlands occurred simultaneously via faulting along the Balcones Escarpment (Kastning, 1983; Barker *et al.*, 1994). Finally, faulting in the Coahuila Folded Belt and uplift of the Tamaulipas Arch (Aranda-Gómez *et al.*, 2005; Chávez-Cabello *et al.*, 2007) separated the Coahuila-Devils sister areas (Fig. 2, sister relation VI).

Highland isolation peaked in the Late Miocene and early Pliocene (Galloway *et al.*, 2011; Fig. 1b). Subsequent climatic cooling re-connected drainage fragments to create "mega-drainages" (sensu Blum & Hattier-Womack, 2009). The Mega Mississippi River, a major conduit for glacial runoff, became the dominant Gulf of Mexico drainage (Saucier, 1994; Blum *et al.*, 2000) and connected the Ozark, Ohio, Loess, Ouachita, and Tennessee and (interglacial) Moraine highlands. The Mega Mobile Bay drainage integrated the Black Warrior, Cahaba, Coosa, and Tallapoosa rivers

(Adams, 1929). The Mega Rio Grande initially connected the Parras, Conchos, Devils, Capitan, Coahuila highlands (Strain, 1970; Hawley, 1993a; Hill, 1996; Mack *et al.*, 2006), but tectonism and aridity later disconnected the Parras highlands (Albritton, 1958). The Mega Brazos River integrated the Calcasieu, Sabine, Trinity, and Brazos rivers, but high sea levels in interglacial periods (such as the present) caused re-fragmentation (Blum & Hattier-Womack, 2009).

DISCUSSION

Methodology

Use of three-item analysis increases potential to discover new patterns (Parenti & Ebach, 2009, 2013) and, in this case, to update existing biogeographical hypotheses without relying on existing interpretations. Incorporation of hierarchical, phylogenetic hypotheses embeds evolutionary hypotheses into the analysis, such that three-item analysis reveals area relations based on evolutionary relations, not just shared traits (Williams & Ebach, 2006; Cao *et al.*, 2007). This allows more direct interplay (i.e., reciprocal illumination) between biological and geological evidence. Nevertheless, this study supports many existing biogeographical hypotheses. For instance, the Mobile highlands were basal to all others in this study and in previous ones (Mayden, 1988; Crandall & Templeton, 1999). However, inclusion of many more highland areas in this study revealed that ancient highlands are often more closely related to younger highlands of the Coastal Plain than to other ancient highlands. In particular, the Ouachita and Ozark highlands were distantly related, consistent with long-standing, pre-glacial isolation.

Subdivision of some areas used in this study would be beneficial because area relations are most informative when three-area relationships are revealed (Parenti & Ebach, 2009). For example, the Ohio and Ozark highlands each lacked sister areas, so separate analysis of drainages of the Ohio highlands (Green, Kanawha, Kentucky, Monongahela) and Ozark highlands (Gasconade, Meramec, Neosho, Osage, White) would more accurately reveal area relations. Further, evidence revealed during map construction suggests the Cumberland, Tennessee, Black Warrior, Cahaba, Coosa, and Tallapoosa rivers should be analyzed separately because each was independent during the Late Tertiary. Biological evidence supports long separation of these drainages (e.g., Starnes & Etnier, 1986; Mayden, 1987; Near & Keck, 2005; Campbell *et al.*, 2008). Given this, it would also be useful to divide the Pine highlands into separate east (ancestral Black Warrior, Cahaba, Coosa, and Tallapoosa rivers) and west (Tennessee River) sections.

A possible weakness of three-item analysis is that it provides only one hypothesis for highland-area relations when more than one pattern may be present. For example, this study identified a sister relation between the Tennessee and Moraine highlands, reflecting colonization of deglaciated regions by fishes related closely to Tennessee highland endemics. However, it is well established that fishes re-colonized glaciated regions from several refugia (e.g., Hoagstrom & Berry, 2006; Berendzen *et al.*, 2008). Thus, a comparative analysis using different clades could discover a different sister area to the Moraine highlands. A potential solution will be to conduct additional studies with different combinations of taxa until a suite of representative patterns is identified.

The goal of three-item analysis is to discover shared patterns (Parenti & Ebach, 2009, 2013), which precludes the use of a priori manipulations, such as time slicing (*sensu* Upchurch & Hunn, 2002). Hence, patterns reported here could include some reticulate area relations. This does not invalidate these relations and, in fact, multiple cycles of speciation that demonstrate similar area relations are strong evidence for a significant biogeographical pattern. In an effort to devise an accurate geobiotic scenario it is ultimately necessary to distinguish a historical sequence of biogeographical events (Morrone, 2009). The broad geographic scale of this study allowed many independent and diverse geological studies to be incorporated for interpretation of the intersection tree. The fact that major nodes correspond to geographically disjunct Middle Miocene to Pliocene events is strong evidence that this period was significant for highland-fish cladogenesis. As more rigorously dated phylogenies (e.g., Near *et al.*, 2003, 2005) become available, it will be possible to better correlate speciation events with rigorously dated geological events (*sensu* Hovenkamp, 2001; Fattorini, 2008).

Synthesis - Ecological Biogeography of highland-fish cladogenesis

It is important to connect biogeographical patterns with ecological paradigms to provide a context for hypothesis generation (Brown *et al.*, 1996). Patterns of cladogenesis in highland fishes resemble a taxon cycle in which species proceed through four stages: (1) geographic expansion, (2) population contraction, isolation, and differentiation, (3) range fragmentation and population extirpation, and (4) restriction to a single area (Ricklefs & Cox, 1972; Ricklefs & Bermingham, 1999). Stage 1 species are colonists from neighbouring regions or derived residents that can expand their distribution through biological release from predators or pathogens (i.e., a conquistador effect). Species in later stages no longer sustain biological release, are more ecologically specialized, and are increasingly confined to interior, high-elevation habitats, in part by low-elevation biological interactions (Brown *et al.*, 1996). Late-stage species persist longer in larger, more insular habitats where they are less vulnerable to stage-1 species and stochastic extirpation (MacArthur & Wilson, 1967; Cox & Ricklefs, 1972).

Evidence consistent with taxon-cycle predictions is present throughout highland-fish literature. For example, ancestral *Campostoma spadiceum* and *Lythrurus snelsoni* of the Ouachita highlands (Pramuk *et al.*, 2007; Blum *et al.*, 2008; Cashner *et al.*, 2010) are putative stage 4 species that originated in the late Tertiary when the Ouachita highlands were unroofed and Ouachita streams were isolated. These species are confined to highland-type habitats and have parapatric distributions with younger, widespread (putative stage 1) *Campostoma plumbeum* and *Lythrurus umbratilis*, respectively (Taylor & Lienesch, 1996; Cashner *et al.* 2010). *Lythrurus umbratilis* has even colonized highland-type habitats lacking *L. snelsoni* (Taylor & Lienesch, 1996). *Notropis suttkusi*, another putative stage 4 Ouachita-highland endemic, is also restricted to highland habitats, although no close relatives have colonized the area (Berendzen *et al.*, 2008).

In general, younger (Quaternary) highland taxa are more widespread than pre-glacial (Tertiary) taxa (Pflieger, 1971; Near & Benard, 2004; Roe *et al.*, 2008). Each clade in this study included widespread, putative stage 1 taxa that are not restricted to rocky, highland-type habitats (Table 2). Some, such as *Cyprinella lutrensis* (Walters *et al.*, 2008; Gido & Propst, 2012) and *Micropterus dolomieu* (Hoagstrom *et al.*, 2007b; Stefferud *et al.*, 2011), are aggressive non-native invaders, which supports the hypothesis of stage-1 biological release (Ricklefs & Bermingham, 1999). Geographical differentiation in some widespread taxa (e.g., Lang & Echelle, 2011; Husemann *et al.*, 2012) suggests they have

progressed to stage 2. Widespread, geographic taxa excluded from this analysis (see results) are likely in stage 1 or stage 2, in which case their distributions should be incongruent with stage 3 and 4 highland endemics. This could explain why their distributions were incongruent with the dominant distributional pattern.

Geological evidence suggests stage 1 progenitors of Tertiary highland endemics could have ranged widely across aggraded, North American landscapes from the Late Eocene to the Early Miocene, consistent with Pflieger (1971) and Mayden (1987, 1988) who hypothesized that an extensive, pre-glacial upland allowed expansion of highland-fish populations. River-channel avulsion and cross grading are common in aggrading areas (e.g., Phillips, 2009) and create inter-drainage connections and stream transfers that must help fishes disperse. In particular, large-scale inter-drainage connections during Middle Miocene drainage rearrangements would have promoted broad geographic spread. Thereafter, uplift and accelerated erosion created novel, sediment-starved, highland habitats that are ecologically unique (e.g., Jenkins & Freeman, 1972; Wilkinson & Edds, 2001; Walters *et al.*, 2003). Highland endemics typically require high stream gradients, low turbidity, stable substrata, and stable water chemistry (e.g., Braasch & Smith, 1965; Taylor & Lienesch, 1995, 1996; Brewer *et al.*, 2007). Molecular evidence supports Late Miocene speciation in at least some highland-fish clades (Near *et al.*, 2003, 2011; Near & Benard, 2004; Webb *et al.*, 2004; Near and Keck, 2005; Domínguez-Domínguez *et al.*, 2011). Descendants of some lowland-type fishes also speciated in highlands at the same time (e.g., Echelle *et al.*, 2005; Niemiller *et al.*, 2012). Small body size (Hardman & Hardman, 2008), discrete sexual selection aided by high water clarity (Mendelson *et al.*, 2007), and character displacement by sympatric congeners (Carlson *et al.*, 2009) potentially accelerated speciation.

Late Miocene to Pliocene resurgence of aggradation would have isolated highland specialists and prolonged isolation would have produced stage-3 species. Increasing aridity could have caused extirpation of some stage 3 populations, especially in smaller highland areas (*sensu* Hoagstrom & Berry, 2006), with remnant taxa restricted to a single area representing stage-4 species. The boundaries of highland-fish distributions often correspond to geomorphic knick points (i.e., falls and rapids) that form a transition between lowland and upland habitats (Hack, 1982). These ecological and, sometimes, physical barriers not only form the downstream boundary of highland-type habitats, but also can cause dispersal limitation that maintains depauperate faunas (Hubbell, 2001; Hoagstrom *et al.*, 2007a). But, knick points are dynamic erosional features that advance inland over time as highlands are converted to lowlands

(Frankel *et al.*, 2007). Thus, highland refuges can shrink over time, which increases susceptibility to stochastic events (MacArthur & Wilson, 1967; Ricklefs and Cox, 1972).

For example, the New River Plateau is a shrinking highland refuge. Some headwaters of the New River have been captured by the Roanoke River (Prince *et al.*, 2011) and downstream knick points are eroding headward at a geologically rapid rate (Prince & Spotila, 2013). Biogeographical evidence suggests other headwaters of the New River drainage were captured by Cumberland or Tennessee river tributaries, although a New River tributary did capture a tributary to the Monongahela River (Jenkins, 1994). In each case, the drainage with a more direct route to sea level had the advantage and, if these processes continue apace, endemics of the New River Plateau will have to survive transfer to other river drainages or invasion by non-endemics to persist. This may be an extreme example, but similar processes occur throughout the study area.

But, emerging alluvial uplands, retreating knick points, and headwater captures make vast, growing territories for stage-1 species and subsequent erosion and dissection of aggraded landscapes produces new highland areas. For example, geological evidence indicates the taxon cycle in Coastal Plain highlands must be relatively recent in response to Plio-Pleistocene evolution of highland-type landscapes (Otvos, 2009; Stults *et al.*, 2010). Late Pleistocene divergence between *Percina kathae* of the Mobile highlands and *P. austroperca* of the Pine highlands (Near & Benard, 2004) is congruent with dispersal of a stage-1 progenitor in the Late Miocene and early Pliocene followed by Plio-Pleistocene separation of these sister areas and putative sister species. Multiple sister-relations of continental highlands with adjacent Coastal Plain highlands in this study indicate geographic dispersal of highland fishes was more restricted in the Plio-Pleistocene than in the Early Miocene, consistent with predictions of the taxon cycle (Ricklefs & Cox, 1978; Ricklefs & Bermingham, 1999).

General cladogenesis among highland areas indicates parallel evolution or (more specifically) parallel taxon cycles. In the same way that anadromous fishes diversify as they colonize post-glacial landscapes (Bell & Andrews, 2007), widespread highland-type fishes apparently diversify as they colonize highland habitats. Plethodontid salamanders exhibit similar patterns (Kozak *et al.*, 2009). This model of speciation has been described as a "raceme" (Bell & Foster, 1994) or "flexible stem" (Wund *et al.*, 2008) wherein a widespread progenitor with high standing genetic diversity or high developmental plasticity gives rise to similar divergent lineages among disjunct, ecological similar areas.

However, highlands are much older than post-glacial landscapes, so highland fishes could diverge far from the progenitor or even outlive it. Like highland salamanders (Kozak & Wiens, 2006), endemic highland fishes exhibit niche conservatism, but emerging (stage 1) species continue to colonize highlands as dispersal opportunities allow, which allows endemic faunas of mixed age to accumulate over time (sensu Hoagstrom & Berry, 2006).

ACKNOWLEDGEMENTS

We thank Michael J. Blum, David A. Etnier, Phillip R. Hollingsworth, Jr., and C. Darrin Hulsey for information on fish distributions and Andrew K. Rindsberg for helpful discussion on ancestral Tennessee and Mobile Bay rivers.

REFERENCES

- Aber, J.S. (1985) Quartzite-bearing gravels and drainage development in eastern Kansas. *TER-QUA Symposium Series*, **1**, 105-110.
- Aber, J.S. (1991) The glaciation of northeastern Kansas. *Boreas*, **20**, 297-314.
- Aber, J.S. (1997) Chert gravel and neogene drainage in east-central Kansas. *Kansas Geological Survey Bulletin*, **240**, 29-41.
- Aber, J.S. (1999) Pre-Illinoian glacial geomorphology and dynamics in the central United States, west of the Mississippi. *Geological Society of America Special Paper*, **337**, 113-119.
- Adams, G.I. (1928) The course of the Tennessee River and the physiography of the southern Appalachian region. *Journal of Geology*, **36**, 481-493.
- Adams, G.I. (1929) The streams of the Coastal Plain of Alabama and the Lafayette problem. *Journal of Geology*, **37**, 193-203.
- Adams, P.N., Opdyke, N.D. & Jaeger, J.M. (2010) Isostatic uplift drive by karstification and sea-level oscillation: modelling landscape evolution in north Florida. *Geology*, **38**, 531-534.
- Albritton, C.C.Jr. (1958) Quaternary stratigraphy of the Guadiana Valley, Durango, Mexico. *Geological Society of America Bulletin*, **69**, 1197-1216.
- Anderson, R.C. (1988) Reconstruction of preglacial drainage and its diversion by earliest glacial forebulge in the upper Mississippi Valley region. *Geology*, **16**, 254-257.
- Anthony, D.M. & Granger, D.E. (2007) A new chronology for the age of Appalachian erosional surfaces determined by cosmogenic nuclides in cave sediments. *Earth Surface Processes and Landforms*, **32**, 874-887.

- Aranda-Gómez, J.J., Housh, T.B., Luhr, J.F., Henry, C.D., Becker, T. & Chávez-Cabello, G. (2005) Reactivation of the San Marcos fault during mid-to-late Tertiary extension, Chihuahua, Mexico. *Geological Society of America Special Paper*, **393**, 509-521.
- Archie, J.W. (1989) Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Systematic Zoology*, **38**, 253-269.
- Bagley, J.C., Mayden, R.L., Roe, K.J., Holznagel, W. & Harris P.M. (2011) Congeneric phylogeographical sampling reveals polyphyly and novel biodiversity within black basses (Centrarchidae: *Micropterus*). *Biological Journal of the Linnean Society*, **104**, 346-363.
- Barker, R.A., Bush, P.W. & Baker, E.T.Jr. (1994) Geologic history and hydrogeologic setting of the Edwards-Trinity aquifer system, west-central Texas. *U.S. Geological Survey Water-Resources Investigations Report*, **94-4039**, 1-51.
- Bartholomew, M.J. & Mills, H.H. (1991) Old courses of the New River: its late Cenozoic migration and bedrock control inferred from high-level stream gravels, southwestern Virginia. *Geological Society of America Bulletin*, **103**, 73-81.
- Bell, M.A. & Andrews C.A. (1997) Evolutionary consequences of postglacial colonization of fresh water by primitively anadromous fishes. *Evolutionary ecology of freshwater animals* (ed. by B. Streit, T Städler, & C.M. Lively), pp. 323-363. Birkhäuser Verlag, Basel, Switzerland.
- Bell, M.A. & Foster, S.A. (1994) Introduction to the evolutionary biology of the threespine stickleback. *The evolutionary biology of the threespine stickleback* (ed. by M.A. Bell & S.A. Foster), pp. 1-27. Oxford University Press, New York.
- Berendzen, P.B., Simons, A.M., Wood, R.M., Dowling, T.E. & Secor, C.L. (2008) Recovering cryptic diversity and ancient drainage patterns in eastern North America: historical biogeography of the *Notropis rubellus* species group (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, **46**, 721-737.
- Blum, M.D. & Hattier-Womack, J. (2009) Climate change, sea-level change, and fluvial sediment supply to deepwater depositional systems. *SEPM Special Publication*, **92**, 15-39.
- Blum, M.D., Guccione, M.J., Wysocki, D.A., Robnett, P.C. & Rutledge, E.M. (2000) Late Pleistocene evolution of the lower Mississippi River valley, southern Missouri to Arkansas. *Geological Society of America Bulletin*, **112**, 221-235.

- Blum, M.J., Neely, D.A., Harris, P.M. & R.L. Mayden (2008) Molecular systematics of the cyprinid genus *Campostoma* (Actinopterygii: Cypriniformes): disassociation between morphological and mitochondrial differentiation. *Copeia*, **2008**, 360-369.
- Boettcher, S.S. & Milliken, K.L. (1994) Mesozoic-Cenozoic unroofing of the southern Appalachian Basin: apatite fission track evidence from middle Pennsylvanian sandstones. *Journal of Geology*, **102**, 655-668.
- Braasch, M.E. & Smith, P.W. (1965) Relationships of the topminnows *Fundulus notatus* and *Fundulus olivaceus* in the upper Mississippi River valley. *Copeia*, **1965**, 46-53.
- Brewer, S.K., Rabeni, C.F., Sowa, S.P. & Annis G. (2007) Natural landscape and stream segment attributes influencing the distribution and relative abundance of riverine smallmouth bass in Missouri. *North American Journal of Fisheries Management*, **27**, 326-341.
- Brown, B.W. (1967) A Pliocene Tennessee River hypothesis for Mississippi. *Southeastern Geology*, **8**, 81-84.
- Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597-623.
- Cao, N., Zaragüeta Bagils, R. & Vignes-Lebbe, R. (2007) Hierarchical representation of hypotheses of homology. *Geodiversitas*, **29**, 5-15.
- Carlson, R.L., Wainwright, P.C. & Near, T.J. (2009) Relationship between species co-occurrence and rate of morphological change in *Percina* darters (Percidae: Etheostomatinae). *Evolution*, **63**, 767-778.
- Campbell, D.C., Johnson, P.D., Williams, J.D., Rindsberg, A.K., Serb, J.M., Small, K.K. & Lydeard, C. (2008) Identification of 'extinct' freshwater mussel species using DNA barcoding. *Molecular Ecology Resources*, **8**, 711-724.
- Cashner, R.C., Matthews, W.J., Marsh-Matthews, E., Unmack, P.J. & Cashner, F.M. (2010) Recognition and redescription of a distinctive stoneroller form from the southern interior highlands. *Copeia*, **2010**, 300-311.
- Cather, S.M., Chapin, C.E., & Kelley, S.A. (2012) Diachronous episodes of Cenozoic erosion in southwestern North America and their relationship to surface uplift, paleoclimate, paleodrainage, and paleoaltimetry. *Geosphere*, **8**, 1177-1206.

- Chávez-Cabello, G., Aranda-Gómez, J.J., Molina-Garza, R.S., Cossío-Torres, T., Arvizu-Gutiérrez, I.R. & González-Naranjo, G.A. (2007) The San Marco fault: a Jurassic multireactivated basement structure in northeastern México. *Geological Society of America Special Paper*, **422**, 261-286.
- Cleaves, E.T. (1989) Appalachian Piedmont landscapes from the Permian to the Holocene. *Geomorphology*, **2**, 159-179.
- Cleaves, E.T. (1993) Climatic impact on isovolumetric weathering of a coarse-grained schist in the northern Piedmont Province of the central Atlantic states. *Geomorphology*, **8**, 191-198.
- Coffey, G.N. (1958) Major glacial drainage changes in Ohio. *Ohio Journal of Science*, **58**, 43-49.
- Cox, R.T. & Van Arsdale, R.B. (2002) The Mississippi Embayment, North America: a first order continental structure generated by the Cretaceous superplume mantle event. *Journal of Geodynamics*, **34**, 163-176.
- Crandall, K.A. & Templeton A.R. (1999) The zoogeography and centers of origin of the crayfish subgenus *Procericambarus* (Decapoda: Cambaridae). *Evolution*, **53**, 123-134.
- Domínguez-Domínguez, O., Vila, M., Pérez-Rodríguez, R., Remón, N. & Doadrio, I. (2011) Complex evolutionary history of the Mexican stoneroller *Campostoma ornatum* Girard, 1856 (Actinopterygii: Cyprinidae). *BMC Evolutionary Biology*, **11**, 153.
- Doosey, M.H., Bart, H.L.Jr., Saitoh, K. & Miya, M. (2010) Phylogenetic relationships of catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution*, **54**, 1028-1034.
- Ebach, M.C., Humphries, C.J., Newman, R.A., Williams, D.M. & Walsh, S.A. (2005) Assumption 2: opaque or intuition? *Journal of Biogeography*, **32**, 781-787.
- Echelle, A.A., Carson, E.W., Echelle, A.F., Van Den Bussche, R.A., Dowling, T.A. & Meyer, A. (2005) Historical biogeography of the new-world pupfish genus *Cyprinodon* (Teleostei: Cyprinodontidae). *Copeia*, **2005**, 320-339.
- Elfrink, N.M. & Siemens, M.A. (1998) Quaternary drainage shifts in Missouri. *Guidebook for the 45th Annual Field Trip*, pp. 48-69. Association of Missouri Geologists, Farmington, Missouri.
- Ervin, C.P. & McGinnis, L.D. (1975) Reelfoot Rift: reactivated precursor to the Mississippi Embayment. *Geological Society of America Bulletin*, **86**, 1287-1295.

- Farris, J.S. (1989) The retention index and the rescaled consistency index. *Cladistics*, **5**, 417-419.
- Fattorini, S. (2008) Hovenkamp's ostracized vicariance analysis: testing new methods of historical biogeography. *Cladistics*, **24**, 611-622.
- Fleeger, G.M., Grote, T., Straffin, E. & Szabo, J.P. (2011) Quaternary geology of northwestern Pennsylvania. *Geological Society of America Field Guide*, **20**, 87-109.
- Frankel, K.L., Pazzaglia, F.J. & Vaughn, J.D. (2007) Knickpoint evolution in a vertically bedded substrate, upstream-dipping terraces and Atlantic slope bedrock channels. *Geological Society of America Bulletin*, **119**, 476-486.
- Frolking, T.A & Pachell, M.A. (2006) Glacial Lake Licking: late-glacial drainage diversion and the formation of Black Hand Gorge, Licking County, Ohio. *Ohio Journal of Science*, **106**, 103-111.
- Gallen, S.F., Wegmann, K.W. & Bohnenstiehl, D.R. (2013) Miocene rejuvenation of topographic relief in the southern Appalachians. *GSA Today*, **23**, 4-10.
- Galloway, W.E. (2008) Depositional evolution of the Gulf of Mexico sedimentary basin. *Sedimentary Basins of the World*, **5**, 505-549.
- Galloway, W.E., Whiteaker, T.L. & Ganey-Curry, P. (2011) History of Cenozoic North American drainage basin evolution, sediment yield, and accumulation in the Gulf of Mexico basin. *Geosphere*, **7**, 938-973.
- Galloway, W.E., Ganey-Curry, P.E., Li, X. & Buffler, R.T. (2000) Cenozoic depositional history of the Gulf of Mexico basin. *AAPG Bulletin*, **84**, 1743-1774.
- Gharrabi, M. & Velde, B. (1995) Clay mineral evolution in the Illinois Basin and its causes. *Clay Minerals*, **30**, 353-364.
- Gido, K.B. & Propst, D.L. (2012) Long-term dynamics of native and nonnative fishes in the San Juan River, New Mexico and Utah, under a partially managed flow regime. *Transactions of the American Fisheries Society*, **141**, 645-659.
- Granger, D.E., Fabel, D. & Palmer, A.N. (2001) Pliocene-Pleistocene incision of the Green River, Kentucky, determined from radioactive decay of cosmogenic ^{26}Al and ^{10}Be in Mammoth Cave sediments. *Geological Society of America Bulletin*, **113**, 825-836.
- Gray, H.H. (1991) Origin and history of the Teays drainage system; the view from midstream. *Geological Society of America Special Paper*, **259**, 43-50.
- Grim, R.E. (1936) The Eocene sediments of Mississippi. *Mississippi Geological Survey Bulletin*, **30**, 1-240.

- Guccione, M.J., Prior, W.L. & Rutledge, E.M. (1990) The Tertiary and early Quaternary geology of Crowley's Ridge. *Field Guide to the Mississippi Alluvial Valley, Northeast Arkansas and Southeast Missouri* (ed. by MJ Guccione and EM Rutledge), pp. 23-44. Friends of the Pleistocene, South-Central Cell.
- Hack, J.T. (1982) Physiographic divisions and differential uplift in the Piedmont and Blue Ridge. *U.S. Geological Survey Professional Paper*, **1265**, 1-49.
- Haenggi, W.T. (2002) Tectonic history of the Chihuahua trough, Mexico and adjacent USA, Part II: Mesozoic and Cenozoic. *Boletín de la Sociedad Geológica Mexicana*, **55**, 38-94.
- Hardman, M. & Hardman L.M. (2008) The relative importance of body size and paleoclimatic change as explanatory variables influencing lineage diversification rate: an evolutionary analysis of bullhead catfishes (Siluriformes: Ictaluridae). *Systematic Biology*, **57**, 116-130.
- Harry, D.L. & Londono, J. (2004) Structure and evolution of the central Gulf of Mexico continental margin and coastal plain, southeast United States. *Geological Society of America Bulletin*, **116**, 188-199.
- Hatcher, R.D.Jr. (1989) Tectonic synthesis of the U.S. Appalachians. *The Appalachian-Ouachita orogen in the United States* (ed. by R.D. Hatcher, Jr., W.A. Thomas, & G.W. Viele), pp. 511-535. Geological Society of America, Boulder, Colorado.
- Hawley, J.W. (1993a) Overview of the geomorphic history of the Carlsbad area. *New Mexico Geological Society, Fall Field Conference Guidebook*, **44**, 2-3.
- Hawley, J.W. (1993b) The Ogallala and Gatuña formations in the southeastern New Mexico region, a progress report. *New Mexico Geological Society, Fall Field Conference Guidebook*, **44**, 261-269.
- Herrick, S.M. & LeGrand, H.E. (1964) Solution subsidence of a limestone terrane in southwest Georgia. *International Association of Scientific Hydrology Bulletin*, **9**, 25-36.
- Highton, R. (1995) Speciation in eastern North American salamanders of the genus *Plethodon*. *Annual Review of Ecology and Systematics*, **26**, 579-600.
- Hill, C.A. (1996) Geology of the Delaware Basin, Guadalupe, Apache, and Glass Mountains, New Mexico and west Texas. *Permian Basin Section-SEPM Publication*, **96-39**, 1-480.

- Hoagstrom, C.W. & Berry, C.R.Jr. (2006) Island biogeography of native fish faunas among Great Plains drainage basins: basin scale features influence composition. *American Fisheries Society Symposium*, **48**, 221-264.
- Hoagstrom, C.W., Gosch, N.J.C., DeWitte, A.C., Berry, C.R.Jr., & Duehr, J.P. (2007a) Biodiversity, Biogeography, and longitudinal fish faunal structure among perennial, warmwater streams of the Cheyenne River drainage. *Prairie Naturalist*, **39**, 117-144.
- Hoagstrom, C.W., DeWitte, A.C., Gosch, N.J.C. & Berry C.R. Jr. (2007b) Historical fish assemblage flux in the Cheyenne River below Angostura Dam. *Journal of Freshwater Ecology*, **22**, 219-229.
- Horberg, L. (1946) Preglacial erosion surfaces in Illinois. *Journal of Geology*, **54**, 179-192.
- Hovenkamp, P. (2001) A direct method for the analysis of vicariance patterns. *Cladistics*, **17**, 260-265.
- Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
- Husemann, M., Ray, J.W., King, R.S., Hooser, E.A. & Danley, P.D. (2012) Comparative biogeography reveals differences in population genetic structure of five species of stream fishes. *Biological Journal of the Linnean Society*, **107**, 867-885.
- Isphording, W.C. (1983) Interpretive mineralogy: examples from Miocene coastal plain sediments. *Gulf Coast Association of Geological Societies Transactions*, **33**, 295-305.
- Isphording, W.C. & Flowers, G.C. (1983) Differentiation of unfossiliferous clastic sediments: solutions from the southern portion of the Alabama-Mississippi Coastal Plain. *Tulane Studies in Geology and Paleontology*, **17**, 59-83.
- Jacobson, R.B., Elston, D.P. & Heaton, J.W. (1988) Stratigraphy and magnetic polarity of the high terrace remnants in the upper Ohio and Monongahela rivers in West Virginia, Pennsylvania, and Ohio. *Quaternary Research*, **29**, 216-232.
- Jenkins, R.E. (1980) *Moxostoma duquesnii* (Lesueur), black redhorse. Atlas of North American Freshwater Fishes (ed. by D.S. Lee, C.R. Gilbert, C.H. Hocutt, R.E. Jenkins, D.E. McAllister & J.R. Stauffer, Jr.), pp. 419-420. North Carolina State University Museum of Natural History, Raleigh.
- Jenkins, R.E. (1994) Biogeography. *Freshwater Fishes of Virginia* (R.E. Jenkins & N.M. Burkhead), pp. 43-99. American Fisheries Society, Bethesda, Maryland.
- Jenkins, R.E. & Freeman, C.A. (1972) Longitudinal distribution and habitat of fishes of Mason Creek, and upper Roanoke River drainage tributary, Virginia. *Virginia Journal of Science*, **23**, 194-202.

- Kaktins, U. & Delano, H.L. (1999) Drainage basins. *The geology of Pennsylvania* (ed. by A.P. Schultz), pp. 379-390. Pennsylvania Geological Survey and Pittsburgh Geological Society, Harrisburg.
- Kastning, E.H. (1983) Relict caves as evidence of landscape and aquifer evolution in a deeply dissected carbonate terrain: southwest Edwards Plateau, Texas, U.S.A. *Journal of Hydrology*, **61**, 89-112.
- Kehew, A.E., Lord, M.L., Kozlowski, A.L. & Fisher T.G. (2009) Proglacial megaflooding along the margins of the Laurentide ice sheet. *Megaflooding on Earth and Mars* (ed. by D.M. Burr, P.A. Carling & V.R. Baker), pp. 104-127.
- Kempton, J.P., Johnson, W.H., Heigold, P.C. & Cartwright, K. (1991) Mahomet Bedrock Valley in east-central Illinois; topography, glacial drift stratigraphy, and hydrogeology. *Geological Society of America Special Paper* **258**, 91-124.
- King, R.E. & Adkins, W.S. (1946) Geology of a part of the lower Conchos Valley, Chihuahua, Mexico. *Geological Society of America Bulletin*, **57**, 275-294.
- Kozak, K.H. & Wiens, J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, **60**, 2604-2621.
- Kozak, K.H., Mendyk, R.W. & Wiens, J.J. (2009) Can parallel diversification occur in sympatry? Repeated patterns of body-size evolution in coexisting clades of North American salamanders. *Evolution*, **63**, 1769-1784.
- Lang, N.J. & Echelle, A.A. (2011) Novel phylogeographic patterns in a lowland fish, *Etheostoma proeliare* (Percidae). *Southeastern Naturalist*, **10**, 133-144.
- Locker, S.D. & Doyle, L.J. (1992) Neogene to recent stratigraphy and depositional regimes of the northwest Florida inner continental shelf. *Marine Geology*, **104**, 123-138.
- MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.
- Mack, G.H., Seager, W.R., Leeder, M.R., Perez-Arlucea, M. & Salyards, S.L. (2006) Pliocene and Quaternary history of the Rio Grande, the axial river of the southern Rio Grande rift, New Mexico, USA. *Earth-Science Reviews*, **79**, 141-162.
- Maley, V.C. & Huffington, R.M. (1953) Cenozoic fill and evaporate solution in the Delaware Basin, Texas and New Mexico. *Geological Society of America Bulletin*, **64**, 539-546.
- Mange, M.A. & Otvos, E.G. (2005) Gulf coastal plain evolution in west Louisiana: heavy mineral provenance and Pleistocene alluvial chronology. *Sedimentary Geology*, **182**, 29-57.

- Markewich, H.W. (1985) Geomorphic evidence for Pliocene-Pleistocene uplift in the area of the Cape Fear Arch, North Carolina. *Tectonic Geomorphology* (ed. by M. Morisawa & J.T. Hack), pp. 279-297. Unwin Hyman, Winchester, Massachusetts.
- Maxwell, R.W.Jr., Origin and chronology of Alabama River terraces. *Gulf Coast Association of Geological Societies Transactions*, **21**, 83-95.
- Mayden, R.L. (1987) Pleistocene glaciation and historical biogeography of North American central-highland fishes. *Kansas Geological Survey Guidebook Series*, **5**, 141-151.
- Mayden, R.L. (1988) Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Zoology*, **37**, 329-355.
- McKeown, F.A., Jones-Cecil, M., Askew, B.L. & McGrath, M.B. (1988) Analysis of stream-profile data and inferred tectonic activity, eastern Ozark Mountains region. *U.S. Geological Survey Bulletin*, **1807**, 1-39.
- Melhorn, W.N. & Kempton, J.P. (1991) The Teays system; a summary. *Geological Society of America Special Paper*, **258**, 125-128.
- Mendelson, T.C., Imhoff, V.E. & Venditti, J.J. (2007) The accumulation of reproductive barriers during speciation: postmating barriers in two behaviorally isolated species of darters (Percidae: Etheostoma). *Evolution*, **61**, 2596-2606.
- Meylan, M.A. & Li, Z. (1995) Geologic mapping of south-central Mississippi: a model for the distribution of Neogene and Quaternary sediments. *Gulf Coast Association of Geological Societies Transactions*, **45**, 435-440.
- Milici, R.C. (1968) Mesozoic and Cenozoic physiographic development of the lower Tennessee River: in terms of the dynamic equilibrium concept. *Journal of Geology*, **76**, 472-479.
- Miller, R.R. (2005) *Freshwater fishes of México*. University of Chicago Press, Chicago.
- Mills, H.H. & Kaye, J.M. (2001) Drainage history of the Tennessee River: review and new metamorphic quartz gravel locations. *Southeastern Geology*, **40**, 75-97.
- Morrone, J.J. (2009) *Evolutionary biogeography*. Columbia University Press, New York.
- Near, T.J. & Benard, M.F. (2004) Rapid allopatric speciation in logperch darters (Percidae: Percina). *Evolution*, **58**, 2798-2808.

- Near, T.J. & Keck, B.P. (2005) Dispersal, vicariance, and timing of diversification in *Nothonotus* darters. *Molecular Ecology*, **14**, 3485-3496.
- Near, T.J., Bolnick, D.I. & Wainwright, P.C. (2005) Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*, **59**, 1768-1782.
- Near, T.J., Kassler, T.W., Koppelman, J.B., Dillman, C.B. & Phillipp, D.P. (2003) Speciation in North American black basses, *Micropterus* (Actinopterygii: Centrarchidae). *Evolution*, **57**, 1610-1621.
- Near, T.J., Bossu, C.M., Bradburd, G.S., Carlson, R.L., Harrington, R.C., Hollingsworth, P.R.Jr., Keck, B.P. & Etnier, D.A. (2011) Phylogeny and temporal diversification of darters (Percidae: Etheostomatinae). *Systematic Biology*, **60**, 565-595.
- Nelson, G. & Ladiges, P.Y. (1992) Information content and fractional weight of three-item statements. *Systematic Biology*, **41**, 490-494.
- Nelson, G. & Ladiges, P.Y. (1996) Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates*, **3167**, 1-58.
- Niemiller, M.L., Fitzpatrick, B.M., Shah, P., Schmitz, L. & Near, T.J. (2012) Evidence for repeated loss of selective constraint in rhodopsin of amblyopsid cavefishes (Teleostei: Amblyopsidae). *Evolution*, **67**, 732-748.
- Opdyke, N.D., Spangler, D.P., Smith, D.L., Jones, D.S. & Lindquist, R.C. (1984) Origin of the epeirogenic uplift of Pliocene-Pleistocene beach ridges in Florida and development of the Florida karst. *Geology*, **12**, 226-228.
- Otvos, E.G. (1981) Tectonic lineaments of Pliocene and Quaternary shorelines, northeast Gulf Coast. *Geology*, **9**, 398-404.
- Otvos, E.G. (1998) Citronelle Formation, Northeastern Gulf Coastal Plain: Pliocene stratigraphic framework and age issues. *Gulf Coast Association of Geological Societies Transactions*, **48**, 321-333.
- Otvos, E.G. (2005) Numerical chronology of Pleistocene coastal plain and valley development; extensive aggradation during glacial low sea-levels. *Quaternary International*, **135**, 91-113.
- Otvos, E.G. (2009) Pleistocene Gulf Coastal Plain: regional stratigraphic and geomorphic contrasts. Intensive erosion related to sea-level decline and hinterland uplift. *Gulf Coast Association of Geological Societies Transactions*, **59**, 585-600.

- Page, L.M. & Burr, B.M. (2011) *Peterson field guide to freshwater fishes of North America north of Mexico*, 2nd edition. Houghton Mifflin Harcourt, Boston.
- Parenti, L.R. & Ebach, M.C. (2009) *Comparative biogeography*. University of California Press, Berkeley.
- Parenti, L.R. & Ebach, M.C. (2013) Evidence and hypothesis in biogeography. *Journal of Biogeography*, **40**, 8013-820.
- Pease, P.P. & Gomez, B. (1997) Landscape development as indicated by basin morphology and the magnetic polarity of cave sediments, Crawford Upland, south-central Indiana. *American Journal of Science*, **297**, 842-858.
- Pflieger, W.L. (1971) A distributional study of Missouri fishes. *University of Kansas Publications, Museum of Natural History*, **20**, 225-570.
- Phillips, J.D. (2009) Avulsion regimes in southeast Texas rivers. *Earth Surface Processes and Landforms*, **34**, 75-87.
- Potter, P.E. (1955) The petrology and origin of the Lafayette Gravel part 2. Geomorphic History. *Journal of Geology*, **63**, 115-132.
- Pramuk, J.B., Grose, M.J., Clarke, A.L., Greenbaum, E., Bonaccorso, E., Guayasamin, J.M., Smith-Pardo, A.H., Benz, B.W., Harris, B.R., Siegfried, E., Reid, Y.R., Holcroft-Benson, N. & Wiley, E.O. (2007) Phylogeny of finescale shiners of the genus *Lythrurus* (Cypriniformes: Cyprinidae) inferred from four mitochondrial genes. *Molecular Phylogenetics and Evolution*, **42**, 287-297.
- Price, R.C. & Whetstone, K.N. (1977) Lateral stream migration as evidence for regional geologic structures in the eastern Gulf Coastal Plain. *Southeastern Geology*, **18**, 129-147.
- Prince, P.S. & Spotila J.A. (2013) Evidence of transient topographic disequilibrium in a landward passive margin river system: knickpoints and paleo-landscapes of the New River basin, southern Appalachians. *Earth Surface Processes and Landforms*, in press.
- Prince, P.S., Spotila, J.A. & Henika, W.S. (2011) Stream capture as driver of transient landscape evolution in a tectonically quiescent setting. *Geology*, **39**, 823-826.
- Ricklefs, R.E. & Bermingham, E. (1999) Taxon cycles in the Lesser Antillean avifauna. *Ostrich*, **70**, 49-59.
- Ricklefs, R.E. & Cox, G.W. (1972) Taxon cycles in the West Indian avifauna. *American Naturalist*, **106**, 195-219.
- Ricklefs, R.E. & Cox, G.W. (1978) Stage of taxon cycle, habitat distribution, and population density in the avifauna of the west indies. *American Naturalist*, **112**, 875-895.

- Roe, K.J., Mayden, R.L. & Harris, P.M. (2008) Systematics and zoogeography of the Rock Basses (Centrarchidae: Ambloplites). *Copeia*, **2008**, 858-867.
- Rosen, D.E. (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*, **27**, 159-188.
- Rubey, W.W. (1952) Geology and mineral resources of the Hardin and Brussels Quadrangles (in Illinois). U.S. Geological Survey Professional Paper, **218**, 1-179.
- Saucier, R.T. (1994) *Geomorphology and quaternary geologic history of the lower Mississippi Valley*. U.S. Army Corps of Engineers, Mississippi River Commission, Vicksburg.
- Schönhuth, S. & Mayden, R.L. (2010) Phylogenetic relationships in the genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **55**, 77-98.
- Schönhuth, S., Blum, M.J., Lozano-Vilano, L., Neely, D.A., Varela-Romero, A., Espinosa, H., Perdices, A. & Mayden, R.L. (2011) Inter-basin exchange and repeated headwater capture across the Sierra Madre Occidental inferred from the phylogeography of Mexican stonerollers. *Journal of Biogeography*, **38**, 1406-1421.
- Self, R.P. (1986) Depositional environments and gravel distribution in the Plio-Pleistocene Citronelle Formation of southeastern Louisiana. *Gulf Coast Association of Geological Societies Transactions*, **36**, 561-573.
- Self, R.P. (2000) The pre-Pliocene course of the lower Tennessee River as deduced from river terrace gravels in southwest Tennessee. *Southeastern Geology*, **39**, 61-70.
- Self, R.P. (2002) Petrology of Cretaceous to Quaternary gravels from the western valley of the Tennessee River. *Journal of the Tennessee Academy of Science*, **77**, 52-56.
- Sheets, R.A. & Simonson, L.A. (2006) Compilation of regional ground-water divides for principal aquifers corresponding to the Great Lakes Basin, United States. *U.S. Geological Survey Scientific Investigations Report*, **2006-5102**, 1-23.
- Staheli, A.C. (1976) Topographic expression of superimposed drainage on the Georgia Piedmont. *Geological Society of America Bulletin*, **87**, 450-452.
- Stanford, S.D. (2010) Onshore record of Hudson River drainage to the continental shelf from the late Miocene through the late Wisconsinan deglaciation, USA: synthesis and revision. *Boreas*, **39**, 1-17.

- Stanford, S.D., Ashley, G.M., Russell, E.W.B. & Brenner, G.J. (2002) Rates and patterns of late Cenozoic denudation in the northernmost Atlantic Coastal Plain and Piedmont. *Geological Society of America Bulletin*, **114**, 1422-1437.
- Starnes, W.C. & Etnier, D.A. (1986) Drainage evolution and fish biogeography of the Tennessee and Cumberland rivers drainage realm. *The zoogeography of North American freshwater fishes* (ed. by C.H. Hocutt & E.O. Wiley), pp. 325-361. John Wiley & Sons, New York.
- Stefferdud, J.A., Gido, K.B. & Propst, D.L. (2011) Spatially variable response of native fish assemblages to discharge, predators, and habitat characteristics in an arid-land river. *Freshwater Biology*, **56**, 1403-1416.
- Strain, W.S. (1970) Late Cenozoic bolson integration in the Chihuahua tectonic belt. *The geologic framework of the Chihuahua tectonic belt* (ed. by K. Seewald & D. Sundeen), pp. 56-59. West Texas Geological Society & University of Texas at Austin.
- Stults, D.Z., Axsmith, B.J. & Yu-Sheng, L. (2010) Evidence of white pine (*Pinus* subgenus *Strobus*) dominance from the Pliocene northeastern Gulf of Mexico Coastal Plain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **287**, 95-100.
- Swofford, D. L. 2003. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Taylor, C.M. & Lienesch, P.W. (1995) Environmental correlates of distribution and abundance for *Lythrurus snelsoni*: a range-wide analysis of an endemic fish species. *Southwestern Naturalist*, **40**, 373-378.
- Taylor, C.M. & Lienesch, P.W. (1996) Regional parapatry of the congeneric cyprinids *Lythrurus snelsoni* and *L. umbratilis*: species replacement along a complex environmental gradient. *Copeia*, **1996**, 493-497.
- Teller, J.T. & Goldthwait, R.P. (1991) The old Kentucky River; a major tributary to the Teays River. *Geological Society of America Special Paper*, **258**, 29-41.
- Thompson, H.D. (1939) Drainage evolution in the southern Appalachians. *Bulletin of the Geological Society of America*, **50**, 1323-1356.
- Thompson, H.D. (1949) Drainage evolution in the Appalachians of Pennsylvania. *Annals of the New York Academy of Sciences*, **52**, 33-62.
- Todd, J.E. (1917) History of Kaw Lake. *Transactions of the Kansas Academy of Science*, **28**, 187-199.

- Torak, L.J., Painter, J.A. & Peck, M.F. (2010) Geohydrology of the Aucilla-Suwannee-Ochlockonee river basin, south-central Georgia and adjacent parts of Florida. *U.S. Geological Survey Scientific Investigations Report*, **2010-5072**, 1-78.
- Upchurch, P. & Hunn, C.A. (2002) "Time": the neglected dimension in cladistic biogeography? *Geobios, Mémoire spécial* 24, 277-286.
- Van Arsdale, R., Bresnahan, R., McCallister, N. & Waldron, B. (2007) Upland complex of the central Mississippi River valley: its origin, denudation, and possible role in reactivation of the New Madrid seismic zone. *Geological Society of America Special Paper*, **425**, 177-192.
- Viele, G.W. & Thomas, W.A. (1989) Tectonic synthesis of the Ouachita orogenic belt. *The Appalachian-Ouachita orogen in the United States* (ed. by R.D. Hatcher, Jr., W.A. Thomas, & G.W. Viele), pp. 695-728. Geological Society of America, Boulder, Colorado.
- Walters, D.M., Leigh, D.S., Freeman, M.C., Freeman, B.J. & Pringle, C.M. (2003) Geomorphology and fish assemblages in a Piedmont river basin, U.S.A. *Freshwater Biology*, **48**, 1950-1970.
- Walters, D.M., Blum, M.J., Rashleigh, B., Freeman, B.J., Porter, B.A. & Burkhead, N.M. (2008) Red shiner invasion and hybridization with blacktail shiner in the upper Coosa River, USA. *Biological Invasions*, **10**, 1229-1242.
- Wayne, W.J., Aber, J.S., Agard, S.S., Bergantino, R.N., Bluemle, J.P., Coates, D.A., Cooley, M.E., Madole, R.F., Martin, J.E., Mears, B.Jr., Morrison, R.B. & Sutherland, W.M. (1991) Quaternary geology of the northern Great Plains. *Quaternary non-glacial geology* (ed. by R.B. Morrison), pp. 441-476. Geological Society of America, Boulder, Colorado.
- Webb, S.A., Graves, J.A., Macias-Garcia, C., Magurran, A.E., Fioghil, D.Ó. & Ritchie, M.G. (2004) Molecular phylogeny of the livebearing Goodeidae (Cyprinodontiformes). *Molecular Phylogenetics and Evolution*, **30**, 527-544.
- Wilkinson, C.D. & Edds, D.R. (2001) Spatial pattern and environmental correlates of a midwestern stream fish community: including spatial autocorrelation as a factor in community analysis. *American Midland Naturalist*, **146**, 271-289.
- Williams, D.M. & Ebach, M.C. (2006) The data matrix. *Geodiversitas*, **28**, 409-420.
- Witzke, B.J. & Ludvigson, G.A. (1990) Petrographic and stratigraphic comparisons of sub-till and inter-till alluvial units in western Iowa: implications for development of the Missouri River drainage. *Midwest Friends of the Pleistocene Field Conference*, **36**, 119-143.

Wund, M.A., Baker, J.A., Clancy, B., Golub, J.L. & Foster, S.A. (2008) A test of the "flexible stem" model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. *American Naturalist*, **172**, 449-462.

Zaragüeta-Bagils, R. & Bourdon, E. (2007) Three-item analysis: hierarchical representation and treatment of missing and inapplicable data. *Comptes Rendus Palevol*, **6**, 527-534.

Zaragüeta Bagils, R., Ung, V., Grand, A., Vignes-Lebbe, R., Cao, N. & Ducasse, J. (2012) LisBeth: new cladistics for phylogenetics and biogeography. *Comptes Rendus Palevol*, **11**, 563-566.

BIOSKETCHES

Chris Hoagstrom, an Associate Professor, is interested in freshwater fish ecology and conservation at all spatiotemporal scales. He is particularly interested in fish community assembly and how fish assemblages and habitat conditions interact.

Visotheary Ung, a PhD student, is working on Southeast Asia biogeography. She is interested in methodology and specialy cladistics approach to historical biogeography.

Kathie Taylor...

Author contributions: C.W.H. conceived the ideas and gathered phylogenetic, distributional, and geological evidence; V.R.U. developed and conducted the three-item analysis; K.T. developed and produced the maps; C.W.H. led the writing; all authors provided comments and feedback and approved the final version of the manuscript

Table 1 Clades of North American fishes used for analysis. Geographical taxa are monophyletic species groups occupying a single highland area (Parenti & Ebach, 2009). References provide phylogenetic and distributional information used in this study. See Appendix S1 for more detail and Miller (2005) and Page & Burr (2011) for more distributional information.

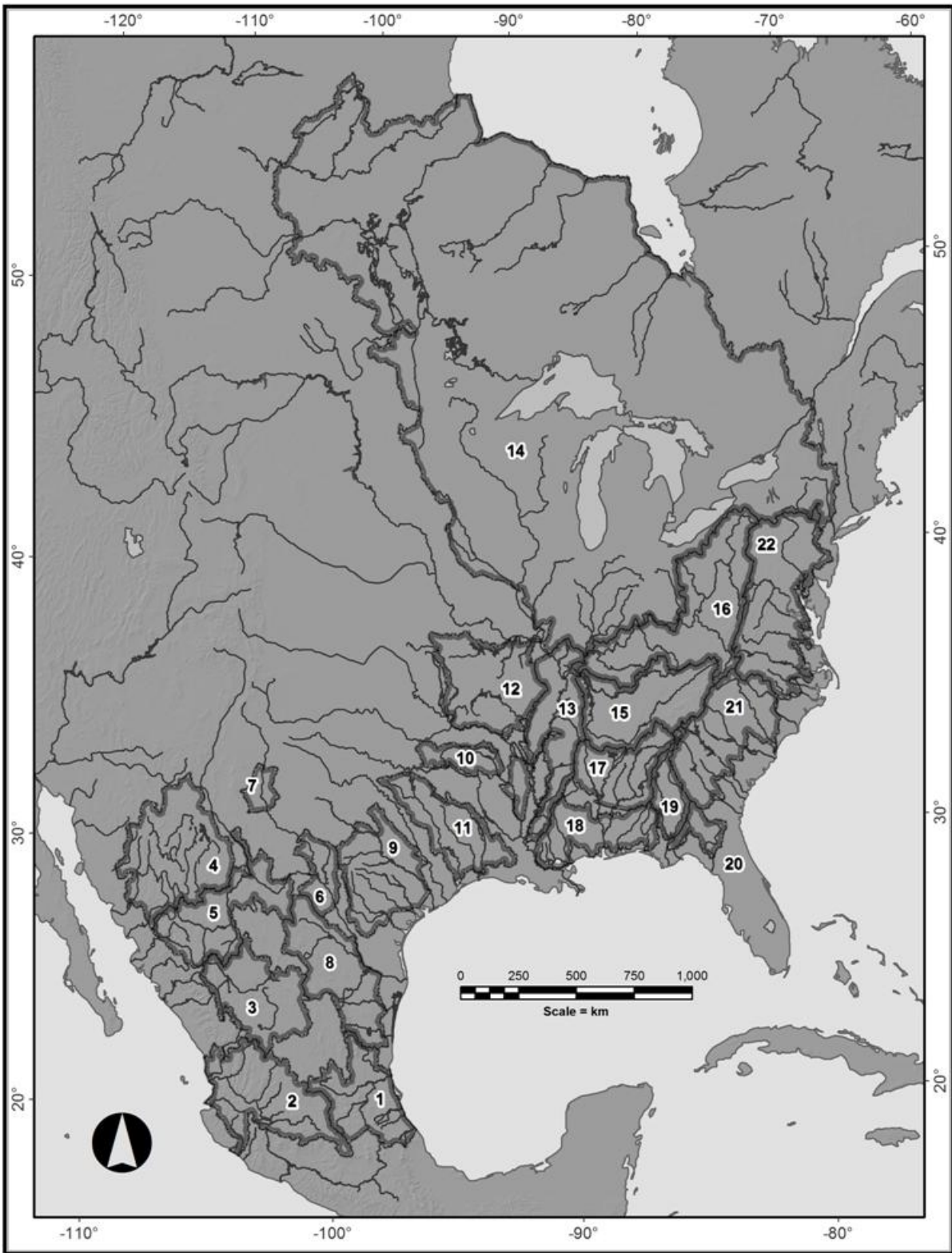
Clade	Description	Geographical taxa (number)	References
<i>Campostoma</i>	Cyprinid genus <i>Campostoma</i>	10	Blum <i>et al.</i> , 2008; Dominguez-Dominguez <i>et al.</i> , 2011; Schönhuth <i>et al.</i> , 2011 M.J. Blum (Tulane University, personal communication, 2012)
<i>Cyprinella</i>	Cyprinid genera <i>Codoma</i> , <i>Cyprinella</i> , & <i>Tampichthys</i>	29	Schönhuth & Mayden, 2010
<i>Moxostoma</i>	<i>M. lachneri</i> clade, genus <i>Moxostoma</i>	9	Jenkins, 1980; Doosey <i>et al.</i> , 2010
<i>Micropterus</i>	Genus <i>Micropterus</i>	8	Near <i>et al.</i> , 2003, 2005; Bagley <i>et al.</i> , 2011
<i>Oligocephalus</i>	Percid sub-genera <i>Oligocephalus</i> & <i>Microperca</i> of <i>Etheostoma</i>	31	Near <i>et al.</i> , 2011; Suttkus <i>et al.</i> , 2012
<i>Pileoma</i>	Sub-genus of <i>Percina</i>	12	Near & Benard 2004; Near <i>et al.</i> , 2011

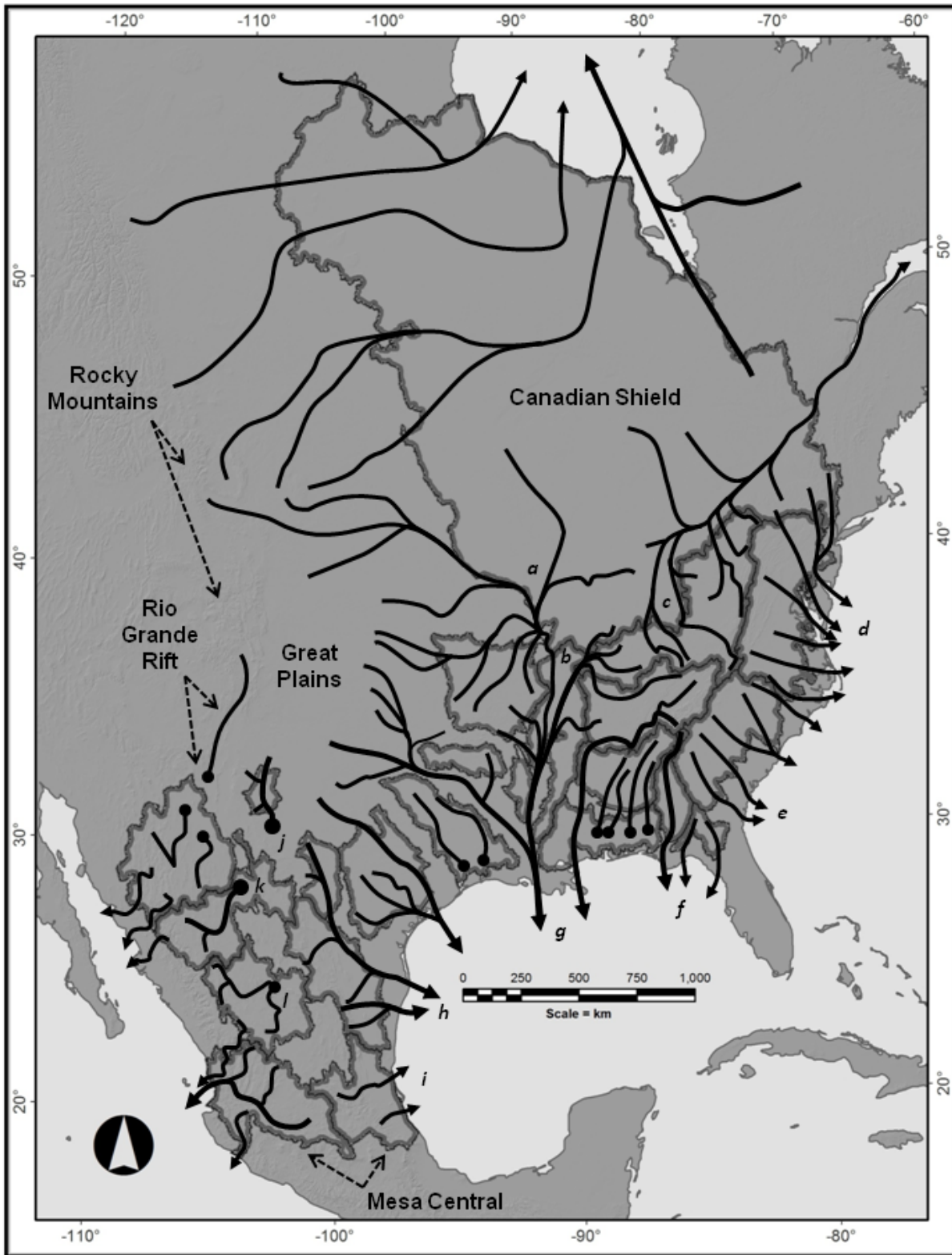
Table 2 Derived members of fish clades used in this study (Table 1) that are not restricted to rocky, highland-type habitats and are widespread across alluvial plains. These are putative stage 1 and stage 2 species within the taxon cycle (Ricklefs & Cox, 1972; Ricklefs & Bermingham, 1999) and are listed by region.

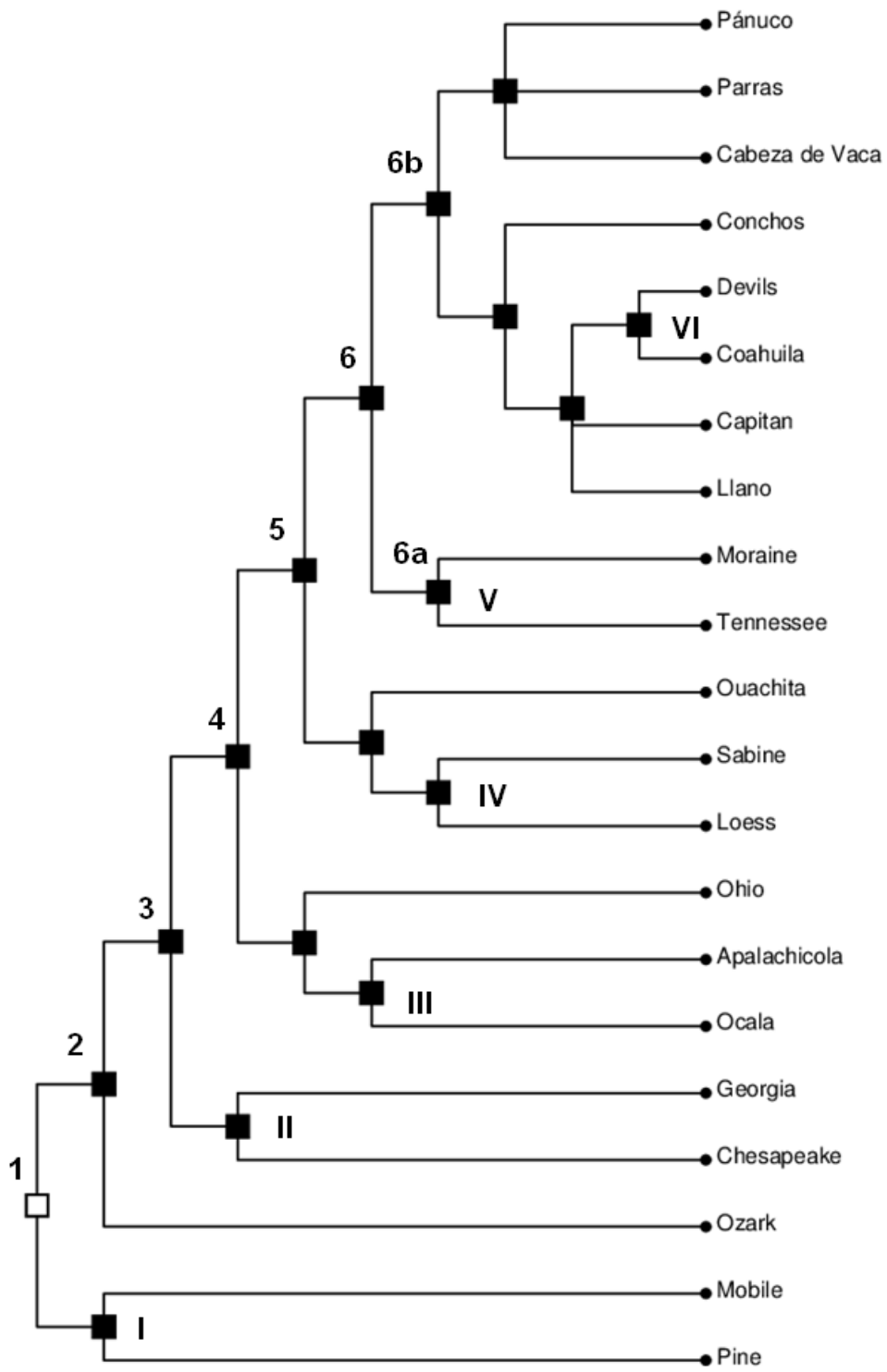
Clade	Coastal Plains	Glacial Plains	Great Plains	Rio Grande Rift
<i>Campostoma</i>		<i>C. pullum</i> <i>C. oligolepis</i>	<i>C. plumbeum</i>	
<i>Cyprinella</i>	<i>C. leedsii</i> <i>C. suavis</i> <i>C. venusta</i>	<i>C. analostana</i> <i>C. lutrensis</i> <i>C. spiloptera</i> <i>C. whipplei</i>	<i>C. lutrensis</i>	<i>C. lutrensis</i>
<i>Moxostoma</i>	<i>M. poecilurum</i>	<i>M. duquesnii</i> <i>M. valenciennesi</i>		
<i>Micropterus</i>	<i>M. floridanus</i> <i>M. punctulatus</i> <i>M. salmoides</i>	<i>M. dolomieu</i> <i>M. salmoides</i>		
<i>Oligocephalus</i>	<i>E. collettei</i> <i>E. proeliare</i> <i>E. swaini</i> group	<i>E. asprigene</i> <i>E. caeruleum</i> <i>E. exile</i> <i>E. spectabile</i>	<i>E. pulchellum</i>	
<i>Pileoma</i>	<i>P. macrolepida</i> <i>P. suttkusi</i>	<i>P. caprodes</i>	<i>P. caprodes</i>	

Figure 1 Highland areas occupied by endemic highland-fish taxa used in this study. (a) Modern river drainages in association with highland areas. Areas are: 1 Pánuco, 2 Lerma, 3 Parras, 4 Cabeza de Vaca, 5 Conchos, 6 Devils, 7 Capitan, 8 Coahuila, 9 Llano, 10 Ouachita, 11 Sabine, 12 Ozark, 13 Loess, 14 Moraine, 15 Tennessee, 16 Ohio, 17 Mobile, 18 Pine, 19 Apalachicola, 20 Ocala, 21 Georgia, 22 Chesapeake. (b) Late Miocene river drainages based on literature review (in text) and results of this study (Fig. 2). Arrow-headed rivers deposited sediments into the ocean, whereas circle-headed rivers deposited sediments inland. Major boundaries to the study area are labelled. Depositional basins referred to in text include: *a* Illinois Basin, *b* Mississippi Embayment, *c* Appalachian Basin, *d* Salisbury Embayment, *e* Southeast Georgia Embayment, *f* Southwest Georgia Embayment, *g* Mississippi-Red-Tennessee Delta, *h* Burgos Basin, *i* Tampico-Misantla Basin, *j* Delaware Basin, *k* Chihuahua Trough, and *l* Parras Basin.

Figure 2 Intersection tree showing evolutionary relations among highland areas (Fig 1) based on a three-item analysis of six highland-fish clades (Table 1, Appendix S1). Major nodes are labelled with Arabic numerals, whereas sister-area relations are labelled with Roman numerals.







Annexe VI : Comments on the International Code of Area Nomenclature (ICAN)

Zaragüeta-Bagils R., Bourdon E., **Ung V.**, Vignes-Lebbe R. et Malécot V. (2009)
Journal of Biogeography, **38** : 1617-1619

a useful approach, both to identify possible null expectations of relationships that differ significantly from zero and to serve as a check on those direct relationships that can be tested against a null expectation of zero.

In conclusion, on average 62% of the variation in the number of extinctions on islands could be explained primarily by the size of the original avifauna if extinction were random. In our original analysis of the real data (Karels *et al.*, 2008), all of the variables explained only 30% of the variation in the number of extinctions. The pattern of extinctions of island avifaunas has considerable non-random structure and this is truly 'something of interest' and importance that requires further study.

TIM J. KARELS^{1*}, F. STEPHEN DOBSON^{2,3}, HEATHER S. ZIMBLER-DELORENZO³ AND AMY L. SKIBIEL³

¹Department of Biology, California State University, 18111 Nordhoff Street, Northridge, CA 91330, USA, ²Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 du CNRS, 1919 route de Mende, F-34293 Montpellier Cedex 5, France and ³Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

*E-mail: karels@csun.edu

ACKNOWLEDGEMENT

We owe T. M. Blackburn thanks for suggesting that we run simulations of randomly generated data to test the efficacy of our path model. Although we disagree with his conclusions concerning the use of path analysis for explaining influences on extinctions in oceanic avifaunas, we are edified by the exchange of ideas.

REFERENCES

- Arbuckle, J.L. (2007) *Amos 16.0 user's guide*. Amos Development Corporation, Spring House, PA.
- Blackburn, T.M., Cassey, P., Duncan, R.P., Evans, K.L. & Gaston, K.J. (2004) Avian extinction and mammalian introductions on oceanic islands. *Science*, **305**, 1955–1958.
- Blackburn, T.M., Cassey, P., Evans, K.L., Gaston, K.J. & Duncan, R.P. (2009) The biogeography of avian extinctions on oceanic islands revisited. *Journal of Biogeography*, **36**, 1613–1614.
- Didham, R.K., Ewers, R.M. & Gemmill, N.J. (2005) Comment on 'avian extinction and mammalian introductions on oceanic islands'. *Science*, **307**, 1411–1412.
- Duncan, R. & Forsyth, D. (2006) Modelling population persistence on islands: mammal introductions in the New Zealand archipelago. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2969–2975.
- Escudero, A., Iriondo, J.M., Olano, J.M., Rubio, A. & Somolinos, R.C. (2000) Factors affecting establishment of a gypsophyte: the case of *Lepidium subulatum* (Brassicaceae). *American Journal of Botany*, **87**, 861–871.
- Hood, G.M. (2000) *PopTools – software for the analysis of ecological models, version 3.0.6*. CSIRO. Available at: <http://www.cse.csiro.au/poptools> (last accessed 28 September 2008).
- Karels, T.J., Dobson, F.S., Trevino, H.S. & Skibiell, A.L. (2008) The biogeography of avian extinctions on oceanic islands. *Journal of Biogeography*, **35**, 1106–1111.
- Kline, R.B. (2005) *Principles and practice of structural equation modeling*. Guilford Press, New York, NY.
- Li, C.C. (1975) *Path analysis: a primer*. Boxwood Press, Pacific Grove, CA.
- Petraitis, P.S., Dunham, A.E. & Niewiarowski, P.H. (1996) Inferring multiple causality: the limitations of path analysis. *Functional Ecology*, **10**, 421–431.
- Steadman, D. & Martin, P.S. (2003) The late Quaternary extinction and future resurrection of birds on Pacific islands. *Earth Science Reviews*, **61**, 133–147.
- Trevino, H.S., Skibiell, A.L., Karels, T.J. & Dobson, F.S. (2008) Importance of causal analysis of threats to oceanic avifaunas: reply to Blackburn *et al.* *Conservation Biology*, **22**, 495–497.
- Editor: Lisa Manne
- doi:10.1111/j.1365-2699.2009.02129.x

On the International Code of Area Nomenclature (ICAN)

ABSTRACT

A recent work proposed a new naming system for biogeographical areas: the International Code of Area Nomenclature (ICAN). This system aims to standardize area names used in biogeography and constitutes a key step towards maturity of this discipline. We point out a number of issues in this first version of ICAN. These include the lack

of distinction between classification and nomenclature, the use of ranks, the criteria of area definition, the conflation of diagnosis and description, the ambiguous use of valid/available, and the system of amendments. We suggest postponing the initial date of application of this code until consistent rules are provided.

Keywords Amendment, area definition, biogeography, classification, description, diagnosis, ICAN, nomenclature, rank, type-locality.

Ebach *et al.* (2008) recently proposed a new naming system for biogeographical areas: the International Code of Area Nomenclature (ICAN). This system aims to standardize area names used in biogeography and constitutes a key step towards maturity of this discipline. However, if ICAN's foundations lack precision, ICAN will actually complicate the naming of areas. We would like to raise some critical observations that will contribute to the development of an internally consistent ICAN.

CLASSIFICATION VS. NOMENCLATURE

First, taxonomic nomenclature is ruled by five codes, not three (as ICAN stipulates). In addition to codes cited by Ebach *et al.* (2008), one can mention the International Code of Nomenclature for Cultivated Plants (Brickell *et al.*, 2004) and the International Code of Virus Classification and Nomenclature (ICVCN) (in Fauquet *et al.*, 2005).

Like the Phylocode, ICAN does not distinguish between nomenclature and classification. In their introduction, Ebach *et al.* (2008) state that '[n]omenclatural rules have been constantly revised during the 250 years since Linnaeus first proposed his classification system'. Despite the fact that nomenclature has been revised since Linnaeus (1736) proposed his nomenclatural rules, current nomenclatural codes date mainly from the 19th century (Strickland *et al.*, 1843; Anonymous, 1847; de Candolle, 1867). A nomenclatural code is not intended to construct classifications, but is used to name entities. In other words, nomenclatural rules concern names, not taxa or areas. ICAN Sec. A Art. 3.2 states that 'ICAN does not endorse any particular method among those in use for validating or classifying areas'. Such a statement is trivial. Likewise, in Sec. C Art. 1.1, it is stated that '[a]rea names may be grouped under more inclusive area names in order to represent a

biogeographical taxonomic hierarchy'. The name 'North America', however, does not contain the name 'New York'. The concept of hierarchy applies to areas, not to names.

A nomenclatural code aims to create objective links between names and concepts (taxa, taxonomic assemblages, temporal intervals or areas) through an object, i.e. a particular specimen called a holotype, whose properties are constant through time. In contrast, the concept associated with a name is subject to scientific discussion. To the greatest extent possible, the link between a scientific concept and the name that refers to that concept must be kept devoid of subjectivity. Similarly, a nomenclatural code must also be normative: decisions taken following its rules must be subject to as little interpretation as possible. Nomenclature should deal with: (1) the objective link between a name and the real world through an object, i.e. the holotype – or, more properly, the onomatophore (the name bearer); (2) synonymy, homonymy and the principle of priority; and (3) definition of publication, validity and availability of names. In its current form, the ICAN draft fails to specify some of these points.

RANKS

In ICAN, the use of ranks is quite arbitrary, because there may be many more areas (nodes) than ranks, at least in some methods such as area cladistics and Brooks parsimony analysis.

It is not stated in ICAN how higher-rank areas inherit the type-locality of names given to included areas. In a ranked taxonomy, the onomatophore of a species is a holotype, the onomatophore of a genus is its type species and that of a family is its type genus. This system guarantees the link between the name of any supraspecific taxon and the holotype without multiplying onomatophores. In ICAN, the type-locality of a higher-rank area can be different from those of included areas. This renders decisions about synonymy/homonymy virtually impossible.

The statement that '[a] rank can only have one valid name' (Sec. C Art. 2.7) is meaningless. ICAN's authors probably meant 'an area can have only one valid name at a given rank'.

AREA DEFINITION

Ebach *et al.* (2008) state that areas of endemism 'have been recognized as the principal units of modern biogeographical analyses'. Each one is viewed as 'the area

delimited by the distribution range of a taxon'. ICAN, however, specifies that only a type-locality and diagnosis or description is needed to define a biogeographical area (Sec. C Art. 2.1). Examples given in the introduction only include geographical descriptions and do not take taxa into account. This contradicts the statement that '[t]he Northern Adriatic would be synonymized with the Central Adriatic, for example, if it were concluded that the presumed endemics live in both the Central and Northern Adriatic'. In this case, taxonomic distributions are used to describe the Adriatic, not geography. In addition, it is uncertain whether the northern Adriatic and central Adriatic would be synonymized if only *some* endemics were found in both areas. ICAN's authors should clarify their criteria, because the consistency of ICAN depends on the rigour of area definition.

Further, we did not find in ICAN any reference to the temporal significance of named areas. For instance, Ebach *et al.* (2008) describe the Adriatic by the presence of the Foci del Po. Yet we do not know where the Foci del Po was placed in the Eocene. We suggest that an area name should be associated with an explicit temporal interval (e.g. Recent/Pleistocene Foci del Po) because current biogeographical studies cannot deal with geographical variations through time, and limit themselves to a single geological age. This will prevent synonymization of area names based on studies concerning organisms of different geological ages.

DIAGNOSIS VS. DESCRIPTION

Ebach *et al.* (2008) conflate diagnosis and description throughout the ICAN draft (including Sec. C Art 2). For instance, they state that 'the name and type-locality must be linked to a diagnosis (ICAN Sec. C Art. 2.3), which may consist of a detailed description of the area'. However, ICAN (Sec. C Art 4.2) states that '[a] name can be rejected if it has the same diagnosis, description, geographical coordinate or distribution as an existing name'. The latter article distinguishes these two concepts without means of comparing them, i.e. without explaining in which case a description may be identical to a diagnosis. According to the Botanical (McNeill *et al.*, 2006; art. 32.2) and the Zoological (International Commission on Zoological Nomenclature, 1999) Codes, a diagnosis is a statement of the characters that distinguish a taxon from other taxa. In contrast, a

description is a list of properties of the taxon itself (Vignes, 1991). In the context of ICAN, a diagnosis should compare the extent of the described area with those of published areas, while the 'description' paragraph should include a list of localities and endemic taxa. In reference to ICAN Sec. C Art 4.2 we suggest that decisions about rejection be taken only for diagnosis vs. diagnosis and description vs. description.

PUBLICATION

ICAN fails to define the date of publication of a name. Ebach *et al.* (2008) state that names have to 'be published in a refereed journal', which implies that monographs or technical/edited books are excluded from the naming system. Sec. C Art. 2.6 states that '[n]ames that are published with a diagnosis and/or descriptions linked to a previous discussion of reference are valid', whereas Sec. C Art. 2.8 states that '[t]he principles of homonymy, synonymy and priority apply to the names within any rank that have valid diagnoses'. Accordingly, it is unclear whether validity applies to names and/or diagnoses. Moreover, Sec. C Art. 2.1 specifies that '[a]n available area name has a type-locality and either a published diagnosis or a description in a refereed publication'. Following this ambiguous statement, an unpublished area name can be available just because a type-locality and a description are published. Likewise, ICAN leaves uncertain which distinction, if any, is made between 'available' and 'valid' throughout the ICAN draft. The two distinct uses of the term valid in the Botanical (valid publication of a name) and Zoological codes (valid name) generate the aforementioned confusion.

AMENDMENTS

Modifications of ICAN rules are made through a series of amendments (Sec. B Art. 2), instead of being directly incorporated into the code. Thus, independent modifications of an article and its subsequent amendments will eventually be impossible to track. In the future, ICAN will be unusable if this system of amendment is maintained.

CONCLUSION

In this paper, we point out a number of inconsistencies and ambiguous statements in the first version of ICAN. We suggest

postponing the initial date of application of this code until internally consistent rules are provided. Members of the Systematic and Evolutionary Biogeographical Association are warmly encouraged to provide a completed version of ICAN. Such a universal naming system will significantly increase the efficiency of communication in biogeography and other related disciplines.

RENÉ ZARAGÜETA-BAGILS^{1*},
ESTELLE BOURDON², VISOHEARY
UNG¹, RÉGINE VIGNES-LEBBE¹ AND
VALÉRY MALÉCOT³

¹UPMC Université Paris 06, Paléobio-
diversité et Paléoenvironnements,

²Collège de France, Mécanismes Adapta-
tifs, UPMC Université Paris 06, Paris and

³AGROCAMPUS OUEST – Centre
d'Angers, Génétique et Horticulture, INRA,
Université d'Angers, Angers, France

*E-mail: rene.zaragüeta_bagils@upmc.fr

REFERENCES

- Anonymous (1847) The rules of American pomology. *Horticulturist Journal of Rural Art and Rural Taste*, 2, 273–275.
- Brickell, C.D., Baum, B.R., Hetterscheld, W.L.A., Leslie, A.C., McNeill, J., Trehane, P., Vrugtman, F. & Wiersema, J. (2004) *International Code of Nomenclature for Cultivated Plants*, 7th edn. ISHS, Leuven.
- de Candolle, A.P. (1867) *Lois de la nomenclature botanique adoptées par le Congrès International de Botanique*. H. Georg, Geneva.
- Ebach, M.C., Morrone, J.J., Parenti, L.R. & Vilorio, A.L. (2008) International Code of Area Nomenclature. *Journal of Biogeography*, 35, 1153–1157.
- Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U. & Ball, L.A. (2005) *Virus taxonomy*. Classification and nomenclature of viruses. 8th Report of the ICTV. Elsevier, San Diego.
- International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature*. The International Trust for Zoological Nomenclature, London.
- Linnaeus, C.V. (1736) *Fundamenta botanica*. S. Schouten, Amsterdam.
- McNeill, J., Barrie, F.R., Burdet, H.M., Demoulin, V., Hawksworth, D.L., Marhold, K., Nicolson, D.H., Prado, J., Silva, P.C., Skog, J.E., Wiersema, J.H. & Turland, N.J. (2006) *International Code of Botanical Nomenclature (Vienna code)*. A.R.G. Gantner, Verlag KG.
- Strickland, H.E., Philipps, J., Richardson, J., Owen, R., Jenyns, L., Broderip, W.J., Henslow, J.S., Shuckard, W.E., Waterhouse, G.R., Yarrell, W., Darwin, C. & Westwood, J.O. (1843) Report of a committee appointed 'to consider the rules by which the Nomenclature of Zoology may be established on a uniform and permanent basis'. *Report of the 12th Meeting of the British Association for the Advancement of Science (Manchester 1842)* pp. 105–121. J. Murray, London.
- Vignes, R. (1991) *Caractérisation automatique de groupes biologiques*. Unpublished PhD Thesis, Université Pierre et Marie Curie, Paris.
- et al., 2007). An explanatory discussion and the text of the ICAN were published subsequently as a Guest Editorial in the *Journal of Biogeography* (Ebach et al., 2008). That editorial was the subject of a critique by Zaragüeta-Bagils et al. (2009) to which we reply briefly here. We are grateful for the attention paid by these authors to the ICAN, particularly by their comment that it '...constitutes a key step toward maturity of [biogeography]'. We offer some comments in anticipation of the benefits of public debate. Our subject headings mirror those in Zaragüeta-Bagils et al.'s (2009) critique.

Editor: David Bellwood

doi:10.1111/j.1365-2699.2009.02106.x

On the International Code of Area Nomenclature (ICAN): a reply to Zaragüeta-Bagils et al.

ABSTRACT

In 2007 the Systematic and Evolutionary Biogeographical Association (SEBA) wrote and ratified the first draft of the International Code of Area Nomenclature (ICAN), which was posted subsequently on the SEBA website. The ICAN was published, along with an explanatory discussion, by Ebach et al. (*Journal of Biogeography*, 2008, 35, 1153–1157), an article that is the subject of criticism by Zaragüeta-Bagils et al. (*Journal of Biogeography*, 2009, 36, 1617–1619). We welcome discussion of the issues raised by these authors and respond to them briefly here. For many reasons, we reject the proposition that implementation of the ICAN be postponed until it is flawless. The ICAN has already been implemented. Further, it is the nature of nomenclatural codes to be proposed and then revised periodically to suit our applications. Most importantly, standardization of area names in biogeography is long overdue.

Keywords Area classification, area names, area nomenclature, area ranks, ICAN, standards in biogeography.

The first draft of the International Code of Area Nomenclature (ICAN) was written and ratified by members of the Systematic and Evolutionary Biogeographical Association (SEBA) and posted on its website (Ebach

CLASSIFICATION VERSUS NOMENCLATURE

The explanatory discussion and the ICAN in Ebach et al. (2008) are explicit about the relationship between classification and nomenclature: the ICAN is concerned with nomenclature, not with particular methods to construct biogeographical classifications. That we mentioned the nomenclatural rules and classification of Linnaeus in the same sentence of the explanatory discussion does not mean that we equate or confuse these concepts (Zaragüeta-Bagils et al., 2009). This is one example of how Zaragüeta-Bagils et al. (2009) unfortunately conflate the explanatory discussion accompanying the ICAN and the ICAN itself. Similarly, they state that the 'ICAN stipulates' (Zaragüeta-Bagils et al., 2009) that taxonomic nomenclature is ruled by three codes. The three nomenclatural codes for natural living organisms – not those cultivated or viral – are noted in the explanatory discussion, not the ICAN.

Zaragüeta-Bagils et al.'s (2009) rejection of the ICAN statement that '[a]rea names may be grouped under more inclusive area names in order to represent a biogeographical taxonomic hierarchy' using the argument that a hierarchy applies to areas, not names, is beside the point. A hierarchy of areas may be represented by words, by a complex map, or, ideally, both.

RANKS

We agree that the use of classificatory ranks is arbitrary and gave several examples in our explanatory discussion of how in practice these may be applied or even ignored. The ICAN is criticized because '...the type-locality of a higher-rank area can be different from those of included areas', and therefore, in the opinion of Zaragüeta-Bagils et al. (2009), it is impossible to make

Annexe VII : Xper²: introducing e-Taxonomy

Ung V., Dubus G., Zaragüeta-Bagils R. & Vignes-Lebbe R. (2010)

Bioinformatics, 26 (5) 703-704.

Xper²: introducing e-taxonomy

Visotheary Ung*, Guillaume Dubus, René Zaragüeta-Bagils and Régine Vignes-Lebbe
UPMC, Université Pierre et Marie Curie, UMR 7207 (CR2P), Centre de Recherche sur la Paléobiodiversité et les Paléoenvironnements, Paris, France

Associate Editor: Alex Bateman

ABSTRACT

Motivation: Computer Aided Identification systems provide users with the resources to relate morpho-anatomic observations with taxa names and to subsequently access other knowledge about the organisms. They have the ability to manage descriptive data and make identifications through interactive keys. They are essential for both authors and users of biodiversity information. Xper² version 2.0 is one of the most user-friendly tools in its category and provides a complete environment dedicated to taxonomic management.

Availability: Xper² software can be freely downloaded at <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/softwares/xper2>

Contact: visotheary.riviere-ung@snv.jussieu.fr

Received on October 15, 2009; revised on November 24, 2009; accepted on December 28, 2009

1 INTRODUCTION

Linnaeus wrote that naming is the first stage of knowledge. Indeed, names of taxa are the main access keys to biological information (Polaszek, 2005). They link the real world of living organisms and the conceptual world proposed by systematists through the description and characterization of taxa. Identifying life forms is crucial for all activities related to biology and their taxonomic descriptions are more than necessary in the study of Biodiversity (Godfray, 2002; Lebbe and Vignes, 1991, 1998). Systematists inventory, study and structure biological diversity as precisely as possible (Matile *et al.*, 1987). All this information may be summarized into knowledge bases to perform identification of specimens for new inventories and monitoring surveys, phylogenetic analyses, as well as biogeographic and ecological studies. A significant increase of this kind of digitized information should be expected in a near future. The deep social and scientific impact of web 2.0 stimulates sharing of digital data and proposals of 'cybertaxonomy' projects to study global biodiversity and climate changes. The generalization and the integration of knowledge-base management systems into current taxonomic work may be the next revolution in taxonomy. International projects are already opening the way of this new, delocalized taxonomy.

2 WHAT IS XPER²?

Xper² is a management system for storage, editing, analysis and on-line distribution of descriptive data. It dynamically creates interactive keys for identifying specimens (Fig. 1). Xper² does

*To whom correspondence should be addressed.

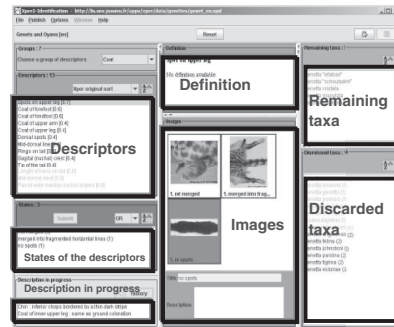


Fig. 1. A screenshot of the identification module (Genets and Oyas knowledge base).

not require any special computer skills. It is written in Java and available on Windows™ Mac™ or Linux in French, English or Spanish versions. It has a user-friendly and intuitive interface aimed at professional taxonomists as well as naturalists who merely want to identify specimens using a ready-made application. Xper²'s users are taxonomists, teachers, fauna and flora experts (for biodiversity monitoring or survey) and ecologists. Formalization of knowledge is the main difficulty for taxonomists. How to structure and organize different types of data? How to keep the scientific sources of the data in order to maintain traceability? Xper² is designed to manage this rich and heterogeneous information. Divided into four modules, Xper² allows taxonomists to either access the descriptive environment, the taxonomic description module, the management tools or the free access key:

- (1) descriptive model: creation and editing of a standardized description (structured into characters, states, groups and logical dependencies);
- (2) taxonomic editor: characteristics of the taxon, i.e. its name, descriptions, comments and pictures;
- (3) management tools (see Section 3.4); and
- (4) free access key: creation of interactive identification keys (IK).

3 NEW FUNCTIONALITIES

3.1 Import/export structural descriptive data

Many international research programs focus on the establishment of standard formats for data and knowledge exchange. Xper² implements the structured descriptive data (Hagedorn *et al.*, 2006), the exchange format for taxonomic descriptions that has emerged from the work of the Biodiversity Information Standards (TDWG, <http://www.tdwg.org>, formerly known as taxonomic database working group). Xper² can seamless import/export data from/to SDD, following international recommendations.

3.2 Nexus export

Xper² extends the descriptive morpho-anatomic work and the taxonomic expertise of taxonomists into phylogenetic studies. A knowledge base can be exported in nexus format and directly be run in most phylogeny programs such as PAUP*.

3.3 HTML export

Xper² provides users with a user-friendly tool to instantly format the knowledge base in a set of HTML files according to a defined cascading style sheet provided by the application. These pages include a form for each taxon displaying its complete description, pictures and any additional comments. By this way, monographies can be easily computed and published in order to serve the entire community. As an example, a website on monitor lizards (<http://lis-upmc.snv.jussieu.fr/varanID>) has been created by a naturalist group from the Paris Museum.

3.4 An accurate knowledge representation and more tools to analyze descriptive data

Xper² offers solutions to control and prevent inconsistencies via a check base function. Furthermore, analyzing descriptive data is one of the most important taxonomists' tasks and Xper² gives tools to carry it out:

- Xper² is able to manage uncertainty by selecting several descriptor-states and using logical operators (logical AND, logical XOR, logical NOT).
- Mismatch threshold: one can define a different threshold according to his expertise level when performing identification.
- Instant comparison of all taxa in a matrix summarizing descriptions.

4 FORTHCOMING DEVELOPMENTS

Management of extended descriptor types are under development. This includes management of numerical data, hierarchical or structured descriptor states, computed values, or any other user-declared type (Chalubert and Vignes, 2006). Associated tools allowing writing descriptions or diagnoses in natural language or proposing different kinds of similarity index among taxa will be implemented in future versions.

5 AVAILABILITY

Xper² is a versatile tool for editing and using taxonomic descriptions. It is free of charge and can be downloaded at <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2>. A mailing-list facility provides users with any help or advice that may be needed. Authors choose how they want to use or publish their knowledge base: distributed on CD (i.e. Mascarene Corals identification tool, Conruyt *et al.*, 2000) or online publication (e.g. XperBotanica—a project to identify the flora of France, <http://lis-upmc.snv.jussieu.fr/xperbotanica>).

6 CONCLUSION

Many international projects such as GBIF, EDIT, BIOTA, Catalogue of Life, Encyclopedia of Life, Key2Nature, etc. have a common purpose: sharing knowledge and providing free web access to Biodiversity Information (Bisby *et al.*, 2002; Ramsey and Janzen, 2002; Tillier and Roberts, 2006). In this context of collaborative work, Xper² fits with current taxonomists' practices and gathers functionalities to perform a suitable *e*-taxonomy!

ACKNOWLEDGEMENTS

The authors thank B. Pavie, C. Gallut and A. Chalubert for their contribution to the initial Xper² version in Java. They thank all the Xper² users and especially Fabienne Audebert, Ona Maioeco, Philippe Gaubert and David Caron for their bases and also members of the EDIT project for their comments on the software. They are also very grateful to Dr Kurt Liittschwager for comments and style improvement.

Funding: Xper² development has been partially funded by the BioInfo 2002 grant of the CNRS and by the French Ministère for Research and New Technology (No. 04L370—Project Xper Botanica, 2005–07).

Conflict of interest: none declared.

REFERENCES

- Bisby, F. *et al.* (2002) Taxonomy at a click of a mouse. *Nature*, **418**, 367.
- Chalubert, A. and Vignes, R. (2006) A New Model for Descriptive Knowledge. *Proceedings of TDWG Annual Meeting 2006*. Available at <http://www.tdwg.org/proceedings/article/view/32>.
- Conruyt, N. *et al.* (2000) IKBS, a knowledge base management system for reengineering systematics: application to corals of the Mascarene Archipelago. *ICRS 2000—Ninth International Conference for Reef Studies*. ICRS, Bali, p. 17.
- GBIF. *Global Biodiversity Information Facility*. <http://www.gbif.fr/>
- Godfray H.C.J. (2002) Challenges for Taxonomy. *Nature*, **417**, 17–19.
- Hagedorn, G. *et al.* (2006). The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.1.
- Lebbe, J. and Vignes, R. (1991) Génération de graphes d'identification à partir de descriptions de concepts. In Kodratoff, Y. and Diday, E. (eds) *Induction Symbolique et Numérique à partir de données*. Cepadues, France, pp. 193–239.
- Lebbe, J. and Vignes, R. (1998). State of the art in computer-aided identification in biology. *Oceanis*, **24**, 305–317.
- Matile, L. *et al.* (1987) Introduction à la systématique zoologique. *Biosystema*, 1–67.
- Polaszek, A. (2005) A universal register for animal names. *Nature*, **437**, 477.
- Ramsey, S. and Janzen, D. (2002) All living things, online. *Nature*, **418**, 262–263.
- Tillier, S. and Roberts, D. (2006) Taxonomy on the fly in a European web project. *Nature*, **440**, 24.

Annexe VIII : IKEY+ : a new single-access key generation web service.

Burguiere T., Causse F., **Ung V.** et Vignes-Lebbe R. (2013)

Systematique Biology, 62(1): 157-161.

Software for Systematics and Evolution

Syst. Biol. 62(1):157–161, 2013
© The Author(s) 2012. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/sys069
Advance Access publication on August 3, 2012

Key+: A New Single-Access Key Generation Web Service

THOMAS BURGUIERE*, FLORIAN CAUSSE, VISOTHEARY ÜNG, AND RÉGINE VIGNES-LEBBE

Laboratoire Informatique et Systématique, Muséum National d'Histoire Naturelle, Département Histoire de la Terre,
UMR 7207 – C.N.R.S. - M.N.H.N. - U.P.M.C., Paris, France

*Correspondence to be sent to: Thomas Burguiere, Laboratoire Informatique et Systématique, M.N.H.N Département Histoire de la Terre,
Bâtiment de Géologie, CP48, 57 rue Cuvier, 75005 Paris, France,
E-mail: thomas.burguiere@upmc.fr

Thomas Burguiere and Florian Causse contributed equally to this article.

Received 2 April 2012; reviews returned 29 June 2012; accepted 25 July 2012
Associate Editor: David Posada

Abstract.—Single-access keys are a major tool for biologists who need to identify specimens. The construction process of these keys is particularly complex (especially if the input data set is large) so having an automatic single-access key generation tool is essential. As part of the European project ViBRANT, our aim was to develop such a tool as a web service, thus allowing end-users to integrate it directly into their workflow.

Key+ generates single-access keys on demand, for single users or research institutions. It receives user input data (using the standard SDD format), accepts several key-generation parameters (affecting the key topology and representation), and supports several output formats.

Key+ is freely available (sources and binary packages) at www.identificationkey.fr. Furthermore, it is deployed on our server and can be queried (for testing purposes) through a simple web client also available at www.identificationkey.fr (last accessed 13 August 2012). Finally, a client plugin will be integrated to the Scratchpads biodiversity networking tool (scratchpads.eu). [Systematics; taxonomy; single-access key; web service; biodiversity informatics.]

During the last decade, taxonomy has been going through a revolution towards cybertaxonomy (Wheeler 2008). This is particularly obvious when looking at the numerous initiatives dedicated to the management and monitoring of biodiversity data such as Global Biodiversity Information Facility (GBIF), European Distributed Institute of Taxonomy (EDIT), BIOTA, Catalogue of Life, Encyclopedia of Life (EOL) or Key To Nature, and now ViBRANT. A significant increase of digitized information should be expected in a near future, and the deep social and scientific impact of web 2.0 stimulates sharing of digital data and proposals of cybertaxonomy projects to study global biodiversity and climate changes. In the current context of distributed knowledge and long-distance collaborations, the European project ViBRANT (<http://vbrant.eu>, last accessed 13 August 2012) has emerged to bridge the existing gap between tools and services, providing a powerful exemplar platform, the Scratchpads, for the entire community.

ViBRANT is about connecting the people, data and science of biodiversity (ViBRANT - Objectives)

As studying and monitoring biodiversity remains a strong challenge for biologists, descriptive data management and identification tools are needed to assist them in their daily work. For instance, we developed a complete descriptive data management tool: Xper² (Ung et al. 2010).

Our contribution to ViBRANT is the implementation of an identification key generation web service, because we also have a significant experience in developing key generation tools such as MyKey (Gérard and Vignes-Lebbe 2010). The most common identification keys (described in Hagedorn et al. 2010) are the printed single-access keys published in almost all taxonomic revisionary work. But these printed keys may be difficult to use while collecting samples in the field or when consulting incomplete specimens. Hence, electronic single-access keys, which can be quickly generated with different parameters in order to bypass potentially problematic steps (i.e., steps with characters for which the specimen studied in the field cannot provide information) are a good alternative.

APPROACH

Web Service

There are currently several tools designed to generate single-access keys such as, among others, the Delta software suite (Dallwitz et al. 1993), Pankey (Pankhurst 1988), and MyKey. These tools are well established and produce valid identification keys, but are not without drawbacks. For instance, both Delta and Pankey need to be installed on the end-user's machine and are not cross-platform compatible (Windows, Apple, and Linux). Mykey is a web application, which makes it platform-agnostic, but the end-users cannot submit their own data

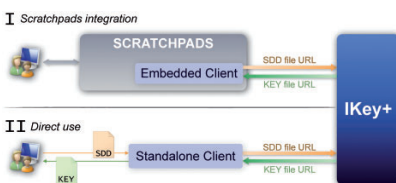


FIGURE 1. IKey+ use cases.

set to generate a key (the data set has to be manually uploaded by someone in our laboratory). Furthermore, these 3 tools do not use the biodiversity description standard file format, Structured Descriptive Data (SDD) (Hagedorn 2006), for data input.

Our objective, within the ViBRANT project, was to design a free and open-source key generation service using the SDD file format that can be used by independent users or through the ViBRANT OBOE service (also available through Scratchpads). It was thus decided to develop a web service (implementing both the SOAP and REST web service communication protocols), which is particularly suitable for these two use cases (cf. Fig. 1) and allows the client to build complex workflows, with a high level of automation.

Architecture

Because the main purpose of ViBRANT is to provide a platform that is both open source and easily reusable, we organized IKey+ in two parts:

- an Application Programming Interface (API) consisting of 3 distinct modules; the Data Model, that is, the computer representation of the descriptive data, the Input-Output (IO) module, which parses the SDD input files and loads the data in the Model and generates the output files, and the Algorithm, which uses the data contained in the Model to generate a key and returns it through the IO module.
- a Simple Object Access Protocol (SOAP) or a REpresentational State Transfer (REST) Service Layer encapsulating the API, which manages the communication between the client and the API.

This allows potential developers to integrate the service into their workflows and adapt it to their specific needs (e.g., integrating the API in a standalone software). The whole application was developed using the J2EE (Java 2, Enterprise Edition) programming environment.

Web Service Input Parameters

IKey+ has several parameters available to the end-user. These parameters allow the end-user to change the

topology of the key (e.g., pruning the key, promoting some characters), change the visual aspect of the key, or the output format. A complete list of these parameters can be found in the user documentation, available here: http://www.identificationkey.fr/resources/docs/identificationKeyGeneratorWS_UserGuide.pdf (last accessed 13 August 2012).

Algorithm

The single-access key generation algorithm is a recursive depth-first graph construction algorithm. Its arguments are a list of taxa, *taxaList* and the list of character considered, *charList* (cf. Supplementary Materials, appendices 1, 2 and 3, available at <http://datadryad.org>, doi:10.5061/dryad.3ft19). The resulting graph (cf. Fig. 2) is a directed acyclic graph, consisting of non-terminal nodes labelled with a character, edges labelled with the states of the character of the previous node, and terminal nodes labelled with taxa. At each step of the algorithm, the best character among those available is selected by the BEST_CHAR function, which iterates over the available characters and returns the character with the greatest discriminant power.

Discriminant Power of a Character

During the last 50 years, estimating a character's discriminant power has been the central point of key-construction algorithms. Many measurement methods have been suggested, and a review of these methods can be found in Gower and Payne (1975), Pankhurst (1991), and Delgado-Calvo-Flores et al. (2006). In a statistical context, examples include the Bayesian probability and generalized entropy, such as the Shannon entropy used in the ID3 algorithm developed by Quinlan (1986) or the Gini index used in Breiman et al. (1984). In a context where there is no probability associated with the states of the characters, one can use the separation coefficient or the variance as measurement methods of a character's discriminant power.

In IKey+, the discriminant power of a given character C is calculated by the DPOWER function and is an estimate of C 's ability to differentiate the taxa of the current list. DPOWER is an extension of the Gyllenberg separation factor (Gyllenberg 1963) and

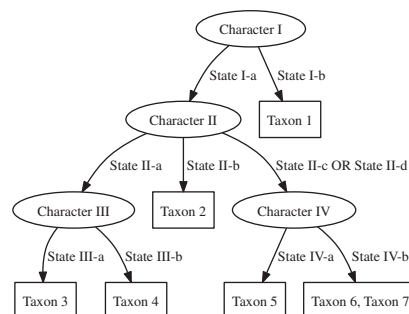


FIGURE 2. Formal identification key example.

is a measurement of the number of pairs of taxa discriminated by C , which is a generalization of the variance. Indeed, the variance of a variable can be computed by comparing all pairs of values (here, comparing all pairs of taxa description for C). With different comparison functions, each adapted to a certain type of character (e.g., a categorical or a numerical character), it is possible to obtain a generalized formula to estimate the discriminant power of a given character, even with polymorphic characters or characters with missing data. For categorical characters, DPOWER can use a binary function (boolean comparison) or other functions such as the Sokal and Michener coefficient (Sokal and Michener 1958) or the Jaccard coefficient (Jaccard 1901). For numerical characters (e.g., a size measured in millimetre), the set of values (i.e., all the values entered for C for all taxa) is split in two intervals. In order to determine the threshold value separating these two intervals, we consider the list of *min* and *max* value of C for each of the remaining taxa. We then choose from this list the value that separates the remaining taxa into two groups of equal size (± 1 taxon). These two intervals are considered as 2 discrete states for the calculation of the discriminant power.

DPOWER iterates over the available taxa (i.e., those that are compatible with the current description), and determines, for each pair of taxa T_a and T_b , the dissimilarity (i.e., the possibility of discriminating T_a and T_b with C) that is based on the number of common states of C for T_a and T_b (n_{11}), the number of states of C that occur only for T_a (n_{10}), the number of states of C that occur only for T_b (n_{01}), and the number of states of C that do not occur for neither T_a or T_b (n_{00}). The discriminant power is then calculated as

$$DP = \sum_a \sum_b \text{SCORE}(n_{11}, n_{10}, n_{01}, n_{00}).$$

Three different SCORE measurements are currently available in iKey+, the Xper coefficient (Ung et al. 2010;

Vignes et al. 1989), the Jaccard coefficient, and the Sokal and Michener coefficient.

BENCHMARKS

Tests

In order to assess the performance of our web service, we conducted a series of tests using the same input file for every test. This file contains a data set representing the subfamily Cichorieae of the plant family Asteraceae with 303 observable characters, 144 taxa, and their descriptions. This data set was generated and curated as an exemplar group for the EDIT project (Hand et al. 2009) [the file is available here: <http://www.infosyslab.fr/vibrant/project/test/Cichorieae-fullSDD.xml> (last accessed 13 August 2012)]. The aim of these tests was to evaluate the overall performance of the algorithm, to assess the impact of the various parameters available to the end-user on performance, and to ensure that iKey+ would be robust in a production environment. We measured the time necessary for the web service to respond to 100 identification keys generation queries, while counting the number of rejected queries due to CPU overload (iKey+ rejects any query received when the CPU load of the host server is greater than 80%). We ran one reference test (described in Supplementary Materials, appendix 4, doi:10.5061/dryad.3ft19), and several tests with variations on the input parameters, the web service communication protocol used or the parallelization setups. The complete list of performance test setups is available in Supplementary Materials, appendix 5, doi:10.5061/dryad.3ft19. We also measured the time necessary to generate a single key, using the reference test configuration described in Supplementary Materials, appendix 4, doi:10.5061/dryad.3ft19. Finally, we tested the usability of iKey+ from an end-user perspective, when generating the Cichorieae identification key using the web interface. We asked 11 persons who were not

involved in the development of *IKey+* to test the web service and the web interface. They were given the SDD formatted Cichorieae data set and were asked to use the web interface to create the identification key. We measured the time needed by each test subject to generate the identification key.

Results

The average length of the paths leading to taxa that were identified by the algorithm is 4.67 steps, with the shortest path being 1 step long, and the longest being 10 steps long. The generation of a single key took roughly 1.8 s. The results of the other tests are shown in Supplementary Materials, appendix 6, doi:10.5061/dryad.3ft19. The reference test took roughly 50 s to complete, that is, 500 ms per query, which is consistent with the time measured for the generation of a single key, because the reference test uses 4 simultaneous threads to generate the keys. In this test, few queries were rejected due to CPU overload. Among the parameters available to the end-user, only the score method parameter had a significant influence: when using the Sokal and Michener score method or the Jaccard score method, the tests took longer to finish (~80 s instead of 50 s). The score method parameter had no impact on the number of rejected queries. Our tests showed that the communication protocol used (SOAP or REST) had no influence on the performance of *IKey+*.

Some taxa do not appear in the generated key, due to insufficient data in the input file. Indeed, the Cichorieae data set we used for our tests was created by an external team. This is because some errors were made during creation of the Cichorieae data set: some taxa with unknown data were not specifically marked as "unknown data", but were left unspecified instead. Regardless, although these ambiguously coded taxa do not appear in the resulting identification key by default, the end-user can choose to have them appear in the key.

When modifying the parallelization of the queries, we observed significant performance variations. As can be expected, when launching 100 queries sequentially (instead of using 4 simultaneous threads launching 25 queries each), the test was 4 times longer, and no queries were rejected. When we augmented the parallelism of the queries (e.g., 25 or 100 simultaneous queries), more queries were rejected (up to 50%). However, when launching 100 simultaneous queries, with a random delay at the beginning of each thread, the results (both in time and number of rejected queries) were comparable to the performance of the reference test. In the usability test, the average time needed to generate the identification key was slightly above 60 s (60.36 s), with the shortest time measured at 26 s, and the longest time measured at 121 s. The complete results of the usability test are available in Supplementary Materials, appendix 7, doi:10.5061/dryad.3ft19.

DISCUSSION

IKey+ is available and can be installed on any J2EE application server (e.g., Apache Tomcat). It can generate single-access keys using a tree or a flat representation in several output formats (HTML, Wiki, SDD, plain-text, etc.).

Our test showed that *IKey+* performs sufficiently well to handle a large data set in a relatively short amount of time and can generate well-optimized key files (average number of steps to identify a taxon: 4.67). This, combined with the web service accessibility, makes it possible to integrate *IKey+* in many workflows that might require a fast and automated key generation process (e.g., a batch key generation script). Furthermore, an end-user can use the web interface available at www.identificationkey.fr to quickly generate a customized identification key, using the numerous parameters available (affecting the topology, representation, file format, etc.).

Finally, our tests showed that *IKey+* is likely to be robust in a production environment (i.e., many simultaneous queries) as it is able to withstand simultaneous key generation queries (e.g., 4 threads launching 25 consecutive queries). It is also protected from cryptic failure, because we implemented a CPU-load-watching mechanism that automatically rejects a query (with an explicit error message) whenever the CPU load exceeds a given threshold (80%). This prevents a crash of the service, or the generation of incomplete or corrupt key files.

CONCLUSION

IKey+ is the first key-generation tool available as a web service with standardized input and output formats. Our test showed that *IKey+* is able to generate keys rapidly and that it can also be used by an end-user with the web interface. Finally, the modular and open-source nature of *IKey+* makes it possible for anyone to reuse its components. For instance, we plan to reuse some components of the API to develop another web service that would provide free-access key identification.

LICENSING

As part of the ViBRANT project, *IKey+*'s source code is freely available and is licensed under the GNU General Public License version 2. It is already available on our google code SVN repository: <http://ikey-plus.googlecode.com/svn/trunk/> (last accessed 13 August 2012). It will be actively maintained by our team for the next 2 years.

SUPPLEMENTARY MATERIAL

Supplementary material, including Algorithms and appendices, can be found at <http://www.sysbio.oxfordjournals.org>.

FUNDING

This work was supported by the European Union funded FP7 ViBRANT Project (Contract number RI-261532, Period, December 2010 to November 2013).

ACKNOWLEDGEMENTS

We sincerely thank Gregor Hagedorn (Julius Kühn Institute, Berlin, Germany) and Andreas Müller (Botanical Garden and Botanical Museum, Berlin, Germany) for sharing their knowledge on the SDD format. We are also grateful to Dave Roberts (Natural History Museum, London, UK) for reviewing an early version of the article and providing style improvements.

REFERENCES

- BIOTA. Available from: URL <http://www.edinburgh.ceh.ac.uk/biota/> (last accessed 13 August 2012).
- Breiman L., Friedman J.H., Olshen R.A. Stone C.J. 1984. Classification and regression trees. Belmont, CA:Wadsworth International Group.
- Catalogue of Life. Available from: URL <http://www.catalogueoflife.org/> (last accessed 13 August 2012).
- Dallwitz M.J., Paine T.A., Zurcher E.J. 1993. User's guide to the delta system: a general system for coding taxonomic descriptions. 4th ed. Available from: URL <http://delta-intkey.com> (last accessed 13 August 2012).
- Delgado-Calvo-Flores M., Fajardo-Contreras W., Gibaja-Galindo E.L., Perez-Perez R. 2006. Xkey: a tool for the generation of identification keys. *Expert Syst. Appl.* 30:337–351.
- EDIT. European Distributed Institute of Taxonomy. Available from: URL <http://www.e-taxonomy.eu/> (last accessed 13 August 2012).
- EOL. Encyclopaedia of Life. Available from: URL <http://eol.org/>.
- GBIF. Global Biodiversity Information Facility. Available from: URL <http://www.gbif.org/> (last accessed 13 August 2012).
- Gérard D., Vignes-Lebbe R. 2010. Mykey: a server-side software to create customized decision trees. In: Nimis P.L., Vignes-Lebbe R., editors. Tools for identifying biodiversity: progress and problems. Edizioni Università di Trieste, Trieste, Italy. p. 107–112.
- Gower J.C., Payne R.W. 1975. A comparison of different criteria for selecting binary tests in diagnostic keys. *Biometrika* 62:665–672.
- Gyllenberg H.G. 1963. A general method for deriving determinative schemes for random collections of microbial isolates. *Ann. Acad. Scient. Fenn. Ser. A IV. Biologica* 1(69):1–23.
- Hagedorn G. 2006. The structured descriptive data (SDD) w3c-xml-schema. Version 1.1 Available from: URL <http://wiki.tdwg.org/twiki/bin/view/SDD/Version1dot1> (last accessed 13 August 2012).
- Hagedorn G., Rambold G., Martellos S. 2010. Types of identification keys. In: Nimis P.L., Vignes-Lebbe R. editors. Tools for identifying biodiversity: progress and problems. Edizioni Università di Trieste, Trieste, Italy. p. 59–64.
- Hand R., Kilian N., Raab-Straube E. 2009. International cichorieae network: Cichorieae portal. Available from: URL <http://wp6-cichorieae.e-taxonomy.eu/portal/> (last accessed 13 August 2012).
- Jaccard P. 1901. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull. Soc. Vaud. Sci. Nat.* 37:547–579.
- J2EE. Java 2, Enterprise Edition. Available from: URL <http://www.oracle.com/technetwork/java/javaee/overview/index.html> (last accessed 13 August 2012).
- KeyToNature. Available from: URL <http://www.keytonature.eu/wiki/> (last accessed 13 August 2012).
- Pankhurst R.J. 1988. Pankey programs. DELTA Newsletter 1:2.
- Pankhurst R.J. 1991. Practical taxonomic computing. Cambridge University Press, Cambridge, UK.
- Quinlan J.R. 1986. Induction of decision trees. *Mach. Learn.* 1:81–106. ISSN 0885-6125. Available from: URL <http://dx.doi.org/10.1007/BF00116251> (last accessed 13 August 2012).
- REST Architecture. Available from: URL <http://www.oracle.com/technetwork/articles/javase/index-137171.html> (last accessed 13 August 2012).
- Scratchpads. Biodiversity Online. Available from: URL <http://scratchpads.eu/> (last accessed 13 August 2012).
- SOAP. Simple Object Access Protocol, W3C Recommendation. Version 1.2. Available from: URL <http://www.w3.org/TR/soap/> (last accessed 13 August 2012).
- Sokal R., Michener C. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.*, (38):1409–1438.
- Ung V., Dubus G., Zaragüeta-Bagils R., Vignes-Lebbe R. 2010. Xper2: introducing e-taxonomy. *Bioinformatics* 26(5):703–704.
- ViBRANTa. Objectives. Available from: URL <http://vbrant.eu/node/1> (last accessed 13 August 2012).
- ViBRANTb. Virtual Biodiversity Research and Access Network for Taxonomy. Available from: URL <http://vbrant.eu> (last accessed 13 August 2012).
- Vignes R., Lebbe J., Darmoni S. 1989. Symbolic-numeric approach for biological knowledge representation: a medical example with creation of identification graphs. In E. Diday, editor, Proceedings of the conference on Data analysis, learning symbolic and numeric knowledge. Nova Science Publishers, Inc. Commack, NY, USA. ISBN 0-941743-64-0. p. 389–398.
- Wheeler, Q.D., editor 2008. The new taxonomy. CRC Press Inc. New York, USA.

Annexe IX : MKey+: Interactive identification webservice

Thomas Burguiere, Antoine Bergamaschi, Florian Causse, **Visotheary Ung**, Nils
Paulhe, Frédéric Lison, and Régine Vignes-Lebbe (2013)

En préparation

MKey+: Interactive identification webservice

Thomas Burguiere*^{†‡}, Antoine Bergamaschi[‡], Florian Causse, Visotheary Ung,
Nils Paulhe, Frédéric Lison, and Régine Vignes-Lebbe

Laboratoire Informatique et Systématique, UMR 7207 – C.N.R.S. - M.N.H.N. - U.P.M.C.
Museum National d'Histoire Naturelle – Département Histoire de la Terre, Paris, FRANCE

May 2013

Abstract

Interactive identification keys are a major tool for systematists who need to identify specimens or to transmit their taxonomic expertise to other scientists or amateurs. Our latest software, MKey+, using the SDD standard file format for the input of a descriptive knowledge base and open, standard, web technologies (HTTP, JSON) allows scientists to perform interactive identification online, and brings interactive identification capabilities to any application capable of sending webservice queries.

keywords: Systematics, Taxonomy, Interactive identification key, webservice, Biodiversity informatics

*to whom correspondence should be addressed

[†]thomas.burguiere@upmc.fr

[‡]Thomas Burguiere and Antoine Bergamaschi contributed equally to this work

Identification keys constitute a major tool for systematist when they need to quickly identify a specimen, using a reference key. The most common identification keys used by scientists (described in Hagedorn et al. 2010) are the printed single-access keys published in almost all taxonomic revisionary work. They consist in a sequence of questions (whose order is fixed by the author of the key) regarding the morphological characters commonly used to describe a single or multiple species. This sequence of questions, depending on their answer, lead the user on a path towards the identification of a specific specimen (*cf.* example in Figure 1). These single-access keys, however, are not without drawbacks.

For instance, the sequence of the questions is necessarily a compromise between the discriminant power of the characters (to reduce the number of steps needed to identify a specimen, the most discriminant characters should appear early in the sequence) and the observation convenience of the characters (if a character which can't be described for a specific specimen comes too early in the sequence, the systematist is unable to identify this specimen).

Multiple-access keys provide a good alternative to single-access keys, since they allow the user to perform an interactive identification *i.e.* to determine the sequence of questions, so that the user may skip a question which can't be answered. These keys require a software component, commonly called a key Player, with which the user chooses the questions order, answer them, and identifies a specimen. Existing software include the Delta software suite (Dallwitz et al. 1993, 2013), the Lucid software suite and our own Xper² software (Ung et al. 2010). All these programs are desktop software, and need to be installed on each user's machine, prior to any interactive identification. Some (Xper² and Lucid) include a web applet, which makes it possible to deploy an interactive key on a web server, but the configuration process is quite cumbersome and the applet technology has substantial security vulnerabilities. There is currently no way for a systematist to upload his complete knowledge base and start an interactive identification right away.

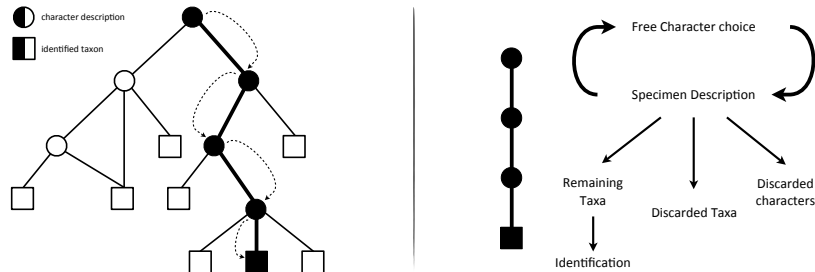


Figure 1: Comparison of a single-access key (left) and a multiple-access key (right). The single-access key is basically a decision tree and the identification process the traversal of this tree.

In the multiple-access key, the identification process consists in the dynamic construction of a traversal path of the potential identification tree.

Within the framework of european project ViBRANT (<http://vbrant.eu>), which aims at bridging the existing gap between scientists, scientific tools and scientific services, we decided to develop an interactive identification software tool, MKey+. We used the same approach used to develop our single-access key generation software, IKey+ (Burguiere et al. 2013). This approach consists of developing a webservice which provides the main functionalities necessary to perform interactive identification, using the SDD standard format (Hagedorn 2006) for data input, that can be readily queried by HTTP requests to exchange JSON-formatted data (Crockford 2006), then developing a simple web client available online. With this approach, a single user is able to perform an interactive identification using the web client, and software developers can add interactive identification capabilities to their software, provided it is capable of interacting with the webservice using HTTP requests.

Methods

Performing an interactive identification using MKey+. The interactive identification process can be decomposed into 4 steps.

- I The user, using any web client, submits the knowledge base to MKey+.
 - II MKey+ returns the list of the characters which the user can describe, sorted by their discriminant power.
 - IIIa The user describes a character with the client, and submits this description to MKey+.
 - IIIb MKey+ returns the list of taxa which are compatible with all the descriptions it has received so far, and the client gets back to step IIIa.
- When there is only one remaining taxon, the user has successfully completed his identification.

Discriminant power of a character. One of the defining feature of the software we develop (Xper², IKey+, and now MKey+) is the calculation of the discriminant power of the characters. This discriminant power is crucial to reduce the number of steps needed to identify a specimen. It indicates the user, at each step of the identification, which of the remaining characters (*i.e.* the characters that haven't been described yet) are the most discriminant for the remaining taxa (*i.e.* the taxa that have not been ruled out at a previous step).

There are many methods to calculate the discriminant power of a character, and a review of these methods can be found in Gower and Payne (1975), Pankhurst (1991) and Delgado-Calvo-Flores et al. (2006). In MKey+, the discriminant power of a given character C is an estimate of C 's ability to differentiate the taxa of a given list. It is an extension of the Gyllenberg separation factor (Gyllenberg 1963) and is a measurement of the number of pairs of taxa discriminated by C , which is a generalisation of the variance. Indeed, the variance of a variable can be computed by comparing all pairs of values (here, comparing all pairs of taxa description for C). This method is the same as the one we used in IKey+ and Xper², and is thoroughly described in Vignes (1991) and Burguiere et al. (2013).

Interacting with MKey+. As stated above, the fact that MKey+ is a webservice makes it possible to interact with in two different ways:

1. a scientist who needs to perform a single identification can use the simple web client we developed ¹ to upload his knowledge base, then start the identification. The only requirements are a modern web browser, an internet connection and an SDD formatted knowledge base.
2. a scientific software developer, without any knowledge of the algorithms involved in the interactive identification process, can write a component which will interact with MKey+ very easily, provided his program is capable of sending HTTP requests. One example of such an integration would be to include an MKey+ client component into the Scratchpads (<http://scratchpads.eu>, Smith et al. 2011) biodiversity networking tool.

MKey+ is fully functional and is available online, and our web client is available for scientists who want to perform interactive identifications right away. Similarly, software developers can freely add an MKey+ client component to their software, a complete technical description of the identification functions is available (*cf.* supplementary data). This brings instant full interactive identification online, which we consider to be a significant improvement over existing tools, which at the very least require that users install a key player on their computer. Furthermore, the fact that MKey+ is a webservice makes it possible to consider connecting it to other third-party webservices, which would extend the capabilities of MKey+.

MKey+ is currently capable of handling two kinds of characters: categorical characters which can be described with discrete states (*e.g.* Shape of the leaf: round / square), and quantitative characters, which can be described with continuous values (*e.g.* length of the tail: 3cm). Future developments are being considered, to give MKey+ the ability to handle other types of characters, such as color space characters, sequence characters, *etc.* . . . Finally, we intend to use MKey+ as the interactive identification engine for the next version of our descriptive

¹available at <http://www.identificationkey.fr/mkey/mkey.html>

data management software, Xper³.

Visotheary Ung, Guillaume Dubus, René Zaragüeta-Bagils, and Régine Vignes-Lebbe. Xper2: introducing e-taxonomy. *Bioinformatics*, 26(5):703–4, Mar 2010. doi: 10.1093/bioinformatics/btp715.

Régine Vignes. *Caractérisation automatique de groupes biologiques*. PhD thesis, Université Pierre et Marie Curie, Apr 1991.

References

- Scratchpads: Biodiversity Online. URL <http://scratchpads.eu/>.
- ViBRANT: Virtual Biodiversity Research and Access Network for Taxonomy. URL <http://vbrant.eu>.
- Thomas Burguiere, Florian Causse, Visotheary Ung, and Régine Vignes-Lebbe. Ikey+: a new single-access key generation web service. *Systematic Biology*, 62(1):157–161, Jan 2013. doi: 10.1093/sysbio/sys069.
- D. Crockford. The application/json media type for javascript object notation (json), July 2006. URL <http://tools.ietf.org/html/rfc4627>.
- M.J. Dallwitz, T.A. Paine, and E.J. Zurcher. User's guide to the delta system: a general system for coding taxonomic descriptions. 4th edition, 1993. URL <http://delta-intkey.com>.
- M.J. Dallwitz, T.A. Paine, and E.J. Zurcher. Principles of interactive keys, Jan 2013. URL <http://delta-intkey.com/www/interactivekeys.htm>.
- M. Delgado-Calvo-Flores, W. Fajardo-Contreras, E.L. Gibaja-Galindo, and R. Perez-Perez. Xkey: A tool for the generation of identification keys. *Expert Systems with Applications*, 30:337–351, 2006.
- J.C. Gower and R.W. Payne. A comparison of different criteria for selecting binary tests in diagnostic keys. *Biometrika*, 62:665–672, 1975.
- H.G. Gyllenberg. A general method for deriving determinative schemes for random collections of microbial isolates. *Ann. Acad. scient.fenn. Ser. A, IV. Biologica*, 1(69):1–23, 1963.
- Gregor Hagedorn. The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.1, 2006. URL <http://wiki.tdwg.org/twiki/bin/view/SDD/Version1dot1>.
- Gregor Hagedorn, Gerhard Rambold, and Stefano Martellos. Types of identification keys. In Pier Luigi Nimis and Régine Vignes-Lebbe, editors, *Tools for Identifying Biodiversity: Progress and Problems*, pages 59–64. Edizioni Università di Trieste, 2010.
- Richard J. Pankhurst. *Practical taxonomic computing*. Cambridge University Press, 1991. ISBN 0-521-41760-0.
- Vincent Smith, Simon Rycroft, Irina Brake, Ben Scott, Edward Baker, Laurence Livermore, Vladimir Blagoderov, and David Roberts. Scratchpads 2.0: a virtual research environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys*, 150:53 – 70, 2011/11/28/ 2011. URL <http://dx.doi.org/10.3897/zookeys.150.2193>.

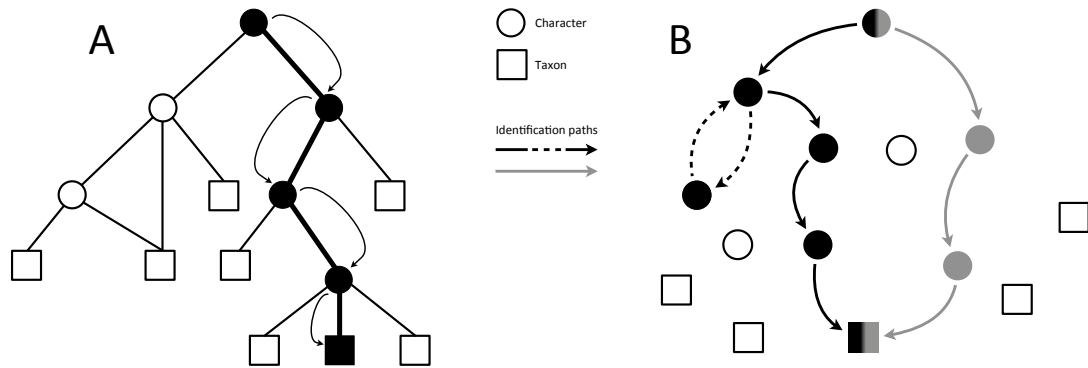


Figure 1: Comparison of a single-access key identification and a multiple-access key identification.

The single-access key (**A**) is essentially a decision tree and the identification process is the traversal of this tree.

In the multiple-access key (**B**), the identification process consists in the dynamic construction of an identification path. A specimen can be identified using different characters, thus creating different identification paths. This is shown in the figure on the right in which both the black and the grey identification paths start from the same character, but depending on the description given by the user, create different identification paths that lead to the same taxon.

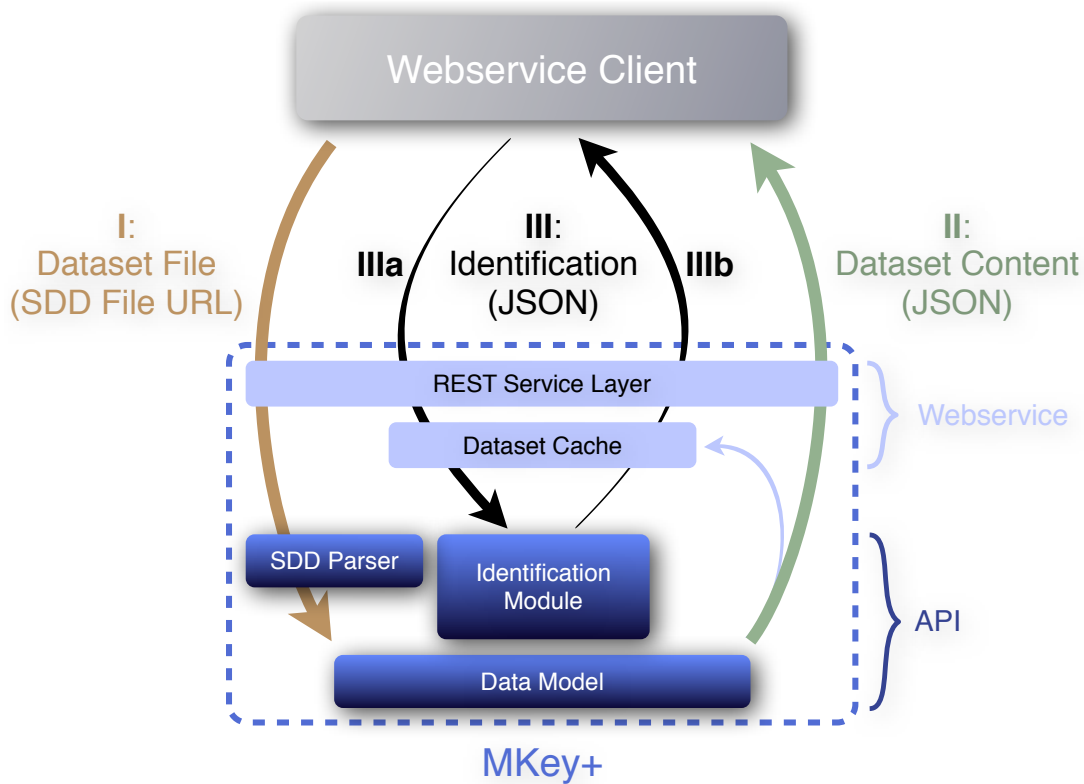


Figure 2: Interactive identification workflow, using MKey+.

Step **I**: the client sends the URL of the SDD file containing the dataset on which the identification will be performed. The SDD file is retrieved by the MKey+ via the REST Service Layer, its content is parsed and loaded in the Data Model.

Step **II**: the Dataset content is sent back to the client as a JSON character string, with the characters sorted by their discriminant power in descending order. The Dataset content is also stored in the Dataset cache.

Step **IIIa**: The specimen is described by the user (via the client), i.e. the user describes one character. This description of a single character is sent back to Mkey+ via the REST Service Layer.

Step **IIIb**: This description is passed to the identification module, which analyses it, returns the list of remaining taxa contained in the dataset that are compatible with this description, and the list of remaining characters sorted by their new discriminant power. The user goes back to step **IIIa**, and when only one taxon remains, the identification is complete.

User No.	Time (s)	Nb Steps
1	145	7
2	125 †	7
3	245	7
4	145	7
5	205	7
6	119	8
7	177	7
8	174	8
9	230	7
10	248 †	7
11	127	7
12	138 †	7
Min	119	-
Max	248	-
Mean	173.17	-
Median	159.50	-
SD	47.89	-

Figure 3: Usability test results

The times marked by a † symbol indicates users who did not identify the correct specie on their first attempt. The time included in this table is the time they needed to finish their second (and successful) attempt.

RÉSUMÉ

La biogéographie historique est la discipline qui s'occupe d'établir les relations de parenté entre aires biogéographiques, comprises comme des ensembles d'aires de distribution de taxons. L'histoire géologique complexe du Sud Est asiatique et de la région Indo-Pacifique en fait une région de choix pour les biogéographes. Nous proposons une nouvelle approche, cladistique, pour l'étude de cette région située autour de la célèbre "Ligne de Wallace", en nous plaçant dans le paradigme de la vicariance où la spéciation apparaît suite à la mise en place d'une barrière géographique (par opposition au modèle dispersaliste à partir d'un centre d'origine). En effet, nous utilisons ici l'analogie existant entre Systématique et Biogéographie pour appliquer des méthodes cladistiques de la Systématique à la Biogéographie. Il s'agit ici de répondre à la question : « *pourquoi existe-t-il une telle distribution des taxons dans cette région du globe ?* » Les résultats attendus sont une meilleure connaissance de l'histoire évolutive de la région ainsi que des traits d'histoire de vie des taxons qui la peuplent. Etant entendu que la distribution conjointe de plusieurs taxons non apparentés ne peut être obtenue par « chance », elle doit forcément rendre compte d'un événement commun (i.e. un événement de vicariance) qui aurait affecté de manière égale ces taxons. L'approche cladistique, que nous proposons n'a été rendue possible que grâce à la finalisation d'un logiciel implémentant la méthode dite d'analyse à 3 éléments (3ia): LisBeth est aujourd'hui disponible en téléchargement librement depuis notre site web et offre de nouvelles perspectives d'analyses à ses utilisateurs, qu'ils soient systématiciens ou biogéographes. Nous montrons que la représentation des caractères offerte par LisBeth autorise de nouvelles inférences quant aux relations de parenté existant entre les aires d'endémismes étudiées. Nous avons ainsi mis en évidence l'émergence d'un « pattern » général de l'analyse conjointe de 35 taxons non apparentés (plantes et animaux).

Dans un deuxième temps nous nous sommes intéressés à une problématique théorique majeure de la biogéographie contemporaine: « *Quid de la dimension temporelle en biogéographie historique ?* ». Comment prendre en compte, sans ambiguïté, des échelles temporelles différentes lors d'analyses diachroniques ? Nous proposons ici un modèle, basé sur la 3ia qui permet de considérer la dimension temporelle et de réaliser, de ce fait, de véritables analyses de biogéographie intégrative.

MOTS-CLEFS : Biogéographie historique, biogéographie intégrative, 3ia, Asie du Sud-Est, Pacifique Ouest, pattern, process.

ABSTRACT

Historical biogeography is the science that deals with relationships between biogeographic areas, understood as set of taxa distributions. The complex geological history of South East Asia and the Indo-West Pacific region makes it a very interesting area for biogeography studies. We propose a new approach, cladistic, for the study of this area around the famous "Wallace's line", placing our study in the paradigm of vicariance where speciation appears following the implementation of a geographic barrier (as opposed to the dispersal model from a center of origin, classically admitted). Indeed, here we use the existing analogy between Systematics and Biogeography to apply cladistic methods for Systematics to Biogeography. This is to answer the question: "*Why is there such distributions of taxa in this region of the world?*" The expected results are a better understanding of the evolutionary history of the region and the life history traits of taxa that inhabit it. Provided that the joint distribution of several unrelated taxa cannot be obtained by "chance", thus it is necessarily due to a common event (ie a vicariant event) which affected equally these taxa. The cladistic approach we propose has been made possible by the completion of a software that implements the so-called 3-element analysis (3ia) method: LisBeth is now freely available for download from our website and offers new perspectives of analysis to its users, whether systematists and biogeographers. We show that the representation of characters offered by LisBeth allows new inferences about the existing relationship between areas of endemism studied. We have highlighted the emergence of a general "pattern" from a joint analysis of 35 unrelated taxa (plants and animals). In a second step we are interested in a major theoretical problem of contemporary biogeography: "*What about the time dimension in historical biogeography?*". How to account unambiguously different time scales during diachronic analyzes? We propose a model based on the 3ia which allows us to consider the temporal dimension and realize, therefore, true analysis of integrative biogeography.

KEY-WORDS : Historical biogeography, integrative biogeography, 3ia, Southeast Asia, West Pacific, pattern, process.