

UNIVERSITE MONTPELLIER II
SCIENCES & TECHNIQUES DU LANGUEDOC

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE MONTPELLIER II

Discipline : Biologie de l'Evolution et Ecologie

Formation Doctorale : Biologie de l'Evolution et Ecologie

Ecole Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosciences, et
Environnement

Présentée et soutenue publiquement par

Julien RENOULT

Le 24 Novembre 2009

Utilisation des marqueurs cytoplasmiques et des discordances cyto-nucléaires pour l'étude des processus évolutifs, démographiques et écologiques

JURY

M. N. Alvarez, Research Leader, Université de Neuchâtel

M. B. Godelle, Professeur, Université Montpellier II

M. F. Kjellberg, Directeur de Recherche, CNRS

M. P.-A. Crochet, Chargé de Recherche, CNRS

Examineur

Examineur

Directeur de thèse

Membre invité

M. J. Chave, Directeur de Recherche, CNRS

M. R. Petit, Directeur de Recherche, INRA

Rapporteur

Rapporteur

REMERCIEMENTS

Cette thèse n'a pu se réaliser que grâce au soutien de mon directeur de thèse, Finn Kjelberg, envers qui je suis pleinement reconnaissant. Un autre grand merci doit revenir à Pierre-André Crochet pour son soutien scientifique et amical durant ces années de thèse. Un grand merci à tous mes collaborateurs scientifiques : Bouchaïb Khadari, Philippe Geniez, Menad Beddek, Paul Bacquet, Laure Benoit, Sylvain Santoni et Alexandre Courtiol. C'est à ces personnes que fera référence, dans la suite du document, le pronom personnel *nous*. Je remercie également Catherine Soler, Martine Hossaert-McKey, Doyle McKey, Martin Schaefer, Marc De Dinechin, Frank Richard pour nos nombreuses interactions amicales et scientifiques.

Un grand merci également à Marie-Pierre, Patrick, Hélène, Laure et Chantal du service des marqueurs génétiques. Sincères remerciements également à Catherine, Marie, Jean-Yves et David pour leur relecture du présent document.

Je remercie les gens qui m'ont accompagné ou aidé sur le terrain : Auguste, Emile, Patricia, Steve et Verohanitra.

Merci à la fine bande de naturalistes pour m'avoir occupé les dimanches : Benoit, PH, Antoine et Cédric.

Bien entendu, je salue vivement tous les gens qui ont égayé mon séjour à Montpellier et ailleurs : Céline, Magali, Adeline, Sophie, Noppol, Mélanie, Laure, Romain, Tippie, Florence, Damien, Alexandre, Alex, Fred, Tom et bien d'autres.

Merci bien sûr à ma famille : mon frère David et ma sœur Emilie, mon père Jean-Yves et ma mère Marie-Ange. Merci aussi à Maya, Ezio et Alexis.

Les plus grands remerciements reviennent bien sûr à Marie, pour son soutien, ses bons conseils, ses bons plats et pour les bons moments passés et à passer ensemble.

SOMMAIRE

Introduction.....	6
I. Compartiments génomiques des cellules vivantes.....	10
I.A. Présentation et origine des organites intracellulaires	11
I.B. Les génomes cytoplasmiques : caractéristiques évolutives.....	13
I.B.1. La transmission maternelle.....	14
I.B.2. L'absence de recombinaison	17
I.B.3. Des taux de mutation extrêmes.....	20
II. Exploitation des caractéristiques architecturales et évolutives des génomes cytoplasmiques : des marqueurs pour retracer l'histoire du vivant	24
II. A. L'ADNmt animal.....	25
II.A.1. L'ADNmt animal utilisé comme marqueur pour les analyses de phylogénie et phylogéographie.....	25
II.A.2. L'ADNmt comme code-barres moléculaire.....	30
II.A.3. Etude de la distribution des lignées génétiques du complexe <i>Podarcis hispanicus</i> à l'aide de marqueurs mitochondriaux.....	31
II.B. L'ADNmt végétal.....	34
II.C. L'ADNcp.....	34
II.C.1. Les marqueurs chloroplastiques en phylogénie et en Phylogéographie.....	34
II.C.2. L'ADNcp comme code-barres moléculaire.....	35
II.C.3. Résolution des nœuds basaux de la phylogénie des <i>Galoglychia</i> grâce aux marqueurs chloroplastiques.....	36
III. Histoire des espèces versus histoire des génomes cytoplasmiques.....	40
III.A. Détection d'une discordance cyto-morphologique : le cas des <i>Podarcis</i> ibériques.....	41
III.B. Détection d'une discordance cyto-nucléaire : le cas des <i>Podarcis</i> ibériques.....	46
III.C. Test de la significativité d'une discordance cyto-nucléaire en phylogénie... ..	49
III.C.1. Méthodes générales.....	49
III.C.2. Test de la significativité de deux discordances phylogénétiquement emboîtées : le cas des figuiers <i>Galoglychia</i>	52
IV. Les causes de discordances cyto-nucléaires.....	57

IV.A. Discordances artéfactuelles.....	58
IV.A.1. Erreurs humaines et effet d'échantillonnage.....	58
IV.A.2. L'artefact d'attraction des longues branches.....	61
IV.A.2.a. L'extraction des longues branches.....	61
IV.A.2.b. L'attraction artificielle des longues branches par le groupe externe.....	61
IV.A.2.c. La simulation paramétrique.....	62
IV.A.2.d. La discordance méthodologique.....	63
IV.B. Discordances biologiques.....	63
IV.B.1. Paralogie.....	63
IV.B.2. Le tri incomplet du polymorphisme ancestral.....	65
IV.B.3. Hybridation introgressive.....	69
V. Les discordances cyto-nucléaires : un outil pour comprendre l'histoire des espèces.....	72
V.A. Histoire évolutive des lézards ibériques <i>Podarcis spp.</i> du sud-est de la Péninsule Ibérique.....	73
V.A.1. L'âge maximal de l'introgession mitochondriale.....	73
V.A.2. Origine de la distribution de la lignée mitochondriale Valencia chez <i>P. liolepis</i>	75
V.A.2.a. Hypothèse démographique.....	76
V.A.2.b. Hypothèse sélective.....	79
V.A.3. L'origine de la lignée Valencia.....	81
V.A.4. Conséquences taxonomiques de la discordance cyto-nucléaire	83
V.B. Exploitation des discordances cyto-nucléaires pour comprendre l'origine des incongruences entre la phylogénie des plantes et la phylogénie des pollinisateurs chez les figuiers <i>Galoglychia</i>	87
V.B.1. Incongruences phylogénétiques, changements d'hôtes et duplication/extinction.....	87
V.B.2. Changements d'hôtes chez les figuiers <i>Galoglychia</i> mis en évidence par les discordances cyto-nucléaires.....	90
Conclusion.....	95
Références Bibliographiques.....	101
Articles.....	130
Annexes : autres manuscrits rédigés durant la thèse.....	213

INTRODUCTION

Ce qui caractérise le mieux la biologie moderne est sans doute sa transdisciplinarité, c'est-à-dire l'association synergique de disciplines ayant pour objectif d'apporter de nouvelles lumières sur le fonctionnement du vivant. Si la plupart des associations disciplinaires ont contribué et continuent de contribuer à l'amélioration des connaissances, certaines auront davantage marqué l'histoire de la biologie. C'est le cas de l'écologie évolutive, qui a pour objectif de comprendre la diversité du vivant à travers l'étude de son organisation dans un contexte spatial et temporel (Fow *et al.* 2001). Tandis que l'écologie traditionnelle s'intéresse à la diversité contemporaine en attribuant *a priori* aux traits, un certain niveau d'adaptation et de capacité à évoluer, l'écologie évolutive considère ces deux points comme des hypothèses à tester. Des connaissances sur l'histoire des espèces, leurs relations de parenté, leur niveau de divergence ou de coévolution, sont bien souvent nécessaires pour tester de telles hypothèses.

C'est précisément l'étude des adaptations et de leur rôle dans la structuration des communautés qui ont motivé mes premiers travaux durant cette thèse. En s'appuyant sur la pauvreté des communautés de frugivores à Madagascar, l'objectif était d'utiliser le système plantes-frugivores pour comprendre les stratégies de signalisation, principalement les signaux visuels mis en jeu dans des réseaux d'interactions polyspécifiques. Une première phase de ce travail a été d'acquérir des données sur l'évolution à la fois de la couleur des fruits et des systèmes visuels de perception des couleurs chez les animaux, ce qui devait passer par la connaissance des relations de parenté entre les plantes et entre les animaux étudiés. Malheureusement, bien qu'un cadre théorique sur l'évolution des signaux visuels dans des réseaux d'interactions multiples ait pu être développé et formalisé, des troubles politiques sur Madagascar m'ont empêché d'effectuer une seconde mission de terrain nécessaire à l'acquisition d'un nombre suffisant de données pour tester mes modèles théoriques. Les résultats de mes travaux sur l'évolution des signaux sont donc encore trop préliminaires pour être présentés dans ce document. Ils devraient cependant paraître prochainement. L'article S1

présenté en Annexe, qui correspond à une réponse à un article publié sur l'évolution de la couleur des poussins en réponse à la sélection par le système visuel des parents, est issu de cette partie de mes recherches.

De ces premiers travaux est resté un questionnement sur le choix des marqueurs moléculaires utilisés pour définir des patrons phylogénétiques et phylogéographiques pour des espèces ou groupes d'espèces chez lesquelles des processus écologiques, démographiques, ou évolutifs sont étudiés. Le choix de tels marqueurs ne peut s'opérer au hasard car les marqueurs ne doivent pas être influencés par la sélection, et ceci d'autant plus que la sélection pourrait porter sur les traits impliqués dans les processus biologiques étudiés. Les caractères morphologiques, qui sont généralement considérés comme n'étant pas neutre, ne pouvaient donc *a priori* répondre à de telles attentes. En revanche, plusieurs marqueurs moléculaires sont reconnus pour leur neutralité et donc leur intérêt en reconstruction phylogénétique et phylogéographique. C'est notamment le cas des marqueurs cytoplasmiques, issus des génomes mitochondriaux et chloroplastiques. Plus particulièrement, les marqueurs mitochondriaux ont été (Avice 2000) et continuent d'être considérés (Zink & Barrowclough 2008) comme des outils de premier choix pour la reconstruction de patrons phylogéographiques et phylogénétiques entre espèces proches. En revanche, l'emploi de ces marqueurs cytoplasmiques est généralement considéré comme d'un intérêt limité pour l'étude des processus démographiques, évolutifs et écologiques, c'est-à-dire l'étude des mécanismes à l'origine des patrons observés (Zink & Barrowclough 2008). Cette dichotomie de l'utilisation des marqueurs cytoplasmiques, selon qu'elle vise l'étude des patrons ou l'étude des processus, est reconnue et employée par un grand nombre d'auteurs. Cependant, d'autres auteurs considèrent que certaines caractéristiques – que j'appellerai par la suite *évolutives* car elles influencent directement la génétique des populations – des génomes mitochondriaux et chloroplastiques rendent, à l'inverse, les marqueurs issus de ces génomes particulièrement inappropriés à l'étude de patrons phylogéographiques et phylogénétiques.

Ce document présente une partie de mes travaux sur l'utilisation des marqueurs cytoplasmiques dans le but d'étudier les patrons phylogénétiques et phylogéographiques ainsi que les processus démographiques, évolutifs et écologiques. Nous montrerons que la fréquence des discordances entre l'histoire des génomes cytoplasmiques et l'histoire des espèces rend inappropriée l'utilisation des marqueurs cytoplasmiques pour la reconstruction des patrons, au contraire des marqueurs nucléaires et morphologiques. En revanche, nous montrerons que ces mêmes discordances peuvent être exploitées pour l'étude des processus.

Le chapitre I présente une synthèse des connaissances sur l'origine des génomes cytoplasmiques ainsi que sur leurs caractéristiques. Ce chapitre est intégralement bibliographique. J'ai choisi d'inclure ce chapitre afin de conserver l'esprit de l'écologie évolutive, qui est d'utiliser les connaissances sur le passé pour comprendre les processus contemporains. Le passé fait ici référence à la théorie endosymbiotique expliquant l'origine des mitochondries et des chloroplastes, tandis que les processus contemporains sont les caractéristiques architecturales et évolutives des génomes de ces organites. Le chapitre II explique comment ces caractéristiques ont fait des génomes cytoplasmiques des marqueurs populaires pour reconstruire des patrons phylogénétiques et phylogéographiques. Dans le chapitre III, nous introduisons la notion de *discordance*, c'est-à-dire le fait que les génomes cytoplasmiques et les espèces peuvent suivre des voies évolutives différentes. Ces discordances constituent un frein à l'emploi des génomes cytoplasmiques dans la reconstruction des patrons. Nous présentons les différentes méthodes permettant de révéler ces discordances. Le chapitre IV présente les causes de discordance et discute les méthodes permettant d'identifier ces causes. Enfin, le chapitre V propose que les marqueurs cytoplasmiques puissent être utilisés dans le but d'étudier des processus démographiques, évolutifs et écologiques. Les chapitres II à V mélangent des synthèses bibliographiques sur des thèmes précis et des illustrations issues de travaux personnels. Ces travaux ont fait l'objet de trois manuscrits rédigés en anglais, consultables à la fin du document.

Chapitre I

COMPARTIMENTS GENOMIQUES DES CELLULES VIVANTES

I.A. Présentation et origine des organites intracellulaires

L'ADN est probablement né il y a plus de 3 milliards d'années, en liaison avec l'apparition des protéines nécessaires à la production et à l'assemblage de ses précurseurs. L'ARN codant pour ces protéines fut alors remplacé par cette nouvelle molécule, plus fiable dans son rôle de stockage de l'information (Eigen & Schuster 1977). Le développement du génome qui s'en suivit ouvrit la voie de la synthèse membranaire (Koonin & Martin 2005 ; mais voir aussi Forterre 2005), et marqua le début de la vie cellulaire. Par ailleurs, l'apparition de la membrane cellulaire marqua un tournant dans l'évolution du vivant, la séquestration de produits codés par le génome donnant naissance à l'individualité et permettant d'augmenter considérablement l'efficacité de la sélection naturelle (Lynch 2006).

Les organismes cellulaires sont aujourd'hui répartis dans trois lignées, les Eubacteria, les Eucaryotes et les Archea. L'origine de ces trois lignées reste largement discutée et seule celle des Eucaryotes, lignée à laquelle appartiennent les taxons étudiés dans ce document, sera brièvement développée. Plusieurs hypothèses sont donc avancées pour expliquer leur origine. L'hypothèse généralement acceptée est que le génome nucléaire de la souche Eucaryote serait un produit chimérique d'une archée et d'une eubactérie (Sogin 1991; Zillig 1991). Plus qu'une simple fusion entre un représentant de chacune de ces deux lignées, l'émergence de la lignée eucaryote pourrait résulter de transferts horizontaux de gènes provenant de contributeurs multiples (Lester *et al.* 2006). D'autres hypothèses alternatives, plus ou moins recevables, ont été avancées pour expliquer l'origine de la cellule eucaryote ; nous n'en feront pas écho.

L'origine des organites des cellules eucaryotes est beaucoup mieux établie que l'origine des cellules eucaryotes elles-mêmes. L'article clef de Lynn Margulis (nommée Sagan à l'époque ; Sagan 1967) permit de faire le point et de remettre à jour l'ancienne

théorie symbiotique de l'origine des mitochondries et des plastes. Des analyses portant sur les gènes mitochondriaux, leur organisation génomique et leur distribution, ont par la suite confirmé les arguments de Margulis et indiquent que les génomes mitochondriaux dérivent d'un ancêtre commun aux α -proteobactéries (Burger *et al.* 1999). L'endosymbiose serait probablement issue d'une invasion unique d'un hôte, appartenant soit à la lignée des Archées (la symbiose ante-daterait ainsi la naissance des eucaryotes), soit à la lignée des eucaryotes, dans les deux cas il y a au moins 1,5 milliard d'années (Ga) (Embley *et al.* 2003). Concernant les plastes, des fossiles de type algues rouges ont confirmé que des eucaryotes multicellulaires portant des plastes existaient il y a 1,2 Ga, tandis que des arguments biochimiques soutiennent l'apparition d'organismes proches des algues actuelles il y a 1,5 Ga (Martin & Russell 2003). Des analyses phylogénétiques, biochimiques et structurales ont pu montrer qu'une unique association symbiotique entre une cyanobactérie phagocytée par un eucaryote hébergeant des mitochondries, il y a 1,2 à 1,5 Ga, a donné naissance aux plastes primitifs dont on retrouve aujourd'hui les descendants dans les algues, les plantes et les glaucophytes (Wetzel *et al.* 2003). De manière remarquable, des plastes se seraient par ailleurs développés suite à une seconde endosymbiose (dite « endosymbiose secondaire ») durant laquelle un eucaryote photosynthétique (de type algue) aurait été phagocyté par un eucaryote non photosynthétique (Muller 1998). Le noyau et le cytoplasme de la cellule phagocytée auraient dégénéré pour donner lieu à une cellule avec un chloroplaste à quatre membranes : deux issues de l'endosymbiose primaire, la membrane cytoplasmique de la cellule phagocytée et la membrane invaginée de la cellule qui phagocyte. Ces chloroplastes sont aujourd'hui présents chez les haptophytes, les heterokontes, les cryptophytes et les chlorarachniophytes. Suite à une réduction du nombre de membranes, on retrouve enfin des chloroplastes à trois membranes chez certains Dinophytes.

Le mode de vie endosymbiotique des organites qui suivit fut responsable de profondes modifications au sein même de leur génome. Notamment, la transition d'un symbiote autonome vers un organite dépendant a entraîné une réduction du génome, avec diminution du nombre de fonctions assurées par ces organites (Dyall *et al.* 2004). De nombreux gènes des endosymbiotes ont donc été perdus et une partie de ceux qui ont été conservés ont été transférés vers le noyau (Gabaldon & Huynen 2003). La phase de transfert nécessita probablement une étape durant laquelle des copies du gène ou du groupe de gènes étaient présentes à la fois dans le noyau et dans l'organite, avant de disparaître de ce dernier. Il est supposé que chez les vertébrés, environ 4000 de ces gènes ont ainsi été transférés de la mitochondrie vers le noyau (Richly *et al.* 2003). Chez les plantes, environ 20% du génome

nucléaire semble provenir du génome chloroplastique initial (Martin *et al.* 2002). Ce phénomène de transfert n'est pas limité à des événements anciens puisque des transferts récents, fréquents et fonctionnels ont été démontrés pour des gènes mitochondriaux (Adams & Palmer 2003) et chloroplastiques (Millen *et al.* 2001) d'angiospermes. Par ailleurs, presque tous les génomes nucléaires possèdent de nombreuses insertions très semblables à des gènes contemporains de l'ADN mitochondrial (ANDmt) (Adams & Palmer 2003; Leister 2005).

Le résultat de ces pertes et transferts se traduit par des génomes cytoplasmiques de taille réduite, avec environ 13 gènes codant pour des protéines dans les génomes mitochondriaux animaux (Gray *et al.* 1999), entre 29 et 59 gènes pour les génomes mitochondriaux des plantes terrestres (Ward *et al.* 1981) et entre 50 et 200 gènes pour les génomes chloroplastiques de ces mêmes plantes (Glockner *et al.* 2000). En terme de taille, les génomes mitochondriaux animaux se situent entre 14 et 20 kb (Gray *et al.* 1998) tandis que ceux des plantes font entre 180 et 1600 kb (Ward *et al.* 1981), et ceux des chloroplastes entre 37 et 220 kb. Ces nombres permettent par ailleurs de constater que la taille d'un génome n'est pas liée au nombre de gènes qu'il abrite. Ceci provient de l'énorme variation dans la quantité d'ADN non codant présente entre les trois types des génomes. Tandis que la plupart des mitochondries animales sont composées de 90-95% d'ADN codant, l'ADNmt des plantes terrestres contient plus de 90% d'ADN non-codant. Plus précisément, l'ADNmt des animaux ne possède aucun intron, excepté chez deux espèces de Cnidaires (Pont-Kingdon *et al.* 1998; van Oppen *et al.* 2002) et une espèce de Placozoaire (Dellaporta *et al.* 2006). L'ADN chloroplastique (ADNcp) occupe une position intermédiaire, renfermant entre 30 et 40% d'ADN non-codant. Comme on le verra plus tard, ces différences dans l'architecture génomique pourraient en partie s'expliquer par des différences dans les taux de mutation.

I.B. Les génomes cytoplasmiques : caractéristiques évolutives

L'objet de la section précédente a été de présenter sommairement l'origine des organites, et de comprendre à partir de cette origine l'architecture de leur génome. Les différences d'architecture génomique entre noyaux, chloroplastes et mitochondries, mais également entre mitochondries animales et végétales, seront exploitées afin de retracer au mieux l'histoire des espèces abritant ces organites comme on le verra dans le Chapitre II. Toutefois, d'autres caractéristiques singularisant les génomes cytoplasmiques par rapport aux génomes nucléaires

auront un impact encore plus important sur l'évolution des organites. Ces caractéristiques sont ici dénommées « caractéristiques évolutives », en raison de leur importance dans les analyses génétiques portant sur l'évolution des organites. Ces caractéristiques sont la transmission maternelle, l'absence de recombinaison et une rapidité d'évolution.

I.B.1. La transmission maternelle

Une des caractéristiques les mieux connues du matériel génétique cytoplasmique est sa transmission monoparentale. Baur (1909) et Correns (1909) furent les premiers à mettre en évidence chez le Pélargonium zonal (*Pelargonium zonale*) et la Belle de nuit (*Mirabilis jalapa*), respectivement, que l'héritabilité du matériel génétique des chloroplastes s'écartait des lois mendéliennes. Cet écart à la loi de Mendel se manifestait par deux observations : la ségrégation rapide des allèles durant la reproduction végétative et la transmission du matériel génétique par un seul parent. Ce ne fut que bien plus tard, dans les années 1950, que le mode de transmission de la mitochondrie fut étudié, grâce à des travaux sur les levures de boulanger (Ephrussi 1953). L'ADN mitochondrial ne fut formellement identifié et localisé que dans les années 60 (voir la revue de Mounolou & Lacroute 2005), et la première preuve directe de transmission monoparentale et maternelle ne fut disponible qu'en 1972 (Dawid & Blackler 1972), chez le xénope *Xenopus sp.*

Les mécanismes à l'origine de l'héritabilité monoparentale des organites sont variés (détaillés dans Birky 1995). Les mécanismes prézygotiques éliminent les organites ou le génome des organites durant la gamétogénèse. Chez certains organismes, la méiose donne lieu à des gamètes identiques (isogamie) tandis que chez d'autres, elle produit de grands gamètes, femelles, et de petits gamètes, mâles (anisogamie). Une conséquence fréquente de cette division et/ou croissance cellulaire inégale est un biais marqué en faveur de la présence d'organites d'origine maternelle dans l'œuf fertilisé. Dans des cas extrêmes, des organites peuvent être complètement exclus des gamètes d'un des deux sexes par cytokinèse inégale (Mose 1961). Les organites peuvent également être détruits une fois dans les gamètes, comme c'est le cas dans les isogamètes de l'algue verte filamenteuse *Temnogyra*. D'autres mécanismes éliminent les organites pendant la fécondation. Chez certains organismes (comme les Tuniciers, Ursprung & Schabtaç 1965), les organites du sperme ne peuvent pénétrer dans l'œuf et parfois même, seuls les noyaux sont transférés durant la reproduction sexuée (comme durant la conjugaison des protozoaires ciliés). Certains mécanismes

zygotiques déterministes éliminent les organites de l'embryon. Un exemple frappant est la dégradation des chloroplastes issus des gamètes mâles chez les algues vertes filamenteuses des genres *Spyrogyra* et *Zygnema*. Chez *Chlamydomonas*, c'est l'ADN chloroplastique issu d'un des deux gamètes (le mt^+) qui subit une dégradation enzymatique. Enfin, une dernière catégorie conduisant à la transmission monoparentale des organites correspond aux mécanismes zygotiques stochastiques. Un de ces mécanismes est la distribution intégrale et aléatoire des organites des deux parents soit dans les cellules embryonnaires soit dans les cellules extra-embryonnaires, lors de la ségrégation de ces deux types cellulaires. Ceci est observé dans les croisements entre *Pelargonium* blancs et verts durant lesquels certains embryons très jeunes ne contiennent aucun chloroplaste, au contraire des cellules extra-embryonnaires (Tilney-Bassett 1970).

Les raisons pour lesquelles de tels mécanismes ont évolué restent relativement obscures. Bromham *et al.* (2003) a fait la revue de deux explications possibles pour les mitochondries. La première, la plus populaire, serait la prévention d'un comportement consanguin lié à la compétition entre ADN mitochondriaux (Hastings 1992) pouvant conduire à des conflits génomiques potentiellement létaux (Hurst & Hamilton 1992). Selon la seconde, il pourrait s'agir d'une adaptation à l'anisogamie qui permettrait de prévenir l'intrusion dans la cellule-œuf d'un ADN mitochondrial spermatique endommagé par une intense activité de respiration (Allen 1996).

La transmission monoparentale semble être le principal type d'héritabilité chloroplastique chez les Angiospermes. Toutefois, les travaux de Baur (1909) ayant permis de mettre en évidence l'existence d'un mode de transmission monoparentale pour le génome chloroplastique ont par la même occasion démontré que ce processus n'était pas exclusif, dans la mesure où certains *Pelargonium* issus de ces croisements possédaient des organites d'origine paternelle ou bien issus des deux parents.

Concernant la transmission entre générations des mitochondries chez les animaux, le concept de transmission strictement maternelle n'a cessé d'être populaire depuis sa découverte en 1972. Ce paradigme a été soutenu par de nombreux travaux n'ayant pu mettre en évidence une quelconque transmission paternelle (Hutchison *et al.* 1974; Hayashi *et al.* 1978; Kroon *et al.* 1978; Avise *et al.* 1979; Francisco *et al.* 1979; Reilly & Thomas 1980). Des observations plus récentes portant sur l'importance des processus stochastiques sont venues soutenir ces premiers travaux. Ainsi, dans les rares cas de transmission paternelle, la très faible proportion d'ADNmt paternel par rapport à l'ADNmt maternel rendrait le premier particulièrement sensible à la perte par dérive génétique. De plus, les effets de la dérive des

lignées génétiques paternelles pourraient être accentués par le goulot d'étranglement que les mitochondries subissent aux stades précoces du développement (Bergstrom & Pritchard 1998; Birky 2001; Shoubridge & Wai 2007; Wolff & Gemmell 2008). Toutefois, depuis maintenant deux décennies, un nombre grandissant d'exemples est venu apporter des preuves concrètes d'hétéroplasmie et de transmission à la descendance d'ADN mitochondrial. Des cas d'héritabilité bi-parentale ont été documentés chez des mammifères, oiseaux, reptiles, poissons, mollusques, nématodes et arthropodes. L'hétéroplasmie serait par ailleurs la norme chez certains bivalves. Une telle hérédité paternelle pourrait somme toute être un phénomène commun, mais sa contribution pourrait être en dessous des limites actuelles de détection (Wolff & Gemmell 2008). Par ailleurs, l'apparente prépondérance de la transmission strictement monoparentale pourrait provenir d'un échantillonnage trop faible dans les expériences ayant étudié le mode de transmission des organites (Birky 2001). Par exemple, si l'on suppose que la transmission paternelle de gènes mitochondriaux est telle que 1% de toutes les mitochondries de la progéniture proviennent du père, cela signifierait qu'environ 1% des descendants bénéficieraient du génotype paternel et qu'un échantillonnage d'environ 300 descendants sans aucun génome paternel serait nécessaire pour démontrer qu'il y a moins de 1% de transmission paternelle (Milligan 1992). Enfin, nous citerons l'exemple anecdotique mais bien documenté de transmission d'ADNmt paternel découverte chez un jeune homme de 28 ans souffrant de myopathie mitochondriale, causé par un ADNmt paternel délétère. L'ADNmt paternel était présent dans les cellules musculaires du patient avec un ratio de 10 génomes mitochondriaux paternels pour 1 maternel (Schwartz & Vissing 2002).

Bien que l'héritabilité paternelle ne soit pas le mode de transmission le plus fréquent des organites cellulaires, il est donc maintenant reconnu que l'héritabilité biparentale et l'hétéroplasmie qui en découlent sont des phénomènes récurrents chez les plantes et les vertébrés. Toutefois, pour influencer l'évolution d'une population d'organites, l'ADNmt paternel doit (i) proliférer considérablement durant l'embryogénèse et (ii) être transmis aux générations futures *via* les lignées germinales ou rester intact dans le cytoplasme de l'oocyte suffisamment longtemps pour que des recombinaisons avec le génome maternel puisse se produire et que les produits qui en dérivent atteignent la lignée germinale. Ces trois dernières années ont vu l'apparition de travaux démontrant que ces deux critères pouvaient s'appliquer à la dynamique de transmission de l'ADNmt paternel chez les insectes (voir par exemple Sherengul *et al.* 2006 et Fontaine *et al.* 2007). Ces travaux sont en accord avec un travail plus ancien (Kondo *et al.* 1990) et confirment que la lignée d'ADNmt paternel peut proliférer dans le soma et la lignée germinale des insectes. Ceci pourrait être dû à un goulot d'étranglement

mitochondrial relativement faible chez les insectes lié à un effectif efficace fort de l'ADNmt ou une division cellulaire moins intense durant l'oogénèse (Ashley *et al.* 1989). La question est désormais de savoir si chez des organismes ayant un faible effectif efficace de l'ADNmt, comme chez les Vertébrés, l'ADNmt paternel peut atteindre les lignées germinales à une fréquence suffisamment élevée, et s'il peut y persister suffisamment longtemps pour influencer l'évolution de la population d'organites. Il apparaît toutefois évident que si une recombinaison avec l'ADNmt de l'autre sexe pouvait avoir lieu, ce mécanisme faciliterait grandement la persistance de gènes mitochondriaux paternels dans les générations suivantes.

Chez les plantes, les démonstrations de transmission biparentale d'ADNmt sont anecdotiques ce qui semble signifier que ce phénomène est nettement plus rare que pour les animaux (Mogensen 1996; Havey 1997; Vaillancourt *et al.* 2004).

Enfin, nous noterons que les mécanismes contrôlant la transmission de l'ADN cytoplasmique peuvent varier en fonction du niveau taxonomique auquel les croisements se produisent. Ainsi, Cruzan *et al.* (1993) ont pu observer une transmission d'ADNcp d'origine paternelle dans des croisements interspécifiques entre *Iris fulva* et *I. hexagona* chez lesquels, normalement, seul l'ADNcp maternel est transmis. De manière plus frappante, Hansen *et al.* (2007) a mis en évidence une transmission maternelle d'ADNcp dans les croisements intraspécifiques et une transmission paternelle dans les croisements interspécifiques chez *Passiflora*. Ces études suggèrent que les mécanismes prévenant la fuite d'ADNcp paternel sont fragilisés dans les cas d'hybridation. Par ailleurs, Corriveau & Coleman (1988) ont supposé que jusqu'à un tiers des espèces d'Angiospermes pouvait subir de la transmission biparentale occasionnelle d'ADNcp par l'intermédiaire de pré-plastes présents dans le pollen.

I.B.2. L'absence de recombinaison

Une deuxième caractéristique de l'ADNmt est son incapacité à recombiner entre molécules. Plusieurs raisons expliquent cette propriété (White *et al.* 2008). Premièrement, pour qu'un événement de recombinaison soit détectable, il faut simultanément que : (i) deux molécules différentes d'ADNmt soient présentes dans la même cellule, (ii) les molécules soient en contact physique suffisant pour permettre la réalisation d'un crossing-over, mécanisme qui ne fait pas partie du cycle de vie normal d'une mitochondrie et (iii) que soit présentes les

enzymes codées par l'ADNn qui permettent la fusion, l'échange et la réparation du génome (Thyagarajan *et al.* 1996; Lakshmipathy & Campbell 1999; Yaffe 1999). Deuxièmement, pour que la molécule recombinante reste intacte, il faut que deux crossing-over se produisent. Enfin, l'alignement des régions homologues peut être rendu difficile à cause de la nature circulaire de l'ADNmt.

La recombinaison étant un mécanisme évolutif important qui permet de réduire le nombre de mutations délétères d'une population et de créer de nouvelles combinaisons d'allèles favorables (Hurst & Peck 1996), l'ADNmt utilise donc d'autres mécanismes évolutifs pour assurer le maintien de la valeur sélective de ses populations chez les individus. Par exemple, Bergstrom & Pritchard (1998) ont montré, à l'aide d'un modèle de génétique des populations, que le goulot d'étranglement mitochondrial augmente l'efficacité de la sélection contre les allèles délétères en augmentant la variance de la valeur sélective parmi les individus de la population.

Le paradigme portant sur l'absence de recombinaison de l'ADNmt fut fragilisé à partir de la fin des années 1990 par des arguments suggérant l'existence d'évènements de recombinaisons à la fois intra- et intermoléculaires chez les animaux (Thyagarajan *et al.* 1996; Lunt & Hyman 1997; Lakshmipathy & Campbell 1999; Tang *et al.* 2000). De plus, en 2004 et 2005, deux équipes ont analysé la distribution du polymorphisme nucléotidique dans les séquences d'ADNmt des animaux et ont apporté un soutien statistique à l'existence de recombinaison mitochondriale à faible fréquence à travers l'ensemble du règne animal (Piganeau *et al.* 2004; Tsaousis *et al.* 2005). Deux nouvelles études ont récemment fourni des preuves convaincantes de recombinaison chez deux espèces de vertébrés : le saumon de l'Atlantique (Ciborowski *et al.* 2007) et un lézard australien *Chlamydosaurus kingii* (Ujvari *et al.* 2007).

S'il y a si peu de preuves empiriques de recombinaison mitochondriale, c'est que sa mise en évidence reste difficile (White *et al.* 2008). Sans prolifération, la fréquence des évènements de recombinaison est probablement trop faible par rapport aux limites de détectabilité des technologies standards (Loeb 2001; Kmiec *et al.* 2006).

La question qui intéresse l'évolutionniste est de savoir dans quelle mesure ces évènements de recombinaison sont susceptibles de modifier ses conclusions. Des simulations ont montré que la recombinaison a vraisemblablement un impact limité sur la capacité qu'ont les algorithmes de phylogénie à reconstruire un arbre tant au niveau spécifique qu'au niveau infra-spécifique (Schierup & Hein 2000; Posada & Crandall 2002). En revanche, nier l'existence d'évènements de recombinaison peut conduire à une vision erronée des

événements de coalescence (Schierup & Hein 2000). Les conséquences potentielles sont : une mauvaise estimation de paramètres comme le temps de divergence, une vision erronée des patrons d'expansion démographique (Beerli & Felsenstein 1999; Innan & Nordborg 2002), une surestimation du nombre de mutations, ou encore un rejet à tort de l'hypothèse d'horloge moléculaire constante (Schierup & Hein 2000; Posada & Crandall 2002).

Nous retiendrons finalement que la recombinaison mitochondriale chez les animaux est un phénomène sous-estimé, et qu'il ne serait pas inapproprié à l'avenir d'utiliser des modèles d'évolution moléculaires jusqu'alors usités dans des analyses d'ADNn. Ces modèles ne nécessitent pas la reconstruction d'un arbre des gènes, ce qui permet la prise en compte de la recombinaison, de la sélection et des processus démographiques (voir par exemple Rannala & Bertorelle 2001).

Chez les plantes, la recombinaison à l'intérieur du génome de l'ADNmt est la règle. Cependant, il existe également un nombre grandissant de cas pour lesquels la recombinaison entre génomes d'ADNmt a été démontrée (Lonsdale *et al.* 1988; Birky 2001; Stadler & Delph 2002; Bergthorsson *et al.* 2003). Bien que ce phénomène semble plus fréquent qu'on ne le croyait il y a peu, sa fréquence précise reste encore inconnue. Les cas de recombinaison d'ADNcp semblent, eux, très marginaux. La recombinaison chloroplastique a été observée pour la première fois chez des algues unicellulaires où la fusion chloroplastique est très rapide (Cavalier 1970). Par la suite, la recombinaison chloroplastique a été évoquée pour expliquer de rares réarrangements structuraux de l'ADNcp chez les Poaceae (Hiratsuka *et al.* 1989; Ogihara *et al.* 1992). Enfin, ce phénomène a pu être observé chez des hybrides expérimentaux de Solanaceae (Trabelsi *et al.* 2005). La question qui reste à traiter est de savoir dans quelle mesure la recombinaison chloroplastique se produit en conditions naturelles chez les plantes à fleurs.

Comme signalé précédemment, l'hybridation entre espèces proches est susceptible d'affaiblir les mécanismes prévenant la transmission d'ADNcp paternel. Les zones hybrides sont donc particulièrement propices à l'hétéroplasmie et donc à la recombinaison entre molécules différentes d'ADNcp. La détection de recombinants d'ADNcp dans la zone hybride de *Pinus contorta* et *P. banksiana* pourrait s'inscrire dans ce contexte (Govindaraju *et al.* 1989). Par ailleurs, des études récentes ont montré que la plupart des ADNcp situés dans les cellules méristématiques étaient associés en complexes structuraux renfermant des concatémères de génomes chloroplastiques ayant une forme linéaire (Bendich 2004; Oldenburg & Bendich 2004). Ces auteurs ont suggéré un nouveau modèle de réplication

rendant obsolète l'ancien modèle basé sur la structure classiquement circulaire de l'ADNcp (Kolodner & Tewari 1975). Chez les Angiospermes, une seule étude semble mettre en évidence l'existence d'un évènement de recombinaison chloroplastique (Heuertz *et al.* sous presse). Pris dans leur ensemble, les travaux portant sur la recombinaison chloroplastique tendent à confirmer que le phénomène est anecdotique chez les plantes à fleurs.

I.B.3. Des taux de mutation extrêmes

Chez les animaux, le taux de mutation de l'ADNmt est beaucoup plus élevé que celui de l'ADNn (Brown *et al.* 1979; Mishmar *et al.* 2003; Kivisild *et al.* 2006). Par exemple, les taux de mutation mitochondriale estimés d'après la phylogénie des Bilatériens sont d'environ 9 à 25 fois supérieurs à ceux estimés pour les génomes nucléaires de ces mêmes lignées (Lynch 2006). Par ailleurs, ces estimations sont probablement des sous-estimations dans la mesure où les seules mesures directes du taux de mutation mitochondriale chez les animaux donnent des valeurs de l'ordre de 10 fois supérieures à celles estimées à partir des études phylogénétiques. Ceci provient sans doute du fait que les mutations silencieuses ne sont pas complètement neutres (Denver *et al.* 2000; Howell *et al.* 2003).

Au moins trois facteurs sont fréquemment avancés pour expliquer le taux de mutation élevé de l'ADNmt des animaux (Lynch *et al.* 2006). Tout d'abord, l'activité respiratoire intense des animaux produit une grande quantité de radicaux libres oxydatifs à l'origine d'un milieu intra-mitochondrial à fort potentiel mutagène (Bogehagen 1999). Deuxièmement, l'ADNmt est continuellement répliqué à l'intérieur de cellules ne se divisant pas, et le taux d'erreur dans l'incorporation des bases (avant correction) est environ 10^3 à 10^4 fois supérieur au taux d'erreur observé dans le génome nucléaire (Johnson & Johnson 2001). Enfin, les gènes codant pour des protéines de réparation de l'ADN ont été transférés vers le noyau dans la majorité des lignées et peu de génomes mitochondriaux actuels codent pour de telles protéines. Le mécanisme de réparation par excision pourrait par ailleurs être entièrement perdu pour l'ADNmt, tandis que la réparation de mauvais appariements est considérablement réduite chez les mammifères comparée aux levures (Gillespie 2000).

Si le taux de mutation mitochondriale est constamment plus élevé que le taux de mutation nucléaire chez les animaux, il existe une importante variation au sein même de ce groupe (Martin *et al.* 1992; Martin & Palumbi 1993; Rand 1994; Mindell *et al.* 1996). Cette variation tend à rendre obsolète le fameux « 2% par millions d'années » utilisé pour calibrer

l'horloge moléculaire, proposé il y a trois décennies pour l'ADNmt de primates (Brown *et al.* 1979) et encore fréquemment utilisé pour diverses études portant sur les mammifères et même sur les oiseaux (voir par exemple Garcia-Moreno 2004, cité dans Nabholz *et al.* 2008). Plusieurs théories ont été proposées pour expliquer la variation du taux de mutation de l'ADNmt entre les différentes lignées animales. Premièrement, l'hypothèse du temps de génération propose que les espèces présentant un temps de génération court subissent un nombre plus élevé de réplifications d'ADN par an et ont donc un nombre d'erreurs par unité de temps plus élevé (Li *et al.* 1987; Ohta 1993). Deuxièmement, l'hypothèse du taux métabolique propose de relier le taux de mutation de l'ADNmt à la production de radicaux libres mutagènes (Martin *et al.* 1992; Martin & Palumbi 1993). Cette hypothèse a été mise en avant avec la découverte d'une relation entre le taux métabolique poids-spécifique, qui varie en fonction du poids et de la température, et le taux de substitutions mitochondriales (Gillooly *et al.* 2005). Enfin, l'hypothèse de la sénescence suggère une autre relation entre le taux de mutation de l'ADNmt et la production de radicaux libres. Ici, les radicaux libres produits tout au long de la vie d'un organisme contribuent, et sont même la cause, de sa sénescence. Cette théorie a trouvé un soutien dans les travaux de Barja & Herrero (2000) démontrant une relation entre la concentration d'un marqueur de dommages oxydatifs et l'espérance de vie maximale chez les mammifères. Des mutations somatiques affectant l'efficacité des protéines respiratoires, résulterait une production plus importante de radicaux libres qui, en retour, augmenteraient le taux de mutation. Ce rétrocontrôle positif conduirait finalement à la sénescence (Kujoth *et al.* 2007). Bien que ces différentes hypothèses ne soient pas exclusives, les travaux récemment publiés de Nabholz *et al.* (2008) apportent un soutien convaincant à la théorie de la sénescence en montrant une relation négative entre la longévité et le taux de mutation de l'ADNmt chez les mammifères.

Le taux de mutation de l'ADN mitochondrial des plantes est considérablement plus faible que celui de l'ADNmt des animaux, et seulement 0,05 fois celui de l'ADNn dans la plupart des lignées végétales (Lynch *et al.* 2006). Cette caractéristique s'applique à l'ensemble du génome mitochondrial (Palmer & Herbon 1988). En outre, les gènes mitochondriaux des plantes évoluent entre 50 et 100 fois plus lentement que leur homologues chez les animaux (Cho *et al.* 2004). Le fait que les eucaryotes unicellulaires possèdent un taux de mutation mitochondriale similaire au taux de mutation nucléaire suggère que les animaux et les plantes ont respectivement acquis leurs taux très élevés et très faibles au cours de l'évolution, et non que l'une de ces deux lignées ait conservé un état ancestral. Toutefois, et

contrairement aux animaux, les facteurs à l'origine de la diminution du taux de mutation de l'ADNmt des plantes restent largement inconnus.

Chez les plantes, les taux de mutation mesurés pour l'ADNcp sont environ deux à quatre fois plus élevés que ceux mesurés pour l'ADNmt, et environ 0,1 fois ceux mesurés pour les génomes nucléaires (Wolfe *et al.* 1987; Gaut *et al.* 1996). Le faible taux de mutation observé dans l'ADNcp, en comparaison à celui observé dans l'ADNn, n'a pas été expliqué de manière convaincante jusqu'à présent. L'absence de recombinaison rendant plus efficaces les balayages sélectifs sur les génomes cytoplasmiques, ce phénomène a été proposé comme une explication alternative au taux d'évolution lent pour expliquer le faible polymorphisme des gènes cytoplasmiques (Maruyama & Birky 1991). Cette explication est peu convaincante à elle seule dans la mesure où elle devrait également s'appliquer au génome mitochondrial des animaux. Notons par ailleurs qu'il existe une importante variation du taux de mutation au sein même du génome chloroplastique (voir Chapitre II). Le taux de mutation semble également varier au sein des Angiospermes. Ainsi, les plantes annuelles semblent évoluer plus rapidement que les plantes pérennes, et ce principalement aux sites non-synonymes (Bousquet *et al.* 1992; Petit & Hampe 2006). Au sein des monocotylédones, les herbes évoluent plus rapidement (3,7 fois plus rapidement pour le marqueur *rbcL*) que les palmiers aux sites silencieux (EyreWalker & Gaut 1997).

Un point intéressant soulevé par Lynch *et al.* (2006) concerne la relation entre le taux de mutation des organites et l'architecture de leur génome. Comme signalé précédemment, le génome mitochondrial des plantes contient une grande quantité d'ADN non-codant tout comme le génome nucléaire, et contrairement au génome mitochondrial des animaux qui lui n'en contient aucun. Le génome chloroplastique se situe en position intermédiaire entre les génomes mitochondriaux animaux et végétaux, avec la présence de quelques éléments non-codants. Enfin, les seuls génomes mitochondriaux animaux possédant des introns – ceux des Cnidaires – ont par ailleurs un taux de mutation tellement faible que le polymorphisme nucléotidique intra-spécifique est virtuellement inobservable (Shearer *et al.* 2002). Il est donc frappant d'observer une corrélation négative entre le taux de mutation et la quantité d'éléments non-codants dans le génome. Selon Lynch et ses collaborateurs, le lien se trouve dans le fait que l'ADN non-codant augmente la vulnérabilité de l'ensemble du génome face aux mutations. Par exemple, des mutations sur les introns peuvent altérer la spécificité de certains de leurs sites nécessaires au processus d'épissage, et augmentent ainsi les chances de

rendre non fonctionnel le gène qui les accueille (Lynch 2002). Ainsi, alors que la variation dans l'architecture de l'ADN nucléaire pourrait principalement être gouvernée par la variation dans la taille efficace (Lynch & Conery 2003; Lynch 2006), la variation dans le taux de mutation serait un facteur majeur expliquant la variation dans l'architecture de l'ADNmt. De manière plus générale, ceci implique que l'évolution de l'architecture des génomes serait gouvernée par des forces non-adaptatives.

Chapitre II

Exploitation des caractéristiques architecturales et évolutives des génomes cytoplasmiques pour retracer l'histoire du vivant

La diversité de l'architecture et des taux d'évolution des différents génomes cellulaires offre de multiples fenêtres temporelles permettant d'étudier des événements plus ou moins anciens sur l'arbre du vivant. Par la suite, nous appellerons « marqueur moléculaire » un fragment d'ADN dont la variation est utilisée pour reconstruire l'histoire du vivant. Des marqueurs moléculaires issus de compartiments génomiques différents, et ayant donc des vitesses d'évolution différentes, pourront ainsi être choisis en fonction de la nature des relations de parenté investiguées : basales dans le cas des études de phylogénie (par exemple entre espèces, genres ou familles) ou terminales dans le cas des études de phylogéographie (entre individus d'une espèce). Le chapitre qui suit présente les principales utilisations des marqueurs issus des différents génomes et souligne le lien entre les caractéristiques architecturales et évolutives qui sont à l'origine de ces utilisations. Un exemple d'utilisation de l'ADNmt et un exemple d'utilisation de l'ADNcp sont finalement donnés afin d'illustrer les considérations générales.

II. A. L'ADNmt animal

II.A.1. L'ADNmt animal utilisé comme marqueur pour les analyses de phylogénie et de phylogéographie

L'ADNmt est la molécule la plus étudiée pour retracer l'histoire des espèces et des populations. Plusieurs facteurs, allant de la facilité d'obtention de l'information à la facilité d'analyse, en passant par la nature de cette information, expliquent un tel succès. Nous allons présenter les principaux de ces facteurs, qui sont tous plus ou moins directement liés aux caractéristiques évolutives et architecturales présentées ci-dessus.

A cause de la nature polyploïde de l'ADN nucléaire, le séquençage de fragments de ce génome ne permet pas toujours d'identifier l'état de caractère d'un site donné. Cette identification nécessite le recours au clonage moléculaire ou l'utilisation de logiciels prédisant l'identité d'un allèle en fonction de sa fréquence dans une population (Scheet & Stephens 2006). Le séquençage de l'ADNmt, qui est lui haploïde, génère directement des haplotypes résolus. Le séquençage de l'ADNmt permet donc en général un gain de temps et d'argent en se détournant du clonage, ou un gain d'exactitude en évitant l'estimation probabiliste de l'état de caractère.

Un second avantage provient du fait que les gènes mitochondriaux ne sont en général pas dupliqués au sein du génome mitochondrial. Ainsi, le clonage qui permet d'éviter l'amplification de séquences paralogues chez certains gènes nucléaires n'est pas nécessaire lorsque l'ADNmt est employé comme marqueur. Rappelons toutefois que le transfert d'un gène mitochondrial vers le noyau passe par une étape intermédiaire durant laquelle le gène est dupliqué et est présent dans les deux types de molécules. Suivant le temps qui s'est écoulé depuis l'évènement de duplication, les deux gènes ainsi que leurs séquences peuvent être plus ou moins identiques. Si cette duplication n'est pas reconnue en tant que telle, l'amplification et l'analyse de tels gènes mitochondriaux hébergés dans le noyau peuvent conduire à une reconstruction phylogénétique erronée (Collura *et al.* 1996; Zhang & Hewitt 1996). Ces deux avantages, ajoutés au fait que l'ADNmt soit présent sous de nombreuses copies dans les cellules (Dowling *et al.* 1990), rend aisée l'acquisition d'informations généalogiques à partir de cette molécule.

La nature non-recombinante de l'ADNmt constitue une qualité essentielle de cette molécule en tant que marqueur moléculaire. Premièrement, comme on l'a vu précédemment, cette propriété permet des estimations de paramètres basés sur la coalescence, comme les temps de divergence, ou encore sur les paramètres démographiques, qui ne sont pas biaisées par les phénomènes de recombinaisons. De plus, l'ensemble de l'ADNmt peut, par cette propriété, être considéré comme une unité généalogique à part entière, c'est-à-dire comme un « super-gène » transmis comme un seul bloc au fil des générations (Avice 2004).

Par ailleurs, il est communément reconnu que le couplage d'un mode de transmission essentiellement maternel et d'un état haploïde réduit d'environ un quart la taille efficace par locus des gènes mitochondriaux (N_e) dans une population diploïde, par rapport aux gènes nucléaires (Palumbi *et al.* 2001). Cette propriété joue un rôle majeur dans l'évolution temporelle de la diversité allélique mitochondriale, mais pour comprendre ce point nous devons brièvement présenter la notion de « tri du polymorphisme ancestral ».

Note sur la taille efficace de l'ADNmt

Le généticien américain Sewall Wright définit le concept de taille efficace de population N_e comme étant « le nombre d'individus se reproduisant dans une population idéale, qui aurait la même quantité de dispersion de fréquences alléliques sous une dérive génétique aléatoire, ou la même quantité d'évènements de reproduction consanguins, que la population réelle considérée » (Wright 1931). Si l'on définit N_g comme étant le nombre réel d'unités génétiques se divisant au niveau de population de taille $2N_e$ pour un locus nucléaire (ce qui correspond environ au nombre de femelles pour un organe transmis maternellement), alors ce nombre efficace de gènes mitochondriaux par locus est habituellement considéré comme étant égal au quart du nombre efficace de gènes nucléaires, en raison de la nature haploïde et de la transmission maternelle de l'ADNmt (Palumbi *et al.* 2001). Cette règle dite du « un-quart » est communément employée dans les analyses utilisant la taille efficace des gènes mitochondriaux. C'est d'ailleurs cette loi que nous avons utilisé pour estimer des temps de divergence basés sur un modèle de coalescence (cf. Chapitre V). Toutefois, certains auteurs ont souligné deux problèmes avec cette règle (Lynch *et al.* 2006). Premièrement, elle suppose un niveau identique d'interférence sélective entre les gènes hébergés dans l'ADNmt et entre ceux hébergés dans le noyau. L'absence de recombinaison mitochondriale chez les animaux est une raison pour laquelle cette assertion n'est pas vraie (Berlin & Ellegren 2001). Dans le même esprit, on peut ajouter que les chromosomes nucléaires contiennent de nombreuses cibles potentielles pour des balayages sélectifs contrairement au génome mitochondrial (mais voir Chapitre V). Deuxièmement, la règle du « un-quart » suppose que les mâles et les femelles ont un rôle équivalent dans la production de progéniture. En pratique, la variance plus élevée du succès reproducteur des mâles par rapport à celle des femelles peut rendre le nombre efficace de gènes nucléaires inférieur au nombre efficace de gènes mitochondriaux (Birky *et al.* 1983). Pour ces raisons, le rapport réel liant la taille efficace des gènes nucléaires à la taille efficace des gènes mitochondriaux ne peut être estimé précisément qu'à l'aide d'études empiriques.

Chez toutes les espèces, les divers allèles d'un locus donné ont leur histoire propre, certains allèles partageant des événements de coalescence récents, et d'autres de plus anciens (Pamilo & Nei 1988). Lors d'un événement de spéciation, les allèles présents dans l'espèce ancestrale vont se répartir aléatoirement dans chacune des deux espèces descendantes (Fig. 1). Il en résulte que certains allèles dans une espèce donnée sont généalogiquement plus proches de certains allèles présents dans une autre espèce qu'ils ne le sont de certains allèles de leur propre espèce. Avec le temps, des lignées alléliques de chaque espèce descendantes vont être aléatoirement perdues par dérive, tandis que de nouveaux allèles vont apparaître par mutation. Finalement, au bout d'un certain temps, une seule lignée ancestrale d'allèles aura persisté

dans une espèce descendante de sorte que la totalité de la variation allélique intraspécifique reflétera des mutations apparues durant la période faisant suite à l'évènement de spéciation. Le temps nécessaire pour que le polymorphisme allélique d'un gène d'une espèce redevienne monophylétique, que l'on appelle également le temps nécessaire pour qu'en moyenne le tri du polymorphisme ancestral soit complet, est estimé à quatre fois la taille efficace de l'espèce descendante. On comprend ainsi que, la taille efficace mitochondriale étant généralement considérée comme étant quatre fois plus faible que la taille efficace nucléaire (mais voir la « Note sur la taille efficace de l'ADNmt »), le tri complet du polymorphisme ancestral sera plus rapidement atteint avec des gènes mitochondriaux qu'avec des gènes nucléaires. Ce processus rend plus fiable l'utilisation de gènes mitochondriaux que celle de gènes nucléaires dans le but de reconstruire, à partir de l'arbre d'un gène, l'arbre des espèces (Hudson & Turelli 2003). Par exemple, certaines théories prédisent que lorsque l'on a 95% de certitude qu'un inter-nœud de l'arbre d'un unique gène mitochondrial ne soit pas affecté par le tri incomplet du polymorphisme ancestral, il faut reconstruire les arbres phylogénétiques de 16 gènes nucléaires indépendants pour retrouver un niveau de certitude équivalent (Moore 1995).

Finalement, le principal avantage de l'ADNmt par rapport à l'ADNn en tant que marqueur moléculaire pour les études de phylogénie et de phylogéographie provient de son taux de mutation élevé. L'ADNmt est présenté depuis 25 ans comme la molécule à évolution rapide par excellence (Brown *et al.* 1979), avec seulement quelques ARN viraux, répliqués par des polymérases dénuées de système de correction, connus pour évoluer plus rapidement. Certaines séquences à l'intérieur de la région de contrôle évoluent même plus rapidement que le reste du génome, et sont donc particulièrement intéressantes pour des analyses nécessitant une résolution fine de structures de populations (Stoneking *et al.* 1991). Un avantage certain pour la lecture et l'alignement des séquences vient du fait que bien que des insertions et délétions ne soient pas rares, la majorité de la variation entre séquences provient de transversions et de transitions (avec un fort biais pour ces dernières ; Aquadro & Greenberg 1983).

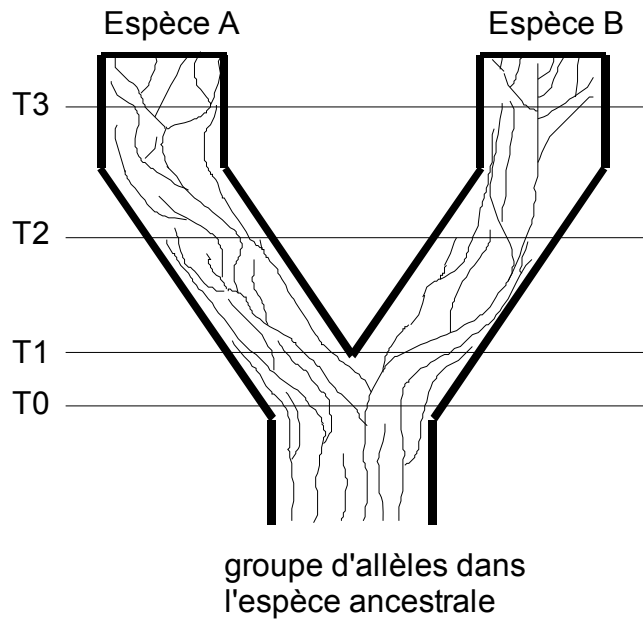


Figure 1. Le tri du polymorphisme ancestral. Les allèles sont représentés par des lignes sinueuses et fines et les espèces par des lignes droites et épaisses. Une espèce ancestrale possède un groupe d'allèles pour un gène donné. Lorsque que les espèces filles A et B commencent à diverger (T_0), le groupe ancestral d'allèles se distribue aléatoirement au sein de ces deux espèces filles. Lorsque les espèces filles deviennent reproductivement isolées (T_1), leur groupe d'allèles respectif est d'abord polyphylétique (entre T_1 et T_2), puis paraphylétique (entre T_2 et T_3). A partir de T_3 , les allèles sont réciproquement monophylétiques. Le temps écoulé entre T_1 et T_3 est le temps nécessaire pour que le tri du polymorphisme ancestral soit complet et est approximativement égal à $4Ne$ générations, avec Ne la taille efficace de la population.

En plus de sa grande utilité en tant que marqueur de l'évolution récente, l'ADNmt a été également largement employé comme marqueur de relations phylogénétiques entre taxa de rangs supérieurs. Une première approche servant à cet effet est la comparaison directe de séquences à évolution plus lente que la moyenne du génome mitochondrial, ou une partie de ces séquences comme les transversions ou les substitutions non-synonymes des régions codant pour des protéines. Des régions particulièrement populaires pour ce genre d'étude sont les gènes codant pour les sous-unités d'ARNr par exemple (Sullivan *et al.* 1995), pour les cytochromes oxydases (e.g. Hebert *et al.* 2003) et le cytochrome *b* (Irwin *et al.* 1991). Il est cependant à noter que chacune de ces régions codantes trouve son optimalité en tant que marqueur dans une fenêtre temporelle de résolution étroite et variable entre taxa. Une seconde approche utilisée pour la résolution des nœuds anciens est l'étude des arrangements de gènes. Dans la plupart des espèces animales, l'ADNmt est constitué des mêmes 37 gènes, mais l'ordre dans lequel ils se succèdent peut différer entre grandes lignées. Parce qu'un

réarrangement de gènes mitochondriaux reste un événement rare, l'ordre des gènes est un marqueur de clades pertinent (voir par exemple Hoffmann *et al.* 1992; Lee & Kocher 1995; Macey *et al.* 1997). Enfin, d'autres aspects de l'évolution du génome mitochondrial, comme l'évolution du code génétique (Wolstenholme 1992), ont pu servir de marqueurs d'évolution ancienne.

II.A.2. L'ADNmt comme code-barres moléculaire

Après un long règne en tant que marqueur de l'histoire des populations (utilisé dans plus de 80 % des études de phylogéographie de l'avant XXI^{ème} siècle, Avise 2000), et alors même que les limitations liées à son emploi étaient de plus en plus mises en avant dans la littérature évolutionniste, l'ADNmt démarra sa seconde ère de gloire avec l'avènement de méthodes d'identification taxonomiques rapides comme le code-barres moléculaire (*DNA Barcoding*). Hebert *et al.* (2003) fut le pionnier de cette entreprise encore aujourd'hui à succès, dont le but est l'utilisation de la séquence d'un court fragment d'ADN pour révéler l'identité de n'importe quelle espèce animale, voire même de permettre la découverte de nouvelles espèces : "Hebert envisages this 'gene species' as a first, mandatory step towards describing real species" (Nicholls 2003). Pour qu'une région de l'ADN puisse faire office de code-barres, elle doit être suffisamment variable, être assez courte pour être amplifiée en une seule réaction et contenir des régions invariables pour y développer des amorces universelles (Stoeckle 2003). En outre, une propriété essentielle d'un bon code-barres moléculaire, nécessaire pour qu'une séquence donnée soit attribuée à une espèce, est la distribution bimodale des distances génétiques, avec un chevauchement le plus faible possible entre les distances génétiques intra- et inter-spécifiques. Diverses études ont montré que le gène mitochondrial de l'enzyme cytochrome oxydase I (*coxI*) répondait au mieux à ces critères (Hebert *et al.* 2003; Hebert *et al.* 2004; Ward *et al.* 2005). Les applications du code-barres moléculaire sont nombreuses et variées, comme l'illustre le projet d'introduire cette méthode dans les services douaniers pour une identification rapide des espèces dont le commerce est réglementé par la Convention sur le Marché International des Espèces en Danger (CITES) (Lahaye *et al.* 2008).

II.A.3. Etude de la distribution des lignées génétiques du complexe *Podarcis hispanicus* à l'aide de marqueurs mitochondriaux

Le bassin Méditerranéen est un des principaux centres de biodiversité dans le monde (Myers *et al.* 2000). La Péninsule Ibérique, en tant qu'ancien grand refuge glaciaire, constitue l'une des trois plus importantes aires de diversité taxonomique dans le bassin Méditerranéen et plus largement dans le Paléarctique Occidental (Hewitt 1996; Taberlet *et al.* 1998; Hewitt 2001). L'herpétofaune ibérique est bien représentative de cette diversité avec 38 espèces dont 8 (soit 21%) endémiques. Cette diversité est par ailleurs encore largement sous-estimée dans la mesure où certaines espèces sont connues pour héberger plusieurs lignées génétiques distinctes et d'origine ancienne suggérant qu'une révision systématique de ces taxa est nécessaire. Ainsi, la persistance de lignées anciennes distinctes a été mise en évidence chez *Timon lepidus* (Mateo *et al.* 1996), *Lissotriton boscai* (Martinez-Solano *et al.* 2006) et *Blanus sp.* (Vaconcelos *et al.* 2006).

Un cas pour lequel une révision systématique apparaît nécessaire concerne le Lézard ibérique *Podarcis hispanicus*, chez qui des travaux utilisant des données d'ADNn, d'ADNmt et morphologiques ont montré que les différentes lignées évolutives ne constituent pas un ensemble monophylétique au regard d'autres lignées déjà élevées au rang d'espèce (Sá Sousa 2000; Harris & Sá Sousa 2001; Harris *et al.* 2002; Harris & Sá Sousa 2002; Pinho *et al.* 2003; Pinho *et al.* 2006). *Podarcis hispanicus* tel que traditionnellement reconnu représenterait donc un complexe d'espèces.

L'étude de Pinho *et al.* (2006) décrit les relations phylogénétiques entre lignées du complexe sur la base de plusieurs gènes mitochondriaux (Fig. 2). Selon cette étude, quatre lignées mitochondriales anciennes semblent être présentes dans le sud-est de l'Espagne. Toutefois, cette étude n'a échantillonné que quelques individus pour chacune des lignées ce qui ne permet pas de décrire leur distribution géographique. Un premier objectif que nous nous sommes fixé afin de mieux comprendre l'histoire évolutive de ce complexe d'espèces est de préciser la répartition géographique de ces unités évolutives. Pour se faire, nous avons séquencé un fragment de la région de contrôle de l'ADNmt chez 69 spécimens répartis dans le sud-est de la Péninsule Ibérique (voir Article 1 pour des précisions méthodologiques). Ce marqueur a été choisi pour son taux d'évolution en général plus élevé que le taux d'évolution moyen du génome mitochondrial, ce qui le rend particulièrement approprié aux études de phylogéographie. Par ailleurs, comme la région séquencée faisait partie des marqueurs utilisés

dans l'étude de Pinho *et al.* (2006), nous avons ajouté à notre alignement les séquences de 14 individus issus de cette étude, qui représentent chacun un des principaux clades présents dans le complexe *P. hispanicus*. Ainsi, il nous fut possible d'attribuer chacun de nos 69 individus à l'un de ces 14 clades. Le résultat est présenté dans la figure 3.

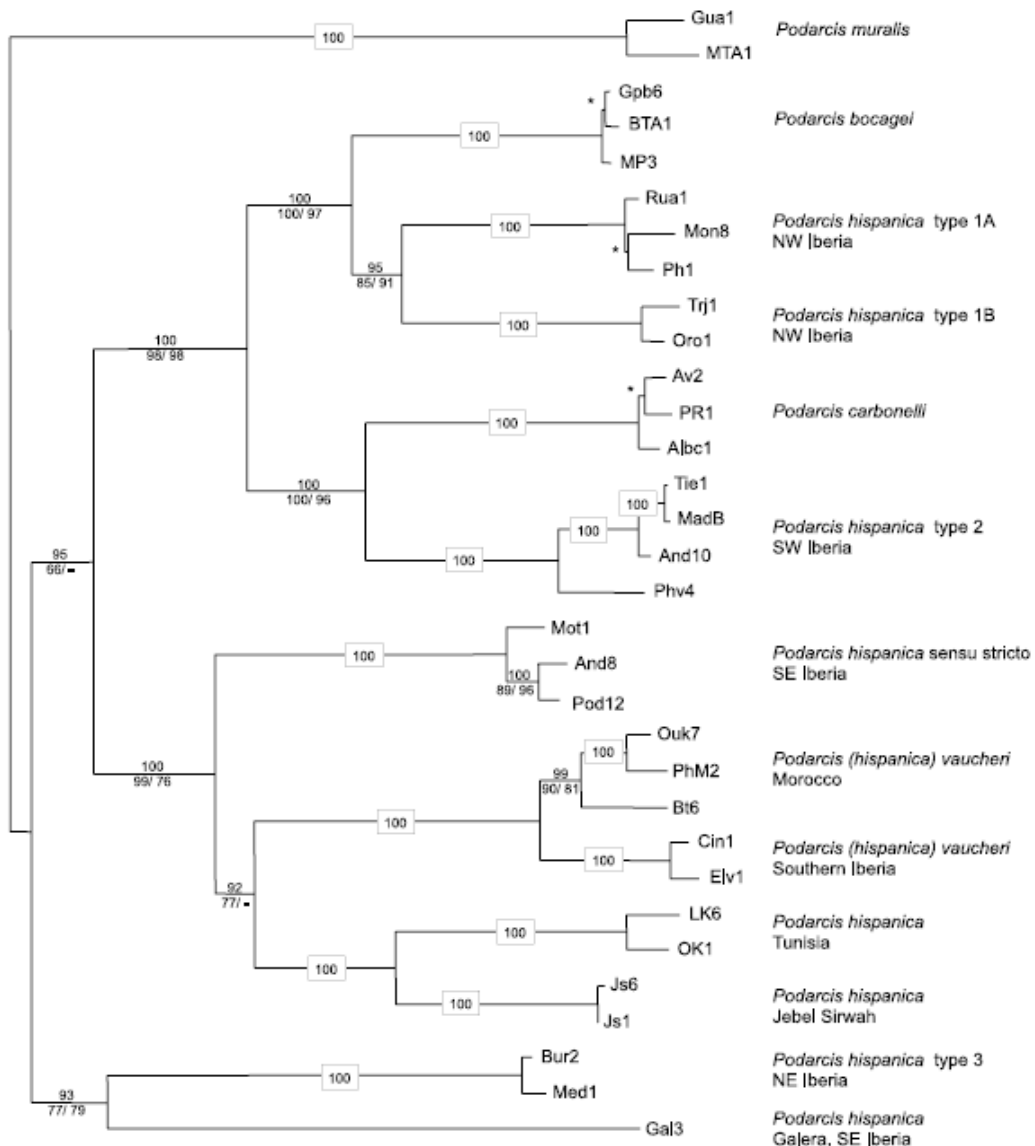


Figure 2. Arbre phylogénétique le plus vraisemblable des lignées mitochondriales anciennes des *Podarcis* de la Péninsule Ibérique présenté par Pinho *et al.* (2006). L'arbre a été enraciné avec *P. muralis*. Les probabilités postérieures bayésiennes sont données au dessus des branches, les valeurs de bootstraps en maximum de vraisemblance et en maximum de parcimonie, respectivement, sont données en dessous des branches. Lorsque ces trois valeurs sont identiques, une seule valeur est donnée en encadré.

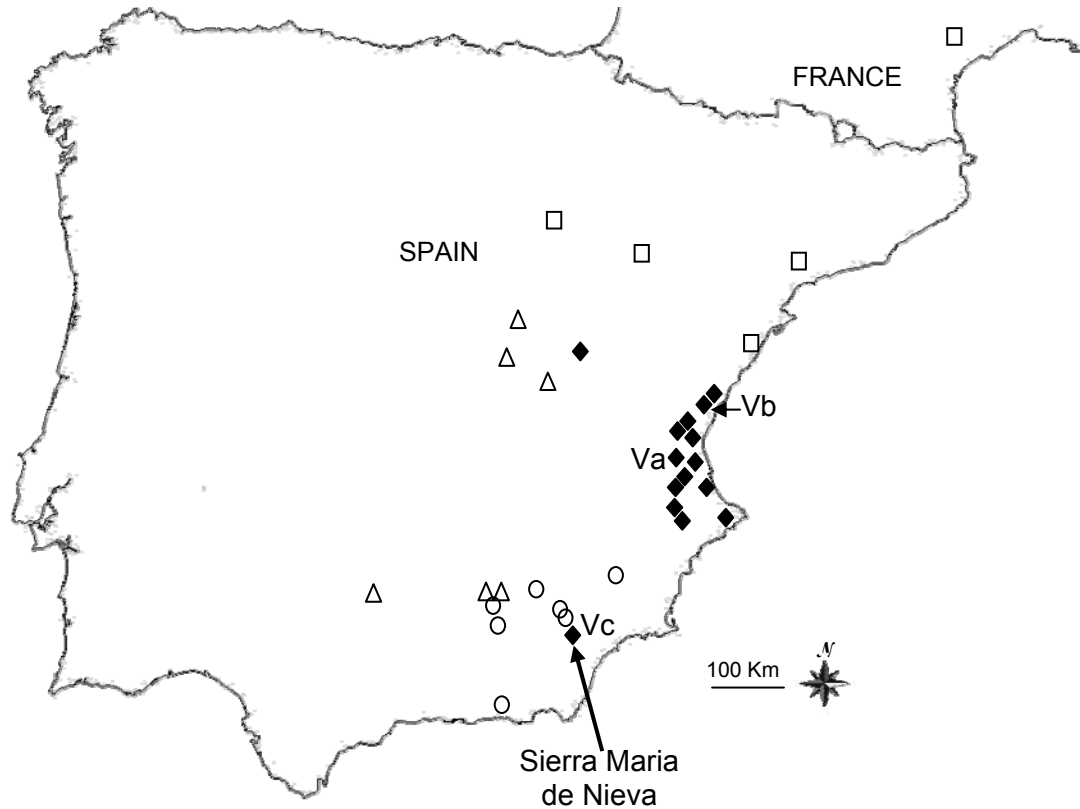


Figure 3. Carte de distribution des lignées mitochondriales profondes dans le sud-est de la Péninsule Ibérique. Les carrés, triangles, cercles et losanges représentent respectivement la lignée mitochondriale Lioplepis, Type 2, Hispanicus et Valencia. Va, Vb et Vc renseignent sur la distribution des trois haplotypes différents présents dans la lignée Valencia.

Les 69 séquences étudiées se sont réparties dans quatre lignées mitochondriales profondes, conformément aux prédictions faites d'après Pinho *et al.* (2006). Chacune des 33 stations échantillonnées ne renfermait que des haplotypes d'une seule lignée mitochondriale. De plus, une unique station mise à part, les quatre lignées se distribuent de manière contiguë et non chevauchante. Nous appellerons ces lignées Lioplepis pour celle située au nord de la région étudiée (équivalente à « *Podarcis hispanica* Type 3 » dans Pinho *et al.* 2006), Type 2 pour celle située à l'ouest, Hispanicus pour celle située au sud (équivalente à « *Podarcis hispanica* Galera » dans Pinho *et al.*) et Valencia pour celle située à l'est (équivalente à « *Podarcis hispanica* sensu stricto » dans Pinho *et al.*). L'exception à l'absence de chevauchement provient d'une station localisée dans la Sierra Maria de Nieva qui accueille 6 individus appartenant tous à la lignée Valencia, alors que cette localité est située au milieu de l'aire de distribution des Hispanicus. Notons dès à présent qu'au sein de la lignée Valencia, trois haplotypes ont été observés dans cette étude : un propre aux individus de la Sierra Maria de

Nieva (Vc), un présent chez les deux uniques individus d'une station située au nord de Valence, et un autre présent chez tous les autres individus de l'aire principale de distribution de la lignée Valencia.

II.B. L'ADNmt végétal

Parce que l'ensemble du génome mitochondrial des plantes évolue de manière lente (Palmer & Herbon 1988), cette molécule n'a été que rarement utilisée en phylogénie et phylogéographie chez les plantes. Toutefois, les séquences d'ADNmt contiennent des introns qui contiennent eux-mêmes des indels évoluant de manière très rapide. Ces indels ont ainsi permis de résoudre des relations de parenté entre taxa très proches (Freudenstein & Chase 2001). Enfin, le nombre grandissant de preuves de recombinaison dans ce génome constitue un handicap supplémentaire par rapport au génome chloroplastique pour une utilisation en phylogénie.

II.C. L'ADNcp

II.C.1. Les marqueurs chloroplastiques en phylogénie et en phylogéographie

L'ADNcp a connu un franc succès auprès des phylogénéticiens à partir des années 80 grâce à plusieurs propriétés le distinguant des autres génomes (Olmstead & Palmer 1994). Historiquement, la faible taille du génome chloroplastique a rendu aisé l'examen du génome complet par l'analyse des sites de restriction (Soltis *et al.* 1992). Tout comme pour l'ADNmt, la plupart des gènes de l'ADNcp sont présents en simple copie ce qui évite l'amplification de gènes paralogues (mais voir Wolfe & Randle 2004). De manière générale, l'ensemble des propriétés liées à la nature haploïde et au mode de transmission maternel de l'ADNmt, qui rendent certains fragments séquencés dans cette molécule propices à leur emploi en tant que marqueurs, sont applicables à l'ADNcp.

Le génome chloroplastique est, à quelques rares exceptions près, constitué de parties répétées et inversées, séparées par une longue et une courte partie présentes sous une copie (dite « simple-copie »). Les premières études portant sur la vitesse d'évolution de l'ADNcp ont suggéré que les parties répétées et inversées évoluaient plus lentement que les parties

simple-copie (Curtis & Clegg 1984), ce qui fut confirmé par la suite : ces dernières évoluent en moyenne 2,3 fois plus rapidement que les premières (Perry & Wolfe 2002). Pour cette raison, les premières recherches en phylogénie utilisant l'ADNcp ont utilisé des gènes des parties simple-copie. Le taux d'évolution toutefois lent de ces parties simple-copie par rapport à l'ADNn a initialement orienté l'utilisation de l'ADNcp vers la résolution de phylogénies basales, notamment grâce à l'emploi de gènes *rbcL*, *ndhF*, *atpB*, et *matK* (Chase *et al.* 1993; Johnson & Soltis 1994; Steele & Vilgalys 1994; Olmstead & Reeves 1995). Par la suite, les régions non codantes de l'ADNcp ont été étudiées afin de trouver des marqueurs plus variables, utilisables dans la résolution des branchements récents. Ainsi, des fragments comme la région *TrnT-TrnL-TrnL-TrnF*, l'espace intergénique *atpB-rbcL* ou l'intron de *trnK/matK* ont été largement utilisés en phylogénie au niveau intra-générique (*e.g.* Bell & Patterson 2000; Bellstedt *et al.* 2001; Klak *et al.* 2003). Chez certaines espèces, la variabilité de l'ADNcp est telle que ce génome a été utilisé dans des études de phylogéographie (*e.g.* Soltis *et al.* 1997).

Récemment, Shaw *et al.* (2005) et Shaw *et al.* (2007) ont étudié la variation de l'ADNcp sur d'abord 21 régions non-codantes puis sur la totalité des régions non-codantes des parties simple-copie de l'ADNcp chez les principales lignées de phanérogames. Une conclusion de ces auteurs est que les régions les plus utilisées en phylogénie (*trnL-trnF* et *trnK/matK*) sont parmi les moins variables des régions non-codantes. Les auteurs proposent que de nouvelles régions, en moyenne beaucoup plus variables que ces régions classiquement utilisées, puissent être la source de nouveaux marqueurs en phylogénie fine et en phylogéographie. Ces études soulignent par ailleurs l'importante variation qui existe entre les différentes régions non-codantes. Les auteurs proposent donc que des études « pilotes » visant à tester l'utilité d'un grand nombre de régions non-codantes avec un nombre limité de séquences soient systématiquement menées pour l'étude phylogénétique ou phylogéographique d'une lignée donnée et pour une « profondeur » donnée.

II.C.2. L'ADNcp comme code-barres moléculaire

A cause du faible taux d'évolution (Cho *et al.* 1998; Cho *et al.* 2004) et des fréquents remaniements structuraux de l'ADNmt chez les plantes (Adams & Palmer 2003), le gène *coxI* ne représente pas un bon code-barres moléculaire pour ce phylum. Les difficultés d'obtention des séquences nucléaires en comparaison aux séquences chloroplastiques ont orienté la

recherche d'un code-barres moléculaire vers ce dernier génome (Kress *et al.* 2005; Chase *et al.* 2007; Ford *et al.* 2009). Les nombreux débats autour du choix d'un code-barres ont mis en avant sept marqueurs candidats : quatre sont des portions des gènes codants (*matK*, *rbcL*, *rpoB* et *rpoC1*) tandis que trois sont des fragments non-codants (*atpF-atpH*, *trnH-psbA* et *psbK-psbI*). Différentes équipes ont testé différentes combinaisons entre ces sept marqueurs (Erickson *et al.* 2008; Lahaye *et al.* 2008; Hollingsworth *et al.* 2009a). Très récemment, un consensus a émergé de ces débats et le groupe de travail sur les plantes appartenant au consortium pour le code-barres moléculaire a présenté ses conclusions : le duo *rbcL+matK* fera office de code-barres moléculaire chez les plantes (Hollingsworth *et al.* 2009b). Sur un jeu de données de plus de 900 séquences couvrant plus de 560 espèces, ce duo a permis l'identification des espèces dans 72% des cas et l'identification du genre dans 100% des cas. Afin d'augmenter le pourcentage d'identification au niveau spécifique, le groupe de travail recommande qu'un ou deux autres marqueurs soient ajoutés à ce duo. L'identité des ces marqueurs additionnels devra varier en fonction du groupe taxonomique et de la qualité des tissus étudiés.

II.C.3. Résolution des nœuds basaux de la phylogénie des Galoglychia grâce aux marqueurs chloroplastiques.

Les plantes du genre *Ficus* (Moraceae) représentent un modèle historique pour l'étude de la coévolution entre plantes et insectes (Herre *et al.* 1996; Machado *et al.* 2001; Weiblen 2001; Jouselin *et al.* 2003; Rønsted *et al.* 2005). Ce genre est exclusivement pollinisé par des guêpes de la famille des Agaonidae. Dans ce système, la règle est qu'une espèce de guêpe ne pollinise qu'une espèce de plante hôte (Ramírez 1970; Michaloud *et al.* 1985). De plus, les sections et sous-sections constituant ce genre apparaissent monophylétiques et pollinisées par des espèces d'un seul genre ou d'un nombre limité de genres de pollinisateurs. Toutefois, des exceptions à la règle « une espèce d'hôte pour une espèce de pollinisateur » semblent assez fréquentes, tant sur la base d'études morphologiques (Ramírez 1970; Michaloud *et al.* 1985) que sur la base de travaux moléculaires (Lopez-Vaamonde *et al.* 2001; Molbo *et al.* 2003; Machado *et al.* 2005; Erasmus *et al.* 2007). Ainsi, Rasplus (1996) a pu estimer d'après la morphologie des insectes et des plantes que 17% des espèces de figuiers Africains sont pollinisées par au moins deux espèces de guêpes et que 15% des espèces de guêpes utilisent au moins deux espèces de plante hôte. Cependant, une section particulière – la section

Galoglychia – semble être la principale source de ces contre-exemples. Etudier en détail la phylogénie des plantes hôtes afin de la comparer avec la phylogénie des insectes pollinisateurs dans cette section constitue alors en enjeu important pour comprendre des processus à l'origine de ces exceptions. Une étude récente a présenté les premiers résultats concernant la taxonomie et la phylogénie moléculaire des insectes pollinisateurs de cette section (Erasmus *et al.* 2007). Il reste donc à compléter l'information existante chez les figuiers.

La section *Galoglychia* section renferme 77 espèces distribuées en six sous-sections (Berg 1986; Berg & Wiebes 1992). Une étude phylogénétique récente basée sur des fragments nucléaires ITS et ETS présente les relations de parenté entre ces sous-sections (Rønsted *et al.* 2007a). Les auteurs ont par ailleurs inclus dans cette étude des espèces d'une section proche, la section *Americana*, constituées exclusivement d'espèces néotropicales. Un résultat intéressant, et même surprenant, est la paraphylie de la section *Galoglychia* au regard de la section *Americana*. En effet, dans les analyses de Rønsted *et al.* (2007a), la section *Americana* est systématiquement branchée à l'intérieur de la section *Galoglychia*. Toutefois, en fonction de la méthode de reconstruction choisie, la position des *Americana* varie. Le branchement des *Americana* à l'intérieur des *Galoglychia* n'est fortement soutenu dans aucune des topologies présentées. Un doute subsiste ainsi sur la monophylie ou non de la section *Galoglychia*. Nous avons alors tenté de résoudre le nœud basal représentant le branchement des deux sections en utilisant des marqueurs moléculaires à taux d'évolution plus lents que ceux des marqueurs ITS et ETS.

Les marqueurs chloroplastiques offrent *a priori* des outils adéquats pour une telle étude phylogénétique. Bien que les marqueurs chloroplastiques aient été les premiers marqueurs utilisés chez les figuiers (Herre *et al.* 1996), leur très faible polymorphisme les a exclus des études suivantes. Notamment, Rønsted *et al.* (2007b) ont étudié la variation nucléotidique des marqueurs chloroplastiques les plus utilisés chez les plantes sur six espèces de *Ficus*. La très faible variation observée a contraint les auteurs à abandonner les marqueurs chloroplastiques dans leurs études phylogénétiques des figuiers. Afin de choisir les marqueurs les plus variables et selon les recommandations de Shaw *et al.* (2007), nous avons donc procédé à une phase de tri des marqueurs chloroplastiques disponibles suivie d'une phase de développement d'amorces spécifiques. La partie « Additional Files » de l'Article 2 présente les détails des différentes phases de développement des marqueurs chloroplastiques.

L'alignement final est constitué de 58 séquences de plus de 3600 pb, représentant 38 espèces de la section *Galoglychia*, et 10 espèces de la section *Americana*. Les méthodes d'alignement et d'analyses phylogénétiques sont présentées en détail dans l'Article 2.

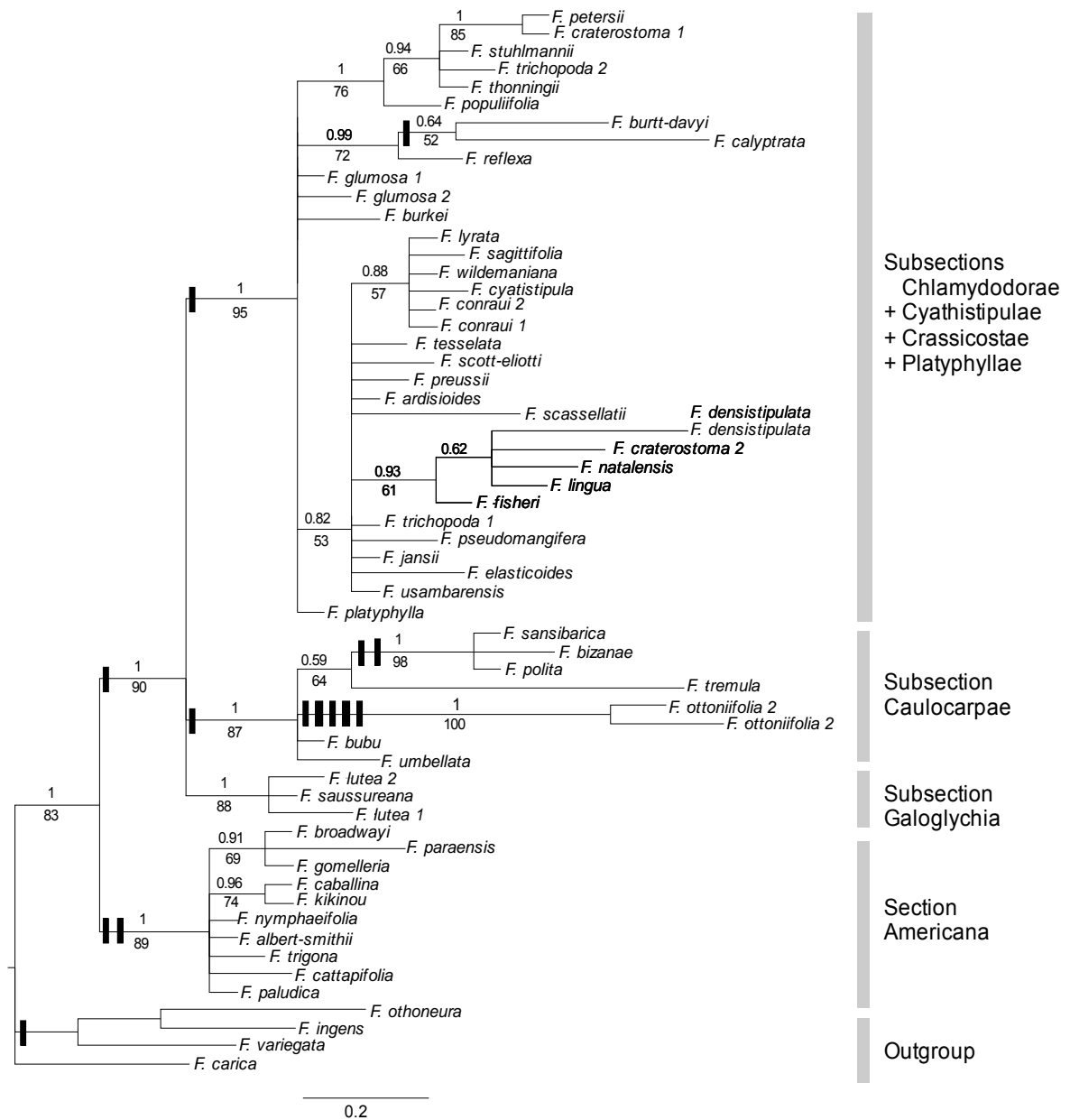


Figure 4. Arbre phylogénétique de la section *Galoglychia* incluant la section *Americana*, reconstruit à partir de marqueurs chloroplastiques par inférence bayésienne. Les valeurs au-dessus et en dessous des branches indiquent respectivement les valeurs de probabilité postérieure bayésienne et de bootstrap en maximum de vraisemblance. Les rectangles noirs indiquent la position d'indels synapomorphiques.

La phylogénie chloroplastique obtenue fut la même quelle que soit la méthode de reconstruction utilisée (Maximum de parcimonies, Maximum de vraisemblance et Inférence Bayésienne). Dans cette phylogénie (Fig. 4), la section *Galoglychia* est monophylétique : la section *Americana* se branche systématiquement en groupe externe à toutes les séquences de *Galoglychia*. En outre, une insertion de 10 pb dans un des cinq fragments chloroplastiques caractérise tous les individus de la section *Galoglychia*, tandis qu'une autre insertion de 10pb et une délétion d'une paire de bases caractérisent les séquences de la section *Americana*. Ainsi, les marqueurs chloroplastiques, au taux d'évolution lent, ont permis de résoudre un nœud basal que des marqueurs au taux d'évolution plus rapide étaient incapables de résoudre.

Les marqueurs chloroplastiques ont aussi permis de résoudre des nœuds plus récents, fortement soutenu par les valeurs de bootstraps et de probabilité postérieures. La section *Galoglychia* présente une trichotomie séparant la sous-section *Caulocarpae*, la sous-section *Galoglychia* et un clade regroupant les sous-sections *Cyathistipulae*, *Crassicostae*, *Chlamydodora* et *Platyphyllae* (appelé clade CCCP par la suite). Enfin, à l'intérieur du clade CCCP, un clade est modérément soutenu par les valeurs de bootstraps mais fortement soutenu par les valeurs de probabilité postérieures : il regroupe *F. petersii*, *F. stuhlmannii*, *F. thonningii* et une accession de chacune des espèces *F. trichopoda* et *F. craterostoma*. Notons que ce dernier clade est étonnant par sa composition, dans la mesure où les deux accessions de *F. trichopoda* et les deux accessions de *F. craterostoma* ne se branchent pas ensemble. En outre, *F. trichopoda* appartient à une sous-section différente de la sous-section des autres espèces du clade. Ces deux points soulignent une discordance entre la phylogénie chloroplastique et la taxonomie traditionnelle des *Ficus*.

Les deux études de cas présentées ici ont fourni des exemples d'utilisation de marqueurs à taux d'évolution variés afin de retracer des relations phylogénétiques d'ancienneté différentes. Dans ces deux cas, l'ADN cytoplasmique a été choisi : ADNmt à évolution rapide dans le premier cas et ADNcp à évolution lente dans le second cas. Ces molécules d'ADN étant physiquement disjointes de l'ADN nucléaire, on peut cependant se demander dans quelle mesure l'évolution des organites retracée à travers l'étude de leur génome est représentative de l'évolution des organismes hébergeant ces organites.

Chapitre III

HISTOIRE DES ESPECES VERSUS HISTOIRE DES GENOMES CYTOPLASMIQUES

Du fait de l'absence de recombinaison au sein des génomes cytoplasmiques, les différents fragments qui les composent sont tous liés, ce qui conduit certains auteurs à considérer chacun de ces génomes comme un « super-gène » (Mallet 2005). Retracer l'histoire des êtres vivants à partir de marqueurs cytoplasmiques revient donc à retracer l'histoire de la mitochondrie, ceci quelque soit le nombre de marqueurs utilisés. Or l'histoire des gènes ne suit pas toujours l'histoire des espèces (Pamilo & Nei 1988; Doyle 1992; Hudson 1992; Moore 1995; Brower *et al.* 1996; Maddison 1997; Page & Charleston 1997; Nichols 2001; Sota & Vogler 2001; Degnan & Rosenberg 2006). Des mécanismes évolutifs comme le tri incomplet du polymorphisme ancestral ou l'hybridation introgressive (développés au Chapitre IV) peuvent en effet rendre l'histoire d'un gène non représentative de l'histoire des espèces. Par la suite, on emploiera le terme de « discordance » pour définir une incohérence entre l'information fournie par un gène (ou un ensemble de gènes issus d'un même génome cytoplasmique) sur l'histoire des espèces et la véritable histoire des espèces. Ce chapitre présente les méthodes utilisées pour détecter et tester la significativité d'une discordance.

III.A. Détection d'une discordance cyto-morphologique : le cas des *Podarcis* ibériques

La détection d'une discordance est basée sur la comparaison de l'information relative à l'histoire évolutive des espèces, apportée par différents marqueurs neutres et indépendants (Doyle 1992). Les caractères morphologiques ont été les premiers à être utilisés pour reconstruire l'histoire des espèces (Stevens 1984). Exploitée dans un contexte

biogéographique, la variation observée au niveau de ces caractères offre un référentiel contre lequel la variation d'un marqueur cytoplasmique peut être confrontée.

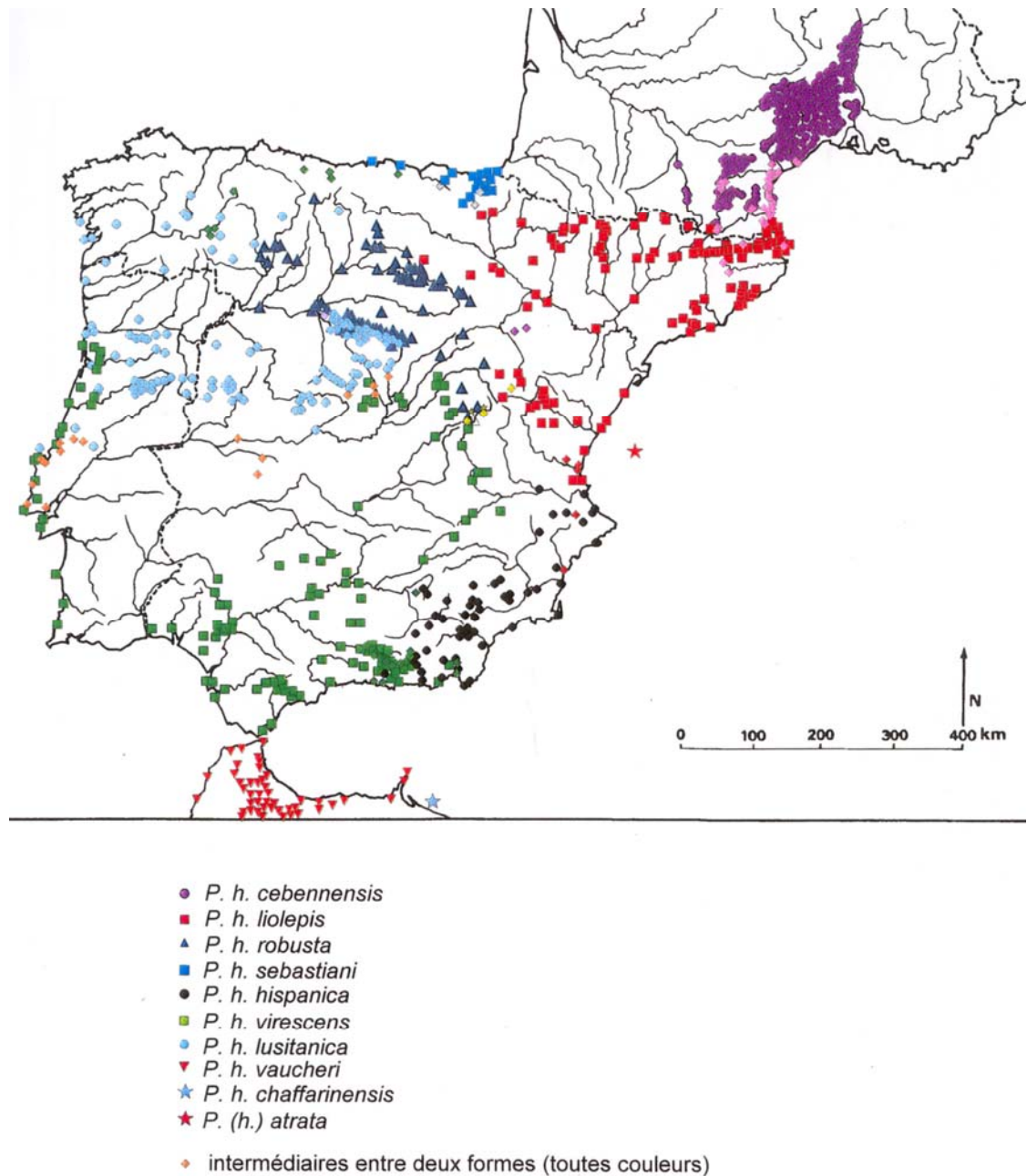


Figure 5. Carte de répartition des morphotypes et des intermédiaires morphologiques de *Podarcis hispanicus* lato sensu dans la Péninsule Ibérique et le sud-est de la France (carte originale extraite de Geniez 2001). Sur cette carte, les morphotypes sont traités comme sous-espèces.

Dans le cas des lézards *Podarcis* du sud-est de la Péninsule Ibérique, nous avons étudié la distribution des trois morphotypes traditionnellement reconnus par la taxonomie traditionnelle : *liolepis*, *virescens* (appelé « type 2 » par la suite), et *hispanica* (appelé *hispanicus* par la suite) (Guillaume 1987; Geniez 2001; Harris & Sá Sousa 2001; Harris *et al.* 2002), dont la distribution est représentée sur la figure 5.

La distribution de ces morphotypes paraît non chevauchante et aucune localité n'est connue pour en accueillir plus d'un seul à la fois. Toutefois, de nombreuses formes intermédiaires semblent être présentes dans les zones de contact, notamment entre les morphotypes *liolepis* et *hispanicus*, dans la région située autour de la ville de Valence (losanges rouges sur la figure 5). La présence de trois morphotypes seulement dans le sud-est de la Péninsule Ibérique semble en contradiction avec la découverte de quatre lignées mitochondriales profondément divergentes dans cette même région (Harris *et al.* 2002; Pinho *et al.* 2006). De plus, la localisation des individus de forme intermédiaire entre *liolepis* et *hispanicus* semble correspondre à la distribution de la lignée mitochondriale Valencia. Afin de déterminer s'il existe une réelle discordance entre la distribution des morphotypes et la distribution des lignées mitochondriales anciennes, il convient donc d'établir une carte plus précise des morphotypes et d'étudier la nature de la variation morphologique observée au niveau de cette zone de contact. Dans un premier temps, et à l'aide d'une analyse discriminante utilisant la variation observée sur 24 caractères morphologiques, nous avons redéfini la limite exacte des morphotypes en postulant que seulement trois d'entre eux étaient présents dans la zone étudiée (*cf.* Article 1 pour les détails méthodologiques). Le résultat est illustré en figure 6.

Conformément aux observations de terrain, la distribution des morphotypes n'apparaît pas chevauchante. En comparant les figures 3 et 6, il est possible de montrer que la distribution du morphotype « type 2 » correspond parfaitement à la distribution de la lignée mitochondriale Type 2. Le morphotype *hispanicus* correspond à la distribution de la lignée mitochondriale Hispanicus à l'exception de la population de la Sierra Maria de Nieva qui présente un morphotype *hispanicus* mais possède un haplotype Valencia. Le morphotype *liolepis* comprend deux lignées mitochondriales profondes (et non monophylétiques) : *Liolepis* dans la partie nord de sa distribution et Valencia dans sa partie sud. En résumé, il existe trois lignées mitochondriales (*Liolepis*, *Hispanicus* et Type 2), chacune présente au sein d'un seul morphotype (*liolepis*, *hispanicus* et « type 2 », respectivement) tandis qu'une

quatrième lignée mitochondriale (Valencia) est retrouvée dans deux morphotypes (*hispanicus* et *liolepis*). La présence de cette quatrième lignée semble donc discordante par rapport à la distribution des morphotypes.

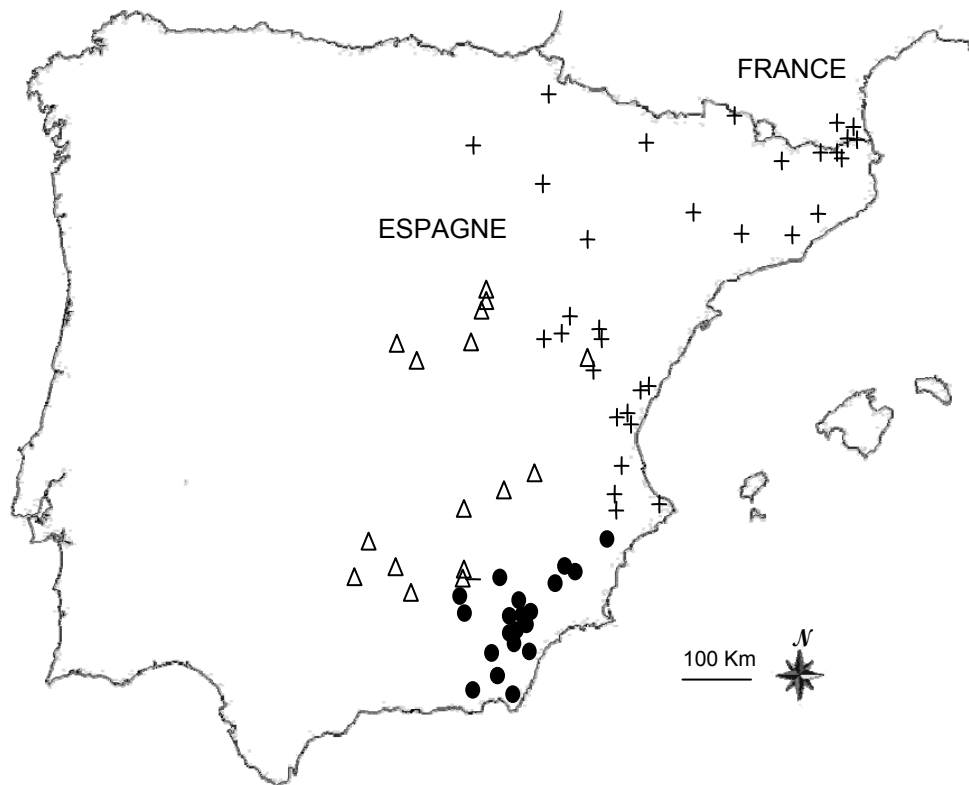


Figure 6. Distribution géographique des trois morphotypes de *P. hispanicus* sensu lato présents dans le sud-est et l'est de la Péninsule Ibérique d'après l'analyse discriminante effectuée avec 24 variables morphologiques. Triangles blancs = type 2 ; ronds noirs = *hispanicus* ; croix noires = *liolepis*.

Pour confirmer la discordance cyto-morphologique, il est néanmoins nécessaire d'étudier si la variation morphologique observée dans la région de Valence (là où est localisée la lignée mitochondriale Valencia) ne correspond pas à un quatrième morphotype. Nous avons alors analysé la variation de la morphologie des *hispanicus* et des *liolepis* le long d'un transect à l'aide d'une analyse en composantes principales (ACP) qui, contrairement à l'analyse discriminante, ne nécessite pas de connaissance *a priori* sur le nombre de morphotypes. Le transect s'étire sur une distance d'environ 1000 km, du sud de l'Espagne au sud de la France, en longeant la côte méditerranéenne (voir Article 1 pour des détails méthodologiques et pour les résultats des tests statistiques). Le résultat, illustré en figure 7, montre qu'il existe une rupture (mise en évidence par PC1) dans la variation morphologique

qui correspond à la zone de transition entre les lignées mitochondriales Valencia et Hispanicus. Cette limite marque donc la zone de contact entre les morphotypes *liolepis* et *hispanicus*. Au niveau de la transition entre Valencia et Liolepis, il existe une variation graduelle de la morphologie sans rupture de pente. Plus précisément, il n'existe pas de différenciation morphologique significative entre les populations localisées de part et d'autre de la limite de répartition de ces deux lignées mitochondriales. Les individus classés comme *liolepis* par l'analyse discriminante ne semblent donc, en effet, renfermer qu'un seul et unique morphotype. Les individus *liolepis* localisés au niveau de la zone de contact avec *hispanicus* semblent présenter un phénotype différent des autres *liolepis*, intermédiaires entre ces derniers et les *hispanicus*.

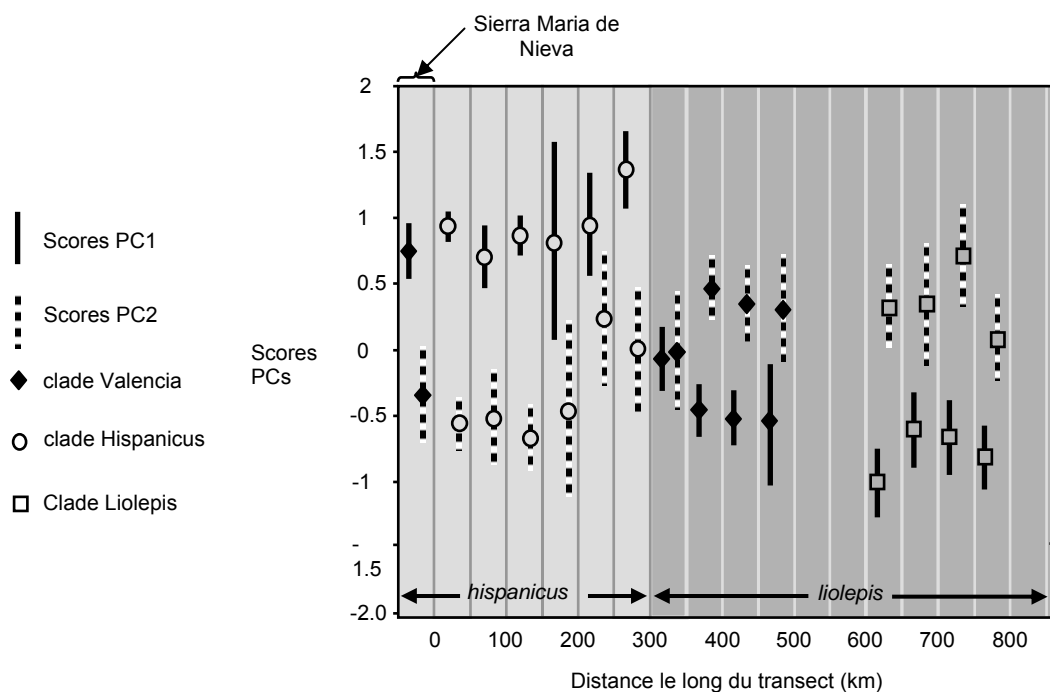


Figure 7. Distribution des scores moyens de la première (PC1) et deuxième (PC2) composante principale calculés à partir de 24 variables morphologiques le long d'un transect parcourant la distribution des morphotypes *liolepis* et *hispanicus*. PC1 et PC2 ont été moyennées pour chaque intervalle de 50 km. Les barres verticales représentent l'intervalle de confiance à 95 %. Les symboles renseignent sur la lignée mitochondriale présente dans l'intervalle considéré.

Cette variation morphologique ne correspond pas à un changement de distribution de lignée mitochondriale mais correspondrait davantage à des traces d'introgessions morphologiques par *hispanicus* couplées à de l'adaptation locale dans une zone de transition climatique (Pérez Cueva 1994; et voir Article 1 pour plus de détails). Par ailleurs, la population de la Sierra

Maria de Nieva dont les individus (n=6) possèdent un haplotype Valencia n'est morphologiquement pas différente des populations *hispanicus* situées aux alentours. En conclusion, il n'existe bien que trois morphotypes de *Podarcis hispanicus* sensu lato dans le sud-est de la Péninsule Ibérique. La présence de la lignée mitochondriale Valencia est responsable de deux discordances entre la morphologie et l'ADNmt : une impliquant les individus de la Sierra Maria de Nieva et l'autre impliquant les individus localisés autour de la région de Valence.

III.B. Détection d'une discordance cyto-nucléaire : le cas des *Podarcis* ibériques

Si les caractères morphologiques peuvent offrir un premier référentiel pour explorer une discordance entre l'histoire d'un génome cytoplasmique et l'histoire des espèces, les fréquents écarts aux conditions de neutralité et d'indépendance observés avec ce type de marqueur imposent la plus grande prudence quant à leur utilisation. Ces écarts impliquent par exemple la pléiotropie ou encore la participation de nombreux caractères dans la constitution de syndromes adaptatifs (Price 1997). En outre, les variations phénotypiques sont davantage susceptibles de refléter des adaptations locales ou de la plasticité phénotypique que l'histoire évolutive des populations (Westeberhard 1989; Avise 1994; Parra-Olea & Wake 2001). Pour ces raisons, il est recommandé d'utiliser une série de marqueurs moléculaires neutres et indépendants pour mettre en évidence une discordance entre l'histoire d'un génome cytoplasmique et l'histoire d'une espèce. La variation observée aux allèles microsatellites, si elle ne permet pas de retracer l'histoire des espèces ou des populations, est toutefois le produit de leur évolution et en offre un reflet indirect. En outre, il est devenu routinier de tester l'indépendance et la neutralité des différents loci microsatellites utilisés. Enfin, leur importante variabilité permet d'étudier la variation génétique entre populations ou entre espèces ayant récemment divergé (Avise 2000). Les marqueurs microsatellites fournissent ainsi un référentiel neutre idéal pour étudier la distribution des populations ou des espèces à laquelle la distribution de la variation cytoplasmique peut être confrontée.

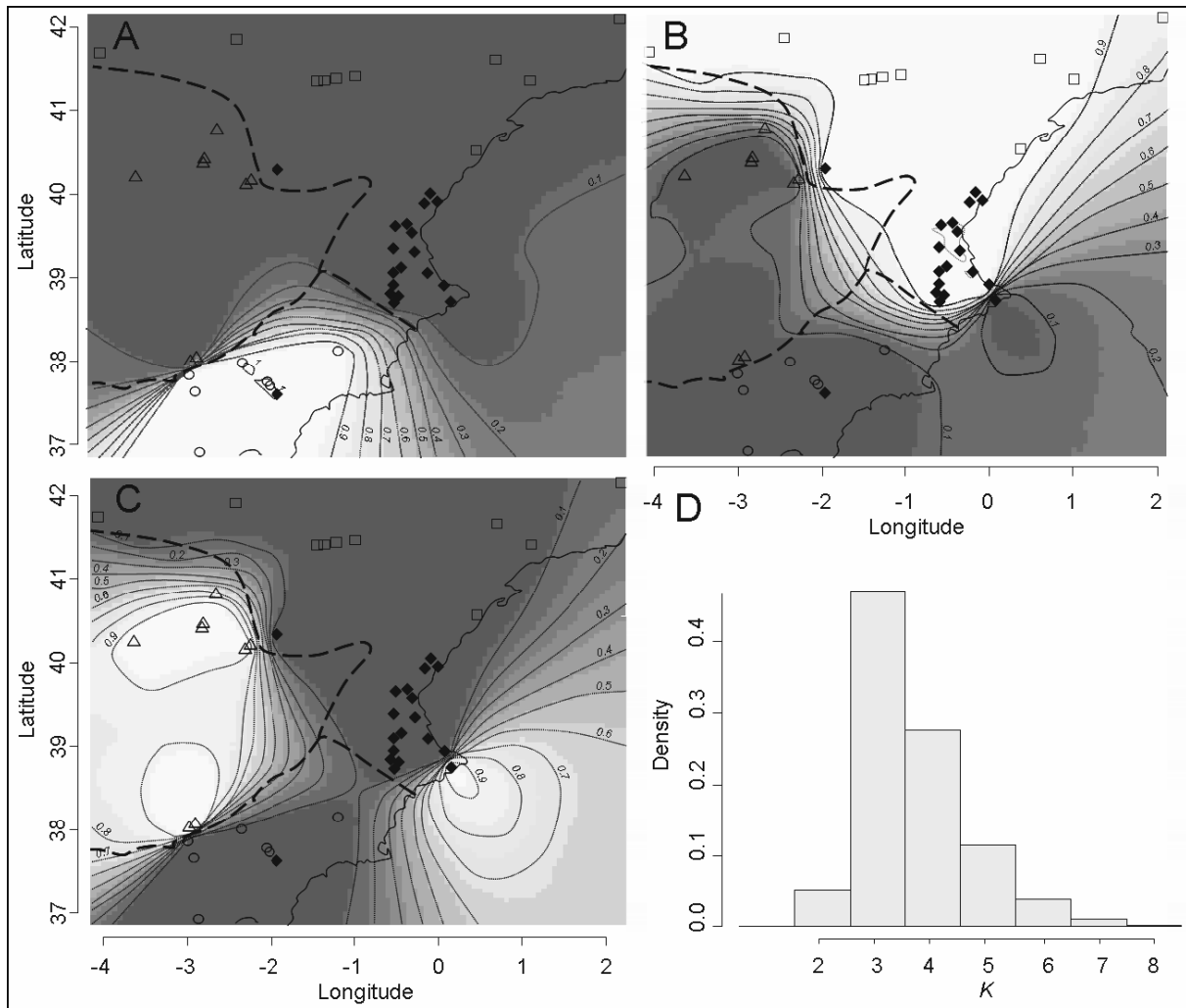


Figure 8. Délimitation des unités génétiquement différenciées à 8 marqueurs microsatellites et un intron nucléaires chez les Lézards ibériques du sud-est et de l'est de la Péninsule Ibérique d'après le logiciel Geneland. Les graphiques A à C représentent l'assignation de chaque pixel à une des trois unités génétiques identifiées : (A) *hispanicus*; (B) *liolepis*; (C) "type 2". Les plus hautes probabilités d'appartenance à une unité génétique sont représentées par les teintes les plus claires. Les symboles indiquent la position de chaque localité échantillonnée et renseignent sur l'identité de la lignée mitochondriale présente à cette même localité (même légende qu'en Figure 7). Les lignes en tirets longs délimitent la distribution des trois morphotypes présents dans la zone. Les échelles correspondent aux coordonnées géo-référencées en degrés décimaux (système de projection WGS84). D. Distribution de la densité postérieure du nombre d'unités génétiques le long de la chaîne MCMC.

Nous avons étudié la distribution de la variabilité génétique neutre à l'aide de huit marqueurs microsatellites et d'un intron nucléaire chez les *Podarcis* du sud-est et de l'est de la Péninsule Ibérique afin de la comparer à la distribution de la variabilité mitochondriale. Nous avons utilisé Geneland (Guillot *et al.* 2005), un logiciel de délimitation d'unités génétiques, spatialement explicite et basée sur l'analyse Bayésienne, afin d'identifier le

nombre d'unités génétiques nucléaires différenciées et pour en localiser les contours géographiques.

Le nombre d'unités génétiques nucléaires neutres présentant la densité postérieure la plus élevée est de trois, conformément au nombre d'unités morphologiques. L'attribution de chaque localité à une de ces trois unités génétiques révèle une structuration géographique de la variabilité génétique nucléaire neutre identique à celle de la variabilité morphologique, à une localité près (voir l'Article 1 pour une discussion sur cette exception). Plus précisément, la localité située dans la Sierra Maria de Nieva (possédant la lignée mitochondriale Valencia) a été attribuée, avec une probabilité postérieure de 100%, à une unité génétique qui regroupe par ailleurs toutes les localités comprenant un morphotype *hispanicus* et une lignée mitochondriale *Hispanicus*. L'absence de différenciation génétique nucléaire neutre entre la population de la Sierra Maria de Nieva et les populations voisines a été confirmée par les valeurs de F_{st} telles que calculées selon Weir & Cockerham (1984 ; résultats présentés dans l'Article 1). De plus, le logiciel n'a jamais recouvert une unité génétique constituée uniquement ou essentiellement des localités possédant la lignée mitochondriale Valencia.

En conclusion, les marqueurs nucléaires et la morphologie convergent pour soutenir l'existence de trois unités évolutives chez les lézards ibériques du sud-est de l'Espagne. Nous appellerons ces trois unités évolutives *P. liolepis* (ayant le morphotype *liolepis* et la lignée mitochondriale *Liolepis*), *P. hispanicus* (morphotype *hispanicus* et la lignée mitochondriale *Hispanicus*) et *P. sp.* « type 2 » (morphotype « type 2 » et la lignée mitochondriale Type 2). La quatrième lignée mitochondriale Valencia est « surnuméraire » dans cette région de l'Espagne et est ainsi responsable de deux cas de discordance cyto-nucléaire.

III.C. Tester la significativité d'une discordance cyto-nucléaire en phylogénie

III.C.1. Méthodes générales

Les analyses phylogéographiques diffèrent généralement des analyses phylogénétiques par la prise en compte de la géographie dans la reconstruction de l'histoire des lignées. L'objet de la phylogénie est la reconstruction des relations de parenté anciennes entre espèces pour laquelle les données biogéographiques actuelles sont généralement de faible utilité. La reconnaissance d'une discordance entre l'histoire d'un gène et l'histoire des espèces se fait alors par la comparaison de l'arbre reconstruit à partir du gène avec l'arbre reconstruit à partir de caractères morphologiques (*e.g.* Yoder *et al.* 2001) et/ou d'autres gènes indépendants (*e.g.* Baum *et al.* 1998; Barker & Lutzoni 2002; Buckley *et al.* 2002). Pour les mêmes raisons que celles développées précédemment, un ensemble de plusieurs gènes nucléaires neutres est une source plus fiable pour la reconstruction de l'histoire des espèces qu'un ensemble de caractères morphologiques (Bortiri *et al.* 2006).

Le test ILD ("Incongruence Length Difference", Farris *et al.* 1994, 1995) est la méthode la plus utilisée en phylogénie pour tester si deux ou plusieurs jeux de données sont combinables dans une seule et même analyse de reconstruction (Bull *et al.* 1993; Cunningham 1997b, 1997a). Par extension, ce test a été largement utilisé pour étudier les discordances entre les histoires évolutives des gènes comme celles causées par des transferts horizontaux (Lecointre *et al.* 1998; Ballard 2000; Sota & Vogler 2001; Escobar-Paramo *et al.* 2004). Le test est basé sur le fait que les jeux de données ayant subi des processus évolutifs différents (=discordants) auront davantage d'homoplasie que des jeux de données ayant une même histoire évolutive. Par conséquent, l'arbre phylogénétique le plus parcimonieux combinant des jeux de données discordants aura une longueur significativement supérieure à la somme des longueurs d'arbres les plus parcimonieux reconstruits à partir de chaque jeu de données séparément. La statistique D du test ILD représente la différence entre ces deux calculs de longueur. En comparant D à une distribution générée en partitionnant aléatoirement le jeu de données combiné, le test ILD fournit un degré de significativité estimant l'erreur de type I, c'est-à-dire la probabilité de rejeter l'hypothèse nulle qui est que les différents jeux de données ne sont pas discordants entre eux (Farris *et al.* 1994).

Le test ILD souffre toutefois d'un excès de conservatisme, c'est-à-dire qu'il signale fréquemment une discordance entre deux jeux de données là où il n'y en a pas (Darlu & Lecointre 2002). Ce taux élevé d'erreur de Type I est provoqué par une sensibilité du test à la

différence de taux d'évolution, d'hétérogénéités de taux d'évolution entre sites et de bruit phylogénétique (*i.e.* les caractères informatifs ne concernant pas les accessions discordantes) entre les différents jeux de données. Des cas d'erreurs de Type II ont également été rapportés (Ramirez 2006). Toutefois, ces erreurs de Type II peuvent facilement être limitées lorsque les sites non-informatifs sont enlevés et le nombre de sites informatifs est suffisamment élevé par rapport au niveau phylogénétique prospecté (Hipp *et al.* 2004).

La statistique D est une mesure moyenne pour l'ensemble de l'arbre. Toutefois, le test ILD peut être également utilisé pour identifier les espèces ou accessions responsables de la discordance entre les jeux de données (Baker *et al.* 2001; Yoder *et al.* 2001; Hipp *et al.* 2004). Par exemple, une augmentation importante du degré de significativité calculé après avoir enlevé une ou plusieurs accessions indique l'implication de cette ou ces accessions dans la discordance, même si le test souffre dans l'ensemble d'un excès de conservatisme. Cette stratégie n'est cependant applicable qu'à des jeux de données de taille limitée ou lorsqu'il y a une suspicion *a priori* sur l'identité de la source de conflit. En l'absence de suspicion, le nombre de clades à extraire un à un augmente exponentiellement avec le nombre d'accessions, ce qui rend la procédure fastidieuse (Ramirez 2006).

Une fois qu'une discordance a été suspectée suite à l'emploi d'un test ILD, d'autres tests peuvent être utilisés pour valider ce résultat. Les tests évaluant l'aptitude d'une topologie alternative à expliquer les données relativement à une topologie originale peuvent être employés à cet effet. La topologie alternative est généralement construite en enlevant toutes les accessions ayant une position supposée discordante sauf une, et en refaisant une analyse phylogénétique en contraignant l'accession discordante restante à avoir un branchement non discordant.

Le test de Wilcoxon est le plus utilisé dans le contexte de la détection de discordances (Templeton 1983). Son principe est de comparer pour chaque caractère le nombre de changements dans la topologie la plus parcimonieuse au nombre de changements dans la topologie alternative. La différence entre le nombre de changements est calculée et ces différences sont classées pour chaque caractère selon leur magnitude et indépendamment de la direction des changements. Le signe positif ou négatif de la différence est alors appliqué, et les rangs positifs et négatifs sont additionnés séparément. Si la somme des rangs nécessitant plus de changements sur la topologie alternative est significativement plus grande que la somme des rangs nécessitant moins de changements, alors les données sont significativement discordantes avec la topologie alternative.

Le test de Kishino-Hasegawa (KH ; Kishino & Hasegawa 1989) estime l'erreur standard et l'intervalle de confiance de la différence de log-vraisemblance entre deux topologies distinctes représentant deux hypothèses alternatives pour expliquer un même jeu de données. A l'origine, Kishino et Hasegawa ont développé ce test pour des arbres spécifiés *a priori*, c'est-à-dire des hypothèses phylogénétiques développées indépendamment du jeu de données disponible. En effet, le test repose sur le fait qu'en obtenant de nombreuses versions du jeu de données (par bootstrapping non-paramétrique), on attend en moyenne que la vraisemblance des deux topologies à tester soit égale, c'est-à-dire que l'espérance de leur différence soit égale à zéro, ce qui correspond à l'hypothèse nulle. Toutefois, lorsque la sélection d'au moins une des deux topologies est dépendante des données, comme c'est le cas lorsque que l'on veut tester une topologie discordante à la topologie ayant le maximum de vraisemblance, l'espérance de la différence devient supérieure à zéro. L'hypothèse nulle du test n'est alors plus valide (Goldman *et al.* 2000). Pour cette raison, le test KH ne devrait pas être utilisé pour tester la significativité d'une discordance.

Shimodaira & Hasegawa (1999) furent les premiers à noter les limites d'application du test KH et développèrent un test non-paramétrique corrigé, le test SH. Ce test compare simultanément toutes les topologies d'un jeu de topologies donné contenant la topologie ayant le maximum de vraisemblance. L'hypothèse nulle testée stipule, dans ce cas, que tous les arbres du jeu de topologies sont également bons dans leur explication des données. L'effet du test est de recentrer l'espérance de sorte à diminuer la significativité associée à la différence de vraisemblance calculée entre la topologie alternative et la topologie ayant la vraisemblance maximale. Le contrôle de l'erreur de Type 1 avec le test SH est excellent (Shimodaira 2002) ce qui signifie que le test ne validera pas une discordance quand elle n'existe pas. Strimmer & Rambaut (2002) notent cependant que le test SH est également biaisé de sorte que le nombre d'arbres inclus dans l'intervalle de confiance tend à augmenter lorsque le nombre de topologies incluses dans le jeu de topologies à tester augmente. Ce biais confère un comportement conservateur du test SH. Le test AU (« Approximately Unbiased ; Shimodaira 2002) est une version modifiée du test SH réduisant le taux d'erreurs de Type 2. Le test AU est ainsi préférable lorsque de nombreux arbres doivent être comparés simultanément. Toutefois, le test de la significativité d'une discordance n'implique que la comparaison de deux arbres. Sachant que le test SH ne souffre pas d'un mauvais contrôle de l'erreur de Type 2 dans le cas précis de la comparaison de deux topologies, mais à l'inverse bénéficie d'un meilleur contrôle de l'erreur de Type 1 que le test AU, le test SH devrait constituer le test de choix pour ce type d'utilisation.

Parce que les méthodes Bayésiennes infèrent des arbres en quantité proportionnelle à leur probabilité postérieure, les outils classiques de statistique fréquentielle comme le test SH ne peuvent être appliqués pour tester des topologies alternatives avec ce type de reconstruction phylogénétique. Il est toutefois possible de tester la significativité d'une topologie alternative en analyse Bayésienne (Buckley *et al.* 2002; Brandley *et al.* 2005). La méthode teste si une hypothèse phylogénétique donnée est absente d'un jeu de topologies renfermant les topologies dont la somme des probabilités postérieures correspond à 95% de la probabilité postérieure cumulée. Si une hypothèse phylogénétique n'est pas présente dans ce jeu de topologies, alors elle peut être rejetée.

III.C.2. Test de la significativité de deux discordances phylogénétiquement emboîtées : le cas des figuiers *Galoglychia*

Comme mentionné précédemment, le branchement polyphylétique des deux accessions de *F. trichopoda* ainsi que des deux accessions de *F. craterostoma* est en désaccord avec la reconnaissance de ces accessions comme appartenant à la même espèce. De même, le branchement de *F. densistipulata*, une espèce de la sous-section des *Cyathistipulae* à l'intérieur d'un clade regroupant des espèces de la sous-section des *Chlamydodorea* est en désaccord avec la taxonomie qui reconnaît la monophylie des *Cyathistipulae* sur la base de caractères morphologiques (Berg & Wiebes 1992). Afin de savoir si ces branchements discordants entre la phylogénie chloroplastique et la taxonomie traditionnelle indiquent la nécessité d'une révision taxonomique, ou sont le fait d'une évolution originale de l'ADN chloroplastique, nous avons reconstruit la phylogénie des individus inclus dans l'arbre chloroplastique à partir de marqueurs nucléaires ITS et ETS (Fig. 9).

Les détails méthodologiques concernant le développement des marqueurs et la reconstruction phylogénétique sont fournis dans l'Article 2. Conformément à une précédente étude sur la phylogénie des *Galoglychia* se basant sur les marqueurs ITS et ETS (Rønsted *et al.* 2007a), et contrairement à la phylogénie chloroplastique, les sous-sections recouvrent leur monophylie. Plus précisément, les deux accessions de *F. craterostoma* et les deux accessions de *F. trichopoda* ne présentent pas de branchement discordant avec leur appartenance respective à une même espèce et *F. densistipulata* est branché au sein d'un clade ne regroupant que des espèces appartenant à la sous-section *Cyathistipulae*.

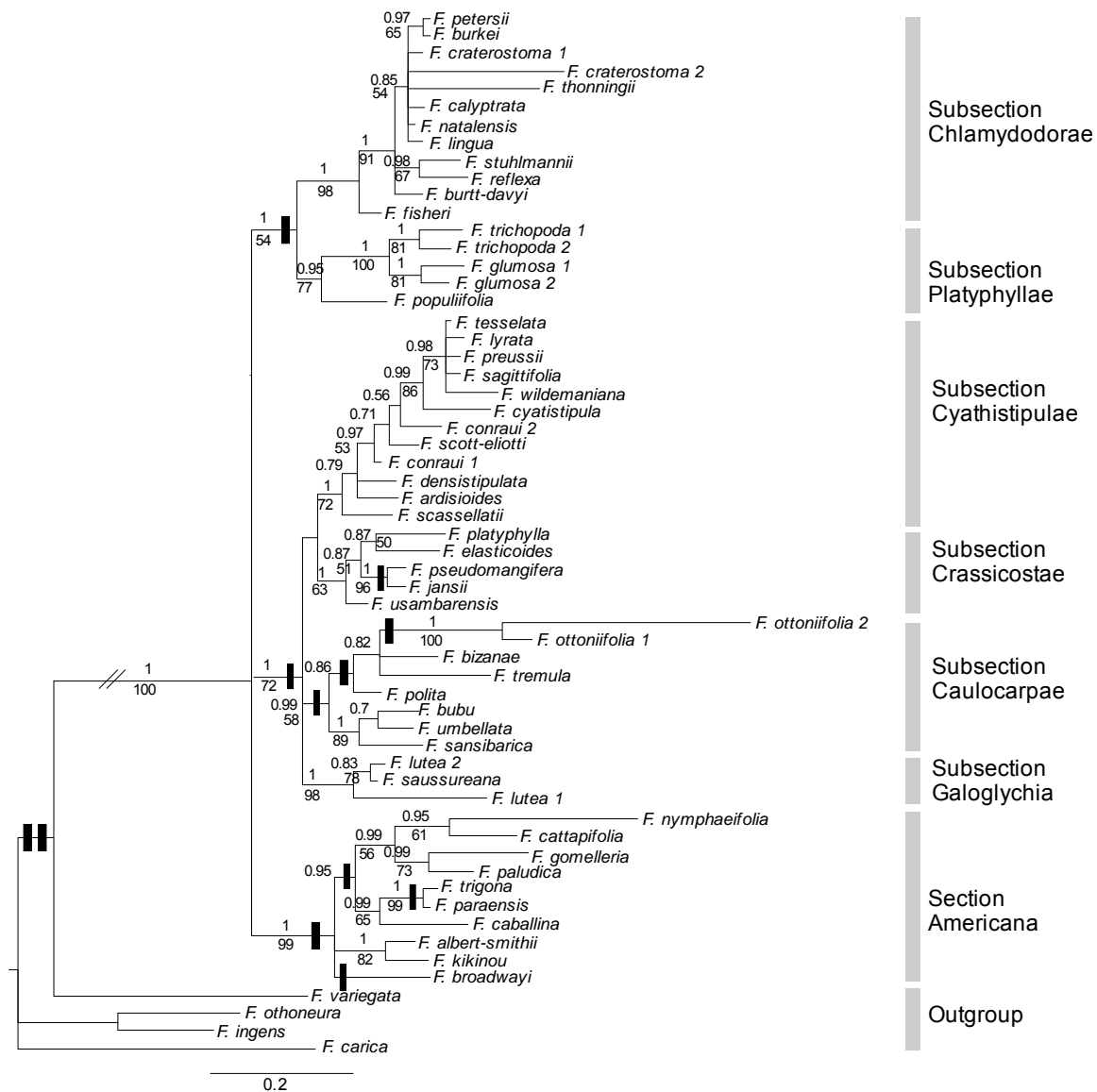


Figure 9. Arbre phylogénétique de la section *Galoglychia* incluant la section *Americana*, reconstruit à partir de marqueurs ITS et ETS par inférence bayésienne. Les valeurs au-dessus et en dessous des branches indiquent respectivement les valeurs de probabilité postérieure bayésienne et de bootstrap en maximum de vraisemblance. Les rectangles noirs indiquent la position d'indels synapomorphiques.

En outre, l'arbre nucléaire propose un branchement des sous-sections entre elles similaire à celui-ci proposé par Rønsted *et al.* (2007a) et différent de celui proposé dans l'arbre chloroplastique. En effet, dans la phylogénie nucléaire, les sous-sections *Cyathistipulae*, *Crassicostae*, *Caulocarpae* et *Galoglychia* forment un groupe monophylétique distinct d'un autre clade incluant les sous-sections *Chlamydorae* et

Platyphyllae. Cette topologie diffère de la topologie chloroplastique qui branche les sous-sections *Chlamydodora*, *Platyphyllae*, *Cyathistipulae* et *Crassicostae* dans un groupe monophylétique (le clade « CCCP »). En conclusion, les phylogénies nucléaires et chloroplastiques apparaissent discordantes à deux niveaux : terminal avec la position conflictuelle des *F. craterostoma*, *F. trichopoda* et *F. densistipulata* ; et basal avec le branchement conflictuel des sous-sections *Cyathistipulae* et *Crassicostae*.

Ces discordances apparentes peuvent toutefois être issues de processus stochastiques liés à la reconstruction phylogénétique ou à l'échantillonnage, ou encore être causées par une différence de résolution phylogénétique entre les arbres chloroplastiques et nucléaires. Ce dernier point mérite une attention particulière dans le cas du branchement conflictuel des *Cyathistipulae* et *Crassicostae* dans la mesure où les deux principaux clades basaux ne sont pas, ou très faiblement, soutenus par les valeurs de bootstraps. Il est donc nécessaire de tester la significativité de ces diverses discordances. Nous avons utilisé le test de Wilcoxon pour comparer la topologie la plus parcimonieuse à une topologie alternative et le test SH pour comparer la topologie la plus vraisemblable à une topologie alternative (Tableau 1). Le fait que la discordance impliquant *F. densistipulata* soit incluse dans la discordance basale (discordances emboîtées) complique la procédure. La méthode classique qui consiste à produire une topologie alternative en enlevant les accessions supposées discordantes sauf une et à refaire une analyse phylogénétique en forçant la discordance supposée qui reste à ne plus en être une, ne peut être appliquée dans ce cas précis. En effet, le retrait de toutes les accessions de *Cyathistipulae* entrainerait par la même occasion le retrait de *F. densistipulata*. Nous avons donc utilisé une autre méthode pour la construction de topologies alternatives, que nous avons appliquée aux quatre discordances. Pour tester chaque discordance, une topologie alternative a été reconstruite en utilisant la topologie originale et en ne déplaçant manuellement qu'une accession ou un clade de sorte à recréer un seul cas de discordance. Par exemple, la discordance basale impliquant *F. densistipulata* a été testée en comparant la longueur/vraisemblance de la topologie chloroplastique la plus parcimonieuse/vraisemblable à la longueur/vraisemblance de la topologie chloroplastique modifiée manuellement de sorte que *F. densistipulata* soit branché avec les *Cyathistipulae* en considérant le jeu de données chloroplastique. Les tests furent répétés en comparant la longueur/vraisemblance de la topologie nucléaire la plus parcimonieuse/vraisemblable à la longueur/vraisemblance de la topologie nucléaire modifiée manuellement de sorte que *F. densistipulata* soit branché comme dans la topologie chloroplastique, en considérant le jeu de données nucléaire. Pour qu'une discordance soit significative il faut que les tests soient significatifs dans les deux sens.

Tableau 1. Résultats des tests évaluant les discordances entre les phylogénies chloroplastiques et nucléaires chez les figuiers *Galoglychia*. Les degrés de significativité d'une ligne donnée correspondent au test comparant la longueur (test de Wilcoxon) ou la vraisemblance (test SH) de la topologie précisée sur la ligne avec la longueur/vraisemblance de l'arbre le plus parcimonieux/vraisemblable de topologie inscrite en caractères gras juste au dessus. Dans les description des topologies, « , » indique que le branchement des sous-sections est laissé non-résolu et « + » que le branchement entre et au sein des sous-sections est laissé non-résolu.

Topologie	Test de Wilcoxon		Test SH	
	longueur	Degré de significativité	-LnL	Degré de significativité
<i>Jeu de données chloroplastiques</i>				
Chloroplastique	143		5956.50	
Nucléaire	195	<0.001	6241.39	<0.001
Chloroplastique - <i>F. trichopoda</i> monophyletique	147	0.045	5981.30	0.021
Chloroplastique - <i>F. craterostoma</i> monophyletique	152	0.003	5999.10	0.001
Chloroplastique - <i>F. densistipulata</i> dans <i>Cyathistipulae</i>	144	0.564	5961.60	0.280
(Americana,(Galoglychia,Caulocarpae,(Cyathistipulae +Crassicostae+Chlamydodora+Platyphyllae))	226		6061.80	
(Americana,(Galoglychia+Caulocarpae+Cyathistipulae +Crassicostae),(Chlamydodora+Platyphyllae))	276	0.016	6340.26	0.005
<i>Jeu de données nucléaires</i>				
Nucléaire	539		5170.68	
Chloroplastique	741	<0.001	6091.60	<0.001
nucléaire - <i>F. trichopoda</i> 2 avec <i>Chlamydodora</i>	560	<0.001	5259.70	0.000
nucléaire - <i>F. craterostoma</i> 2 avec <i>Cyathistipulae</i> et <i>Crassicostae</i>	560	<0.001	5211.23	0.001
nucléaire - <i>F. densistipulata</i> dans <i>Chlamydodora</i>	559	<0.001	5256.40	<0.001
(Americana,(Galoglychia+Caulocarpae+Cyathistipulae +Crassicostae),(Chlamydodora+Platyphyllae))	818		5407.92	
(Americana,(Galoglychia,Caulocarpae,(Cyathistipulae +Crassicostae+Chlamydodora+Platyphyllae))	962	0.001	5890.20	0.002

La discordance basale impliquant la position des sous-sections *Cyathistipulae* et *Crassicostae* ainsi que deux des trois discordances terminales se sont révélées significatives selon les tests de Wilcoxon et les tests SH. En revanche, toujours d'après ces deux tests, la discordance impliquant *F. densistipulata* n'est pas significative lorsque les tests ont été conduits à partir des données chloroplastiques et en utilisant la topologie chloroplastique originale. Concrètement, ceci signifie que le jeu de données chloroplastiques ne rejette pas l'hypothèse que *F. densistipulata* puisse être branché au sein des *Cyathistipulae*. En

conclusion, il existe au moins trois cas de discordances cyto-nucléaires dans la phylogénie des figiers *Galoglychia* impliquant un cas de discordance basale et deux cas de discordances terminales.

Chapitre IV

LES CAUSES DE DISCORDANCES CYTO-NUCLEAIRES

Lorsqu'une discordance cyto-nucléaire est détectée puis sa significativité testée, l'identification de sa cause est une étape nécessaire à la poursuite d'une étude phylogénétique ou phylogéographique. En effet, certaines causes pourront être contrées en modifiant les méthodes d'analyse, contrairement à d'autres. Par ailleurs, certaines causes de discordance cyto-nucléaire pourront aider à mieux comprendre l'histoire des populations ou des espèces étudiées (voir Chapitre V). Les discordances cyto-nucléaires peuvent avoir des causes très variées et sont ici classées en deux catégories, selon qu'elles sont artéfactuelles et causées par l'analyste, ou causées par un phénomène biologique. On notera par ailleurs que les causes de discordances cyto-nucléaires sont similaires aux causes générant une discordance entre les histoires de deux gènes telles qu'inférées par les méthodes de reconstruction phylogénétique ou phylogéographique, et encore plus généralement aux causes de discordance entre l'histoire inférée d'un gène et l'histoire de l'espèce ou de la lignée chez qui ce gène a été étudié.

IV.A. Discordances artéfactuelles

IV.A.1. Erreurs humaines et effet d'échantillonnage

Une source évidente de discordance entre phylogénies cytoplasmiques et nucléaires est l'erreur humaine, qui englobe de multiples sources d'erreurs allant de l'erreur d'identification ou d'étiquetage, à la concaténation de séquences issues d'accessions différentes pour les alignements multi-marqueurs. Ces erreurs peuvent survenir à n'importe quelle étape de l'étude et dans n'importe quel échantillon et c'est précisément la nature stochastique de ce type d'erreur qui peut permettre leur identification. Ainsi lorsque qu'une discordance porte sur le positionnement d'un clade (c'est-à-dire plus de deux espèces monophylétiques), la

probabilité que des erreurs humaines puissent avoir affecté l'ensemble des espèces du clade par malchance est très faible. Par exemple, dans le cas des figuiers de la section *Galoglychia*, le fait que toutes les accessions de la sous-section *Cyathistipulae* soient impliquées dans une position discordante rend peu probable le fait que l'erreur humaine puisse être à l'origine de cette discordance. Par extension, lorsque la discordance porte sur le branchement d'une unique accession, la probabilité que le branchement discordant suive une prédiction ou une logique biologique est inférieure à la probabilité que le branchement se fasse à n'importe quel endroit dans l'arbre phylogénétique. Ainsi, toujours en suivant l'exemple des figuiers *Galoglychia*, au moins deux accessions (*F. trichopoda* et *F. craterostoma*) ont un branchement discordant entre les topologies chloroplastiques et nucléaires. Comme nous le verrons au Chapitre V, le fait que ces deux événements de discordance affectent le même clade (le clade CCCP) suggère une explication biologique qui soutient en retour l'idée que ces discordances ne soient pas le fait d'une erreur de la part de l'analyste.

Une deuxième source pouvant conduire à des topologies erronées est un échantillonnage insuffisant (Wendel & Doyle 1998). Van der Niet et Linder (2008) proposent une méthode pour tester si un échantillonnage limité peut être la cause des discordances observées : dans ce cas une réduction de l'échantillon ayant servi à construire les phylogénies devrait augmenter le nombre d'évènements de discordance. Le test repose sur une procédure de Jack-knife durant laquelle 20%, 40%, 60% puis 80% des accessions sont successivement retirées du jeu de données. Pour chaque niveau de Jack-knife, 10 matrices cytoplasmiques et 10 matrices nucléaires sont générées et pour chaque matrice, 100 répliques de bootstraps sont effectuées. Le nombre de cas de discordances entre les matrices cytoplasmiques et nucléaires est compté et moyenné pour chaque niveau de Jack-knife sur la base d'une valeur critique de bootstraps (par exemple 90%). Le nombre moyen de discordances est finalement divisé par le nombre total d'accessions incluses dans le niveau de Jack-knife (Fig. 10). Si un échantillonnage insuffisant est susceptible d'être une source de discordance, alors une corrélation négative entre le nombre de discordances et le nombre d'accessions sera être observée.

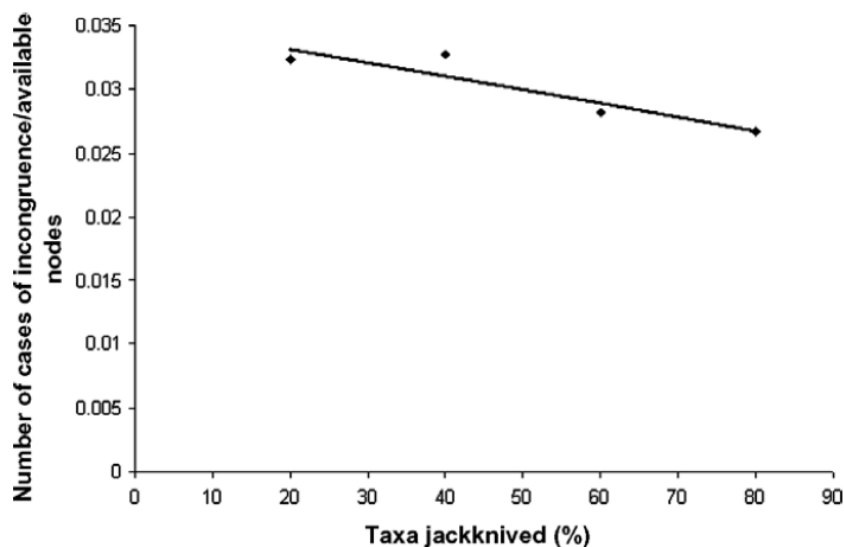


Figure 10. Relation entre le nombre de taxons échantillonnés et le nombre de cas de discordances cyto-nucléaires dans l'étude de Niet et van Linder (van der Niet & Linder 2008).

Ces auteurs ont appliqué leur méthode à un jeu de données sur la phylogénie d'un groupe d'orchidées. Leur nombre de discordances soutenues par plus de 90% des valeurs de bootstrap était corrélé positivement au nombre de taxons échantillonnés, rejetant ainsi l'hypothèse qu'un échantillonnage insuffisant puisse être la cause des discordances cyto-nucléaires observées. Leur résultat pourrait s'expliquer par le fait que plus le nombre d'accessions enlevées est élevé, plus les chances de supprimer les accessions discordantes augmentent.

Cette méthode paraît séduisante lorsqu'à la fois le nombre d'accessions et le nombre de discordances révélées par les analyses sont relativement élevés. Dans l'exemple des figuiers *Galoglychia*, seulement trois cas de discordances ont été objectivés. Par ailleurs, la faible résolution de nombreux nœuds terminaux contraint également l'application de cette procédure de Jack-knife. Toutefois, dans cet exemple, le fait que la topologie nucléaire soit similaire à celle publiée précédemment, ceci bien que 35% des accessions soient différentes, souligne qu'à la fois la nature de l'échantillonnage et le nombre d'échantillons, ont peu de chance de rendre artificiel l'arbre obtenu, du moins le nucléaire.

IV.A.2. L'artefact d'attraction des longues branches

L'artefact d'attraction des longues branches (ALB) est un artefact des méthodes de reconstruction phylogénétique qui provoque le regroupement des accessions qui évoluent le plus rapidement sans que ce regroupement ne reflète leur vraie relation de parenté (Felsenstein 1978). Différentes méthodes ont été proposées pour détecter une ALB, plus ou moins informatives et plus ou moins applicables selon la nature du jeu de données (Bergsten 2005), dont les principales sont ici brièvement présentées.

IV.A.2.a. L'extraction des longues branches

Cette méthode repose sur le principe basique que, pour qu'une longue branche soit attirée, il faut qu'il y ait une seconde longue branche dans la phylogénie. Ainsi, si deux branches longues sont groupées ensemble dans une topologie, il est possible de faire deux analyses supplémentaires, chacune en enlevant une des deux longues branches et en laissant l'autre en place, afin d'observer le nouveau branchement de chacune des deux longues branches. Si une ALB n'affecte pas la topologie originale, dans ce cas les branches auront des places inchangées dans les nouvelles analyses. Si une ALB est à l'origine d'un groupement artificiel des deux longues branches, cette méthode renseigne sur la position réelle (c'est-à-dire sans l'artefact) des longues branches. La méthode fut proposée pour la première fois par Siddall & Whiting (1999) puis appliquée à diverses études (Pol & Siddall 2001; Bergsten & Miller 2004; Hampl *et al.* 2004). Facile à mettre en œuvre, elle permet de fournir les premiers arguments à charge ou à décharge d'une ALB et est conseillée dès que cet artefact est suspecté. On notera toutefois sa plus grande limite, qui est qu'une topologie optimale ne peut être trouvée que lorsqu'un jeu de données est complet.

IV.A.2.b. L'attraction artificielle des longues branches par le groupe externe

Wheeler (1990) nota qu'un groupe externe particulièrement éloigné pouvait se comporter comme une séquence aléatoire et pouvait ainsi avoir la faculté d'attirer artificiellement la

branche la plus longue du groupe interne. Une manière de tester ce phénomène consiste à générer artificiellement 100 séquences aléatoires ayant la même longueur que la séquence du groupe externe, puis à échanger la ou les séquences du groupe externe par les séquences aléatoires, une à une, et enfin à relancer une analyse phylogénétique pour chacun des remplacements. Si l'arbre est enraciné de la même manière que dans la topologie originale dans plus de 5% des analyses, on peut alors suspecter que le groupe externe agit comme un attracteur de longues branches. Bien entendu, le point faible de ce test, qui n'est pas négligeable, est qu'il ne permet pas de dissocier une ALB d'un cas où la branche la plus longue du groupe interne constituerait réellement la racine de ce groupe. Ceci est d'autant plus problématique que sous l'hypothèse de l'horloge moléculaire, toutes les topologies asymétriques devraient avoir les branches basales les plus longues (Heard 1992; Rogers 1993; Colless 1995). Finalement, même si le test présenté ci-dessus est peu convaincant sur le plan théorique, toutes les racines déterminées par un groupe externe distant ou aléatoire devraient être considérées comme suspectes. Une analyse complémentaire au test est la conduite d'une analyse phylogénétique excluant tout groupe externe, ce qui permet de manière très simple d'observer l'influence de ce groupe sur les branchements au sein du groupe interne (Holland *et al.* 2003).

IV.A.2.c. La simulation paramétrique

Proposée pour la première fois par Huelsenbeck *et al.* (1996), cette méthode a été appliquée à de très nombreuses études afin d'étudier le rôle potentiel d'une ALB (par exemple Maddison & Maddison 2004; Wilcox *et al.* 2004; van der Niet & Linder 2008). Le principe du test est le suivant : (i) estimer les paramètres du modèle d'évolution des séquences du jeu de données, en se basant sur un arbre dans lequel les longues branches ne sont pas des branches sœurs, (ii) produire par simulation un grand nombre (*e.g.* 100) de jeux de données à partir de l'arbre et du modèle d'évolution ci-dessus et (iii) analyser les jeux de données un à un avec la même méthode que celle utilisée dans l'analyse originale. Si les longues branches se branchent ensemble dans un nombre important d'analyses, cela signifie que les branches en question sont suffisamment longues pour s'attirer mutuellement de manière artificielle. Ce test, de même que le précédent, ne permet toutefois pas d'exclure la possibilité que deux longues branches soient réellement des branches sœurs. Cette méthode reste cependant très

intéressante et un module du logiciel Mesquite (Maddison & Maddison 2004) en a automatisé la procédure.

IV.A.2.d. La discordance méthodologique

Cette dernière manière de suspecter une ALB repose sur le fait que toutes les méthodes de reconstruction phylogénétique ne sont pas également sensibles au problème de l'ALB. Ainsi, les méthodes de reconstructions basées sur la parcimonie sont plus sensibles à l'ALB que les méthodes prenant en compte des taux d'évolutions ou des longueurs de branches inégaux (Pol & Siddall 2001; Swofford *et al.* 2001). Concrètement, une phylogénie reconstruite par méthode de Maximum de vraisemblance ou par Inférence Bayésienne aura moins de chance de subir une ALB. Ainsi, la comparaison de topologies obtenues par différentes méthodes de reconstruction phylogénétique peut permettre la détection d'une ALB.

Cependant, des analyses par méthode de distance ou par maximum de vraisemblance peuvent également être sujettes à l'ALB, en particulier lorsqu'une hypothèse faite sur un modèle d'évolution est violée (Chang 1996; Gaut *et al.* 1996; Lockhart *et al.* 1996). La recherche d'une discordance méthodologique n'est donc pas une méthode infaillible pour l'identification d'une ALB.

IV.B. Discordances biologiques

IV.B.1. Paralogie

La duplication d'un gène conduit à la présence d'un gène orthologue (le gène d'origine) et d'une ou plusieurs copies paralogues. Pouvant être physiquement disjointes, les copies paralogues et orthologues peuvent subir des pressions évolutives différentes (Doyle 1992). Par ailleurs, des mutations peuvent survenir dans les copies orthologues et conduire à leur extinction ou à l'impossibilité de les amplifier par PCR. Au final, si des séquences paralogues sont incluses sans le savoir dans une reconstruction phylogénétique, l'arbre obtenu mélangera des informations sur l'histoire des organismes hébergeant le gène et sur les l'histoire des événements de duplication (Wendel & Doyle 1998).

Le problème lié à l'amplification de copies paralogues en reconstruction phylogénétique peut être particulièrement important avec certains marqueurs. Par exemple, la région de l'ITS de l'ARN ribosomal, très largement employée chez les plantes, est connue pour être présente sous de multiples copies dans de nombreuses lignées (Hartmann *et al.* 2001; Ko & Jung 2002). Buckler *et al.* (1997) ont ainsi démontré la présence de copies paralogues d'ARN ribosomal chez *Gossypium*, *Nicotiana*, *Tripsacum*, *Zea* et dans la famille des Winteraceae. C'est ainsi que certains auteurs soutiennent que l'évolution imprévisible de l'ITS est suffisamment contraignante pour que ce marqueur ne soit pas un bon outil en reconstruction phylogénétique (Alvarez & Wendel 2003).

Comme précisé au Chapitre I, la duplication de gènes cytoplasmiques est un phénomène exceptionnel. Le problème associé à l'amplification de séquences paralogues devra alors se poser pour les jeux de données issus de marqueurs nucléaires. Il est notable que, bien souvent, un raisonnement logique basé sur la comparaison de la phylogénie nucléaire avec la morphologie ou la distribution géographique, rend évident le fait que la phylogénie nucléaire n'est pas responsable de la discordance. Ceci permet de rejeter d'emblée l'hypothèse de la paralogie.

Par exemple, des arguments taxonomiques et biogéographiques forts soutiennent l'idée que c'est la phylogénie chloroplastique et non nucléaire qui est source de conflit dans le cas des figuiers *Galoglychia*, malgré un faible soutien statistique des nœuds nucléaires. Premièrement, la monophylie des sous-sections au sein de la section *Galoglychia* suit parfaitement la taxonomie classique (Mildbraed & Burret 1911; Berg & Wiebes 1992). Deuxièmement, les deux principaux clades à l'intérieur de la section *Galoglychia* sont concentrés sur des régions floristiques distinctes du continent Africain (Rønsted *et al.* 2007a). Le premier clade, incluant les sous-sections *Platyphyllae* et *Chlamyodorae* se concentre essentiellement sur l'Afrique de l'est et s'étend sur Madagascar et les îles voisines (Comores, Mascareignes, Aldabra et Seychelles). De nombreuses espèces du clade (surtout des *Platyphyllae*) se trouvent en savane et sont adaptées à un climat sec. Les espèces de ce clade spécifique à la forêt humide sont peu nombreuses (ex. : *F. calyptrata* et *F. recurvata*) et quelques autres peuvent vivre à la fois en forêt humide et en forêt sèche (*F. natalensis*, *F. craterostoma* et *F. thonningii*; Berg & Wiebes 1992). Le second clade, incluant les sous-sections *Caulocarpae*, *Galoglychia*, *Cyathistipulae* et *Crassicostae*, est essentiellement centré sur l'Afrique de l'ouest et l'Afrique centrale. Par ailleurs, une majorité d'espèces est liée au milieu forestier humide. A la vue de ces deux types d'arguments (biogéographiques et

taxonomiques) il est très improbable que la phylogénie nucléaire ne reflète pas l'histoire évolutive des espèces de figuiers *Galoglychia*.

Par ailleurs, nous avons compté le nombre d'événements de duplication et d'extinction nécessaires pour expliquer le patron de discordance cyto-nucléaire, à l'aide du logiciel GeneTree 1.3.0 (Page & Charleston 1997; Page 1998). Le résultat obtenu montre que si on considère que les fragments ITS et ETS constituent une seule unité évolutive et si on exclut tout événement d'introgession, alors un évènement de duplication à la base de la section *Galoglychia* suivi de trois évènements d'extinction, chacun à la base d'une paire de sous-sections, sont nécessaires pour retrouver un branchement des sous-sections tel qu'observé dans la phylogénie chloroplastique (Fig. 11).

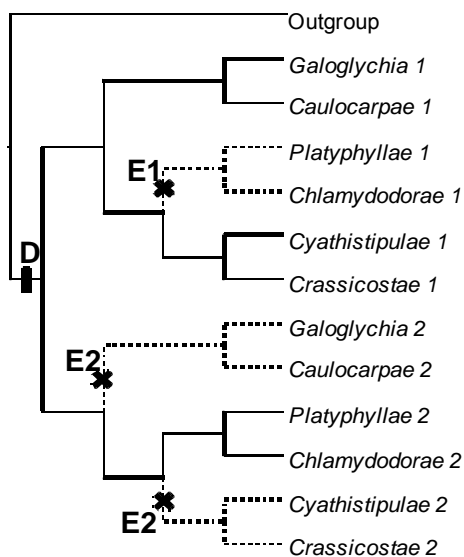


Figure 11. La discordance cyto-nucléaire chez les *Galoglychia* expliquée par la duplication et l'extinction d'ITS et ETS. Le scénario nécessite un évènement basal de duplication (D) suivi d'un évènement d'extinction de l'allèle E1 et de deux évènements d'extinctions de l'allèle E2.

Le nombre d'évènements de duplication et d'extinction nécessaires à la réconciliation des topologies nucléaires et cytoplasmiques n'est pas déraisonnable. Toutefois, la monophylie des sous-sections est difficilement compatible avec l'amplification de copies paralogues d'ITS et d'ETS qui évoluent généralement de manière indépendante (Alvarez & Wendel 2003). Sous l'hypothèse de paralogie, les extinctions auraient dû systématiquement se produire chez les ancêtres des sous-sections et non chez ses descendants, ce qui est peu vraisemblable.

IV.B.2. Le tri incomplet du polymorphisme ancestral

Les risques de discordance entre arbres de gènes qui résultent de la nature stochastique du processus de coalescence ont été documentés depuis plusieurs décennies (voir, par exemple Tatenò *et al.* 1982). Le tri du polymorphisme ancestral étant théoriquement complet après $4N_e$ générations (N_e correspondant à la taille efficace de la population), la différence de taille efficace des gènes nucléaires et cytoplasmiques aboutit au fait que le tri ne sera pas complet au même moment pour les gènes de ces différents compartiments cellulaires. C'est ainsi que, comme il l'a été précisé précédemment, les allèles des gènes cytoplasmiques deviennent monophylétiques avant les allèles des gènes nucléaires (Pamilo & Nei 1988). Les arbres phylogénétiques reconstruits à partir de marqueurs nucléaires ont donc davantage de risques de présenter de la polyphylie ou de la paraphylie liée à un tri incomplet du polymorphisme ancestral que les arbres reconstruits à partir de marqueurs cytoplasmiques (Funk & Omland 2003). De surcroît, les risques de discordances liées au tri incomplet du polymorphisme ancestral sont plus importants lorsque des radiations d'espèces récentes sont étudiées (Moran & Kornfield 1993, 1995; Klein & Payne 1998).

L'attribution d'un cas de discordance cyto-nucléaire à du tri incomplet de polymorphisme ancestral est une tâche périlleuse. En effet, ce phénomène peut générer des patrons de discordance similaires à ceux produits par de l'hybridation (Holder *et al.* 1999). Des méthodes basées sur des simulations de coalescents ont récemment été développées pour tester l'hypothèse du tri incomplet du polymorphisme ancestral (Buckley *et al.* 2006). Le principe général de ces méthodes repose sur la comparaison du patron de discordance entre plusieurs arbres de gènes indépendants. Prenons pour exemple une étude dans laquelle trois gènes nucléaires et un gène mitochondrial sont séquencés. L'hypothèse nulle du test est l'hypothèse de tri incomplet du polymorphisme ancestral selon laquelle le patron de discordance entre les quatre gènes est indépendant (puisque le processus de tri des allèles est purement aléatoire pour chaque gène). L'hypothèse alternative est l'hypothèse d'hybridation introgressive, selon laquelle le patron de discordance entre gènes n'est pas indépendant. La première étape du test consiste à utiliser l'arbre des espèces (une accession par espèce choisie au hasard), avec des longueurs de branches optimisées et une profondeur d'arbre imposée, afin de reproduire par simulation 100 ensembles de 4 arbres des gènes (un ensemble par gène réellement séquencé). Le processus de simulation part de chaque copie du gène et reproduit la coalescence de ces copies de gènes selon un modèle simple de panmixie, à taille efficace de population et temps de génération fixé. Les simulations peuvent être aisément effectuées à

l'aide du logiciel Mesquite (Maddison & Maddison 2004). Ensuite, le nombre d'ensembles parmi les 100 où le taxon (ou clade) discordant est branché différemment dans deux arbres sur quatre (ce qui correspond à l'hypothèse alternative de distribution non aléatoire du patron de discordance) doit être compté. Sous l'hypothèse nulle, une fréquence de patrons non aléatoires n'excédant pas 5% est attendue. Si la fréquence de patrons non aléatoires dépasse ce seuil, alors l'hypothèse de tri incomplet du polymorphisme ancestral est rejetée.

Il est intéressant de noter que ce test basé sur des simulations nécessite la connaissance *a priori* de la taille efficace de la population ancestrale, de la profondeur de l'arbre des espèces ainsi que du temps de génération. La taille efficace N_e d'une population est très difficilement déductible d'après de seules connaissances sur la biologie de l'espèce. Cette taille est calculable en utilisant des modèles de coalescence et un large échantillon d'individus appartenant à la population pour laquelle N_e doit être estimée (Wang & Whitlock 2003; Wang 2009). Ce type d'échantillon est généralement différent de celui utilisé en reconstruction phylogénétique. Ainsi, N_e doit généralement être extraite de la littérature, à partir d'études de génétique des populations appartenant à la lignée étudiée. Par ailleurs, si une estimation précise de N_e pour une population actuelle est parfois délicate, son estimation pour la population ancestrale l'est encore plus car une fraction seulement de la variation génétique de la population actuelle est informative pour prédire le passé (Wang 2005). De plus, les méthodes utilisées pour estimer la N_e ancestrale nécessitent davantage d'hypothèses dont certaines ne sont très vraisemblablement pas valides, comme l'hypothèse d'horloge moléculaire ou l'hypothèse d'absence totale de recombinaison intra-génique. Pour ces raisons, il est généralement recommandé de répéter le test en faisant varier N_e selon un facteur allant jusqu'à 10^6 . En outre, un seul arbre des espèces est utilisé pour simuler des coalescents de gènes ayant des N_e vraisemblablement différents (par exemple entre gènes mitochondriaux et nucléaires). Un second problème associé à l'emploi de ce test est la connaissance de la profondeur de l'arbre. Il est possible d'attribuer une profondeur sur la base de données fossiles, toutefois, ceci suppose que l'arbre des espèces a une profondeur égale à l'arbre des gènes. En réalité, le coalescent de l'arbre des gènes est vraisemblablement plus profondément que celui de l'arbre des espèces (Edwards & Beerli 2000). Le temps de génération, ou âge moyen des individus femelles à la reproduction, est habituellement le mieux connu des paramètres à définir *a priori*. Cependant, tout comme la taille efficace de la population, la méthode ne permet pas de faire varier le temps de génération entre les différentes parties de l'arbre. Or, lorsque de nombreuses espèces ou populations issues de la radiation d'une lignée sont étudiées, la diversité des forces de sélection à l'origine de la radiation aboutit

fréquemment à des formes très variées au sein de cette lignée (*e.g.* Losos *et al.* 1998). La taille étant par ailleurs fréquemment corrélée au temps de génération (Calder 1984), le temps de génération est susceptible de varier entre espèces ou populations de la lignée.

Afin de déterminer si le tri incomplet du polymorphisme ancestral est source de discordance entre les phylogénies chloroplastiques et nucléaires chez les figuiers *Galoglychia*, cette méthode pourrait paraître appropriée. Toutefois, la section *Galoglychia* renferme des espèces issues de milieux très différents, allant de la savane de plaine (*e.g.* *Ficus platyphylla*) à la forêt équatoriale montagneuse (*e.g.* *F. amadiensis*), ce qui conduit vraisemblablement à une importante variabilité de la taille efficace au sein de la section. En outre, la section accueille des formes de vie très variées, allant d'arbres émergents de la canopée (principalement dans la section *Caulocarpae*) à de petits arbustes de sous-bois, ou perchés en haut d'autres arbres (principalement dans les sous-sections *Chlamydodora* et *Cyathistipulae*). La taille des arbres est un facteur connu pour influencer la structure génétique spatiale (données non publiées sur les figuiers) et donc la taille efficace de la population, ainsi que le temps de génération (King 1996). De nombreuses conditions nécessaires à l'application du test, qui sont basées sur la simulation de coalescents, ne sont donc pas remplies avec le jeu de données sur les figuiers *Galoglychia*. Plus généralement, ces conditions seront rarement remplies dans les études phylogénétiques où les taxons étudiés sont le fruit d'une diversification importante. Tant que les effets des écarts aux conditions d'application n'auront pas été étudiés et objectivement évalués, l'interprétation des résultats issus du test devrait être empreinte de précautions.

Concernant la discordance cyto-nucléaire faisant intervenir la position des *Cyathistipulae* et *Crassicostae*, il est notable que dans la phylogénie chloroplastique, la branche supportant le clade CCCP (longueur = 0.1774) est plus longue que les branches supportant des clades chez qui la monophylie réciproque a été atteinte, comme le clade des *Caulocarpae* (0.1699), des *Galoglychia* (0.1269) ou encore des *Americana* (0.1756). Ainsi, on peut supposer que le temps écoulé avant l'apparition d'événements de spéciation au sein du clade CCCP a été suffisamment long pour avoir permis un tri complet du polymorphisme ancestral. Ce raisonnement souffre du fait qu'il suppose une corrélation positive entre le nombre de générations et la longueur de branche, ainsi qu'un N_e constant. Concernant la première supposition, le taux de mutation est corrélé (bien qu'imparfaitement) au temps de génération (Nabholz *et al.* 2008), ce qui signifie qu'il existe une relation indirecte entre le nombre de générations et la longueur de branche. Concernant la constance de N_e au travers du temps pour la lignée étudiée, nous venons de voir que cette hypothèse était peu

vraisemblable. Finalement, ce raisonnement, basé sur la comparaison de longueurs de branches est très simple par rapport au test de simulation et repose, tout comme celui-ci, sur l'hypothèse de N_e constant, mais ne repose pas sur d'autres connaissances requises *a priori* comme la valeur du temps de génération et la valeur de N_e . C'est pourquoi, en dépit de son évidente limitation, nous avons utilisé ce raisonnement pour discuter l'hypothèse du tri incomplet du polymorphisme ancestral. Au final, bien qu'il nous ait été impossible de formellement exclure cette hypothèse, elle ne représente pas l'hypothèse favorite pour expliquer la discordance cyto-nucléaire mettant en jeu le branchement des *Cyatistipulae* et *Crassicostae*.

D'autres arguments logiques peuvent être mis en avant pour rejeter l'hypothèse du tri incomplet du polymorphisme ancestral. Par exemple, dans la phylogénie mitochondriale des lézards *Podarcis* du sud-est ibérique, la lignée Valencia est incluse au sein d'un grand clade accueillant les lignées mitochondriales africaines et ibériques de *P. vaucheri*. Un autre clade important regroupe les lignées Liolepis et Hispanicus. L'ancêtre commun à ces deux clades est daté à 7,2 millions d'années (Pinho *et al.* 2006). Il est raisonnable de supposer que ce temps est suffisamment long pour invoquer un tri complet du polymorphisme ancestral, à moins que les populations n'aient des tailles efficaces démesurément importantes, de l'ordre de 2 à 7 millions pour des temps de génération allant de 1 à 3 ans. De plus, sous l'hypothèse d'un tri incomplet, il serait largement improbable de retrouver des haplotypes de Valence dans *P. hispanicus* et *P. liolepis* mais aucune trace des lignées mitochondriales de Type 2, Type 1 et *P. bocagei*. Finalement, dans cet exemple aussi, l'hypothèse du tri incomplet du polymorphisme ancestral semble peu probable pour expliquer la présence de la lignée mitochondriale Valencia dans *P. hispanicus* et *P. liolepis*.

IV.B.3. Hybridation introgressive

La dernière source de discordance entre les arbres des différents gènes d'une espèce peut provenir de flux de gènes inter-spécifiques. Chez les eucaryotes, l'hybridation, ou croisement entre groupes ou taxons génétiquement distinguables produisant des hybrides viables, est une première étape permettant de tels flux de gènes. L'hybridation est un phénomène commun tant chez les plantes que chez les animaux (Arnold 1997). Vingt-cinq pourcent des plantes de la flore britannique sont connues pour pouvoir s'hybrider avec au moins une autre espèce

(Stace 1975, 1997), tout comme 9,3% des oiseaux de la planète (Grant & Grant 1992), 11% des papillons européens (Guillaumin & Descimon 1976) ou 6% des mammifères européens (MacDonald & Barrett 1995). Toutefois, un évènement d'hybridation correspond généralement à emmener des génomes parentaux au centre d'une zone hybride, où les loci neutres sont en déséquilibre de liaison avec des loci sous sélection, empêchant ainsi la libre diffusion de ces loci neutres à l'intérieur de l'une ou l'autre des espèces (Barton & Hewitt 1985, 1989). Les effets de l'hybridation s'arrêtent alors là. Ces loci neutres peuvent cependant s'extraire de cet environnement génétique défavorable grâce au mécanisme de recombinaison (Martinsen *et al.* 2001). S'ils se retrouvent dans un environnement génétique favorable, sélectionné positivement, ou un environnement complètement neutre, alors les allèles étrangers de ces loci pourront diffuser à l'intérieur de l'espèce d'accueil par l'intermédiaire de croisements entre les hybrides et l'un des parents (Raufaste *et al.* 2005). L'invasion d'un génome par du matériel génétique étranger est appelé introgression (Mallet 2005). L'introgression peut aboutir à une discordance entre arbres de gènes en introduisant des lignées alléliques phylogénétiquement divergentes de part et d'autre des limites entre espèces (*e.g.* Patton & Smith 1994; Shaw 1999).

Les effets d'une hybridation introgressive sont particulièrement sensibles pour les loci situés sur des génomes cytoplasmiques. En effet, comme mentionné précédemment, tous les loci d'un de ces génomes peuvent introgresser comme un seul bloc du fait de l'absence de recombinaison (Smith 1992). Ainsi, l'analyse d'un seul fragment cytoplasmique révélera à lui seul l'origine hétérospécifique du génome entier auquel il appartient (Martinsen *et al.* 2001). De plus, les loci cytoplasmiques introgressent en moyenne plus profondément dans l'espèce d'accueil que les loci nucléaires, car cette introgression n'est pas contrainte par des liaisons avec d'autres loci soumis à sélection (Barton & Jones 1983; Tegelstrom 1987). La différence de flux entre gènes nucléaires et cytoplasmiques a été bien étudiée chez les *Helianthus* (Rieseberg *et al.* 1990; Rieseberg *et al.* 1991). Dans ce genre, des loci chloroplastiques ont introgressé jusqu'à 10 fois plus loin que les loci nucléaires. Des génomes mitochondriaux exogènes ont pu être notés à plusieurs centaines de kilomètres de la zone hybride actuelle, sans que des traces d'ADN nucléaire exogène n'aient été enregistrées, comme dans le cas d'une souris scandinave (Gyllensten & Wilson 1987).

La meilleure signature d'une introgression est le partage chez des individus sympatriques d'haplotypes mitochondriaux qui sont par ailleurs géographiquement localisés, chez des espèces qui divergent par leur morphologie, ou d'après d'autres gènes que les gènes mitochondriaux. Ce schéma de distribution géographique d'allèles est difficilement

interprétable autrement que par des flux de gènes interspécifiques récents ou en cours (Funk & Omland 2003). Cependant, la reconnaissance formelle d'une hybridation introgressive comme cause de discordance cyto-nucléaire est plus délicate lorsque l'événement d'introgession est ancien et que la barrière aux flux géniques entre espèces est devenue imperméable. Plus le temps écoulé depuis la fin des échanges de gènes est important, plus les lignées d'allèles introgressés peuvent avoir une position phylogénétique basale et moins elles auront de chance de révéler une association géographique avec la région d'origine de ces lignées. Par ailleurs, si le flux de gènes mitochondriaux est élevé et bidirectionnel, il est parfois impossible de retrouver quelle est la lignée d'origine de chacune des espèces. A son extrême, la lignée d'origine peut s'éteindre ce qui ne laisse plus d'indice pour révéler l'origine hétérosécificque de la lignée restante (Bernatchez *et al.* 1995). Ce phénomène est facilité par la faible taille efficace des loci cytoplasmiques (Takahata & Slatkin 1984).

Dans bien des cas, l'hypothèse d'introgession sera l'hypothèse emise par défaut pour expliquer une discordance soutenue par des arguments écologiques ou biogéographiques. Dans leur étude sur les discordances cyto-nucléaires au sein du genre d'orchidées *Satyrium*, Van der Niet et Linder (2008) discutent l'hypothèse d'introgession sur la base de la distribution géographique actuelle des espèces. Pour chaque cas de discordance, le nombre d'évènements de dispersion et d'extinction sont comptés et la vraisemblance est discutée. Cette méthode présente des limites liées à son importante subjectivité et est par ailleurs difficilement transférable à n'importe quel modèle d'étude. Le Chapitre V montre comment des arguments écologiques et biogéographiques ont pu être exploités pour soutenir l'hypothèse d'hybridation introgressive dans le cas des *Podarcis* ibériques, ainsi que chez les figuiers de la section *Galoglychia*.

Chapitre V

LES DISCORDANCES CYTO-NUCLEAIRES : UN OUTIL POUR COMPRENDRE L'HISTOIRE DES ESPECES

Les discordances cyto-nucléaires sont souvent perçues comme un piège en reconstruction phylogénétique et phylogénomique qu'il convient de détecter et d'éviter afin de retracer correctement l'histoire du vivant (Degnan & Rosenberg 2009). Les discordances cyto-nucléaires peuvent également être considérées comme un outil servant à mieux connaître l'histoire et la biologie des espèces, comme l'isolement reproducteur au niveau des zones hybrides (Rieseberg *et al.* 1990; Rieseberg *et al.* 1991), la distribution géographique passée des espèces (Melo-Ferreira *et al.* 2005; Melo-Ferreira *et al.* 2007) ou encore l'infection d'arthropodes par des endoparasites (Jiggins 2003; Hurst & Jiggins 2005). Dans ce chapitre, nous allons développer deux exemples d'exploitation de discordances cyto-nucléaires. Le premier cas, portant sur les lézards ibériques du genre *Podarcis*, utilise de manière classique les résultats d'un patron d'introgession mitochondriale pour comprendre la distribution passée des espèces et leurs interactions dans les zones de contact. Le deuxième cas utilise de manière plus originale la découverte de traces d'introgession chloroplastique chez les figuiers de la section *Galoglychia* pour étudier l'histoire des interactions entre plantes hôtes et guêpes pollinisatrices.

V.A. Histoire évolutive des lézards ibériques *Podarcis spp.* du sud-est de la Péninsule Ibérique

V.A.1. L'âge maximal de l'introgession mitochondriale

La lignée mitochondriale Valencia qui a introgressé à la fois *P. hispanicus* et *P. liolepis*, est présente sous la forme de trois haplotypes chez les individus échantillonnés et pour le fragment de la région de contrôle séquencé. Un haplotype Va est présent chez tous les

individus *P. liolepis* de la région située autour de la ville de Valence, excepté dans une localité où les deux individus séquencés présentaient un haplotype Vb différant de Va par une unique mutation. La population de *Podarcis hispanicus* localisée dans la Sierra Maria de Nieva présente un unique haplotype Vc différant de Va par deux substitutions.

En dépit du nombre limité de mutations séparant les haplotypes mitochondriaux Valencia présents chez *P. liolepis* de ceux présents chez *P. hispanicus*, nous avons calculé le temps de divergence entre ces deux groupes. Si l'on fait l'hypothèse que les lignées Valencia ont été isolées après qu'elles aient introgressé leurs espèces hôtes (ce qui n'est pas improbable étant donné l'absence de flux de gènes mitochondriaux entre *P. liolepis* et *P. hispanicus*) alors le temps de divergence calculé représente l'âge maximal possible des deux événements d'introgession. Le temps de divergence a été estimé par méthode de coalescence avec le programme MDIV (Nielsen & Wakeley 2001). Le taux d'évolution de la région de contrôle de l'ADNmt chez les reptiles n'est pas bien connu. Le fragment séquencé évolue au même taux que le *cytochrome b* chez un genre proche (Crochet *et al.* 2004) mais pourrait évoluer plus rapidement chez les *Podarcis* (Pinho, pers. com.). D'un autre côté, le temps de divergence est probablement surestimé parce que nos populations étudiées ne satisfont vraisemblablement pas l'hypothèse d'égalité des tailles efficaces (Chapman *et al.* 2007). Afin de prendre en compte toutes ces incertitudes, nous avons estimé le temps de divergence avec deux taux d'évolution différents encadrant vraisemblablement le taux d'évolution réel : 1% et 2% de divergence par million d'années.

Le temps de divergence entre les haplotypes mitochondriaux Valencia présents chez *P. liolepis* et *P. hispanicus* a été estimé à 120 000-240 000 ans. Cet âge correspond à une période du Pléistocène durant laquelle le climat subissait des variations cycliques (Avice 2000). Durant les phases froides de cette période, la Péninsule Ibérique jouait un rôle de refuge glaciaire (Weiss & Ferrand 2006). L'alternance de phases d'expansion et de rétractation des aires de distribution des espèces, causée par ces variations climatiques, a été responsable d'évènements répétés de mise en contact puis de séparation de nombreuses unités évolutives (Hewitt 1996; Taberlet *et al.* 1998). Les phases de séparation ont généré la mise en place de divergences génétiques entre lignées parfois accompagnées d'un début d'isolement reproducteur, tandis que les phases de mise en contact ont été propices aux hybridations entre des lignées fraîchement différenciées et incomplètement isolées sur le plan reproductif (Holder *et al.* 1999). Il est donc très probable que ce soit dans ce contexte que la mise en contact entre le taxon à l'origine de la lignée Valencia et d'un côté *Podarcis liolepis*, de

l'autre *P. hispanicus*, ait eu lieu et ait pu permettre l'introduction de la lignée mitochondriale Valencia chez ces deux espèces.

Le calcul du temps de divergence présenté ici repose sur l'hypothèse que le taxon à l'origine de la lignée Valencia se soit hybridé indépendamment avec *P. liolepis* et *P. hispanicus*. De manière alternative, l'hybridation aurait pu se produire avec l'une des deux espèces seulement. L'hybridation entre *P. liolepis* et *P. hispanicus* aurait par la suite permis l'introduction de la lignée Valencia dans l'autre espèce. Cette hypothèse de transmission indirecte d'une lignée mitochondriale a été par exemple proposée pour expliquer la présence d'ADNmt du lièvre variable *Lepus timidus* chez le lièvre commun *L. europaeus* (Melo-Ferreira *et al.* 2005; Melo-Ferreira *et al.* 2007). Durant les périodes glaciaires du Quaternaire, le lièvre variable était en contact avec le lièvre commun et le lièvre ibérique *L. granatensis* dans le nord de l'Espagne. Ces deux dernières espèces étaient également en contact dans cette région. L'absence d'allèles autosomaux de lièvre variable chez le lièvre commun en Péninsule Ibérique suggère que ces deux espèces ne s'hybridaient pas, à l'inverse du lièvre ibérique, avec le lièvre commun. Ainsi, l'ADNmt du lièvre variable présent chez le lièvre commun aurait pu être transmis par l'intermédiaire du lièvre ibérique. Choisir entre l'hypothèse de deux hybridations indépendantes ou l'hypothèse d'une transmission indirecte n'est pas évidente dans le cas des lézards ibériques du fait que le taxon à l'origine de la lignée de Valence n'a pas encore été identifié, mais également à cause de la faible variabilité du fragment d'ADNmt étudié. Toutefois, si la lignée mitochondriale Valencia avait été transmise indirectement, il est probable que cette lignée aurait été présente de manière continue entre la Sierra Maria de Nieva et la zone de contact entre *P. liolepis* et *P. hispanicus*, ce qui n'est pas le cas. L'hypothèse de deux événements indépendants d'hybridation est donc celle que nous soutenons le plus.

V.A.2. Origine de la distribution de la lignée mitochondriale Valencia chez P. liolepis

La présence d'une lignée mitochondriale étrangère à l'intérieur d'une espèce n'est pas un phénomène rare, comme signalé précédemment. Toutefois, deux observations sont frappantes avec l'exemple des lézards ibériques : (i) le fait que cette lignée soit la seule représentée dans une vaste aire géographique et (ii) la quasi absence de polymorphisme allélique au sein de la lignée Valencia. Le second point peut être objectivé en comparant la diversité allélique entre

les différentes lignées mitochondriales du complexe d'espèces de *Podarcis* ibériques (Tableau 2) : la diversité haplotypique estimée dans la lignée Valencia présente chez *P. liolepis* est effectivement la plus basse des diversités haplotypiques estimées chez les lignées du complexe d'espèces présentes dans la région. Deux hypothèses principales peuvent être avancées pour expliquer cette étonnante distribution de la lignée Valencia chez *P. liolepis* : l'hypothèse démographique et l'hypothèse sélective (Fig. 12).

Tableau 2. Estimations de la diversité des séquences dans les lignées mitochondriales du complexe d'espèces de *Podarcis hispanicus* du sud-est de la Péninsule Ibérique. Hisp, *Hispanicus*; Lio, *Lirolepis*; Val, Valencia; Val (*P. liolepis*), ANDmt Valencia présent chez *P. liolepis*. n_s , nombre de localités échantillonnées; n_i , nombre d'individus analysés; n_h , nombre d'haplotypes d'ADNmt observés; π , diversité haplotypique. L'écart-type est donné entre parenthèses.

Group	n_s	n_i	n_h	h	π (%)
Hisp, Type 2, Lio, Val	39	78	21	0.851 (0.028)	1.8 (1.1)
Hisp	8	12	5	0.758 (0.093)	2.1 (1.3)
Type 2	7	19	5	0.298 (0.133)	0.9 (0.7)
Lio	6	13	7	0.864 (0.079)	3.5 (1.9)
Val	18	34	4	0.458 (0.094)	0.8 (0.5)
Val(<i>P. liolepis</i>)	15	26	2	0.153 (0.092)	0.2 (0.2)

V.A.2.a. Hypothèse démographique

La distribution actuelle de la lignée Valencia pourrait représenter l'ancienne distribution du taxon à l'origine de cette lignée. *P. liolepis*, situé au nord, aurait alors envahi ce taxon. Une compétition directe entre *P. liolepis* et le taxon Valencia, accompagnée d'hybridations sur le front de colonisation, aurait pu balayer le génome nucléaire de ce dernier mais pas le génome mitochondrial.

Des modèles théoriques récents soutiennent fortement l'idée que, lorsque qu'une espèce indigène est envahie par une espèce colonisatrice, l'introgession d'allèles neutres se produit presque exclusivement de l'espèce indigène vers l'espèce colonisatrice (Currat & Excoffier 2004; Currat *et al.* 2008). Ce phénomène s'amplifie au fur et à mesure que l'espèce colonisatrice progresse, ce qui conduit finalement à des fréquences très élevées d'allèles indigènes dans l'espèce colonisatrice. Le phénomène démographique décrit par ces modèles pourrait ainsi aisément expliquer le schéma observé chez les lézards ibériques.

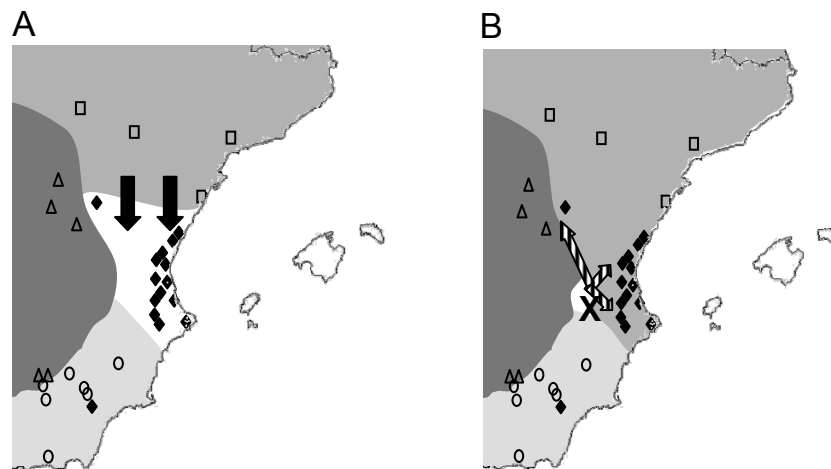


Figure 12. Hypothèse démographique (A) et hypothèse sélective (B) expliquant la distribution actuelle de la lignée mitochondriale Valencia. Les carrés, triangles, cercles et losanges représentent respectivement la lignée mitochondriale Liolepis, Type 2, Hispanicus et Valencia. Le fond de carte représente la distribution supposée des taxons pour chaque hypothèse : gris clair = *P. hispanicus* ; gris foncé = *Podarcis. sp. « type 2 »* ; gris intermédiaire = *P. liolepis* ; blanc = *Podarcis. sp. « Valencia »*. Les flèches noires représentent la colonisation de *P. liolepis* dans l'aire de distribution de *Podarcis. sp. « Valencia »*. Les flèches hachurées représentent la diffusion de la lignée mitochondriale Valencia par balayage sélectif, à la suite d'un évènement d'hybridation unique représenté par une croix.

Cependant, ces mêmes simulations montrent que la fixation, c'est-à-dire le remplacement complet de l'ADNmt indigène dans l'espèce colonisatrice, ne peut se produire qu'à la suite d'un grand nombre d'évènements d'hybridation (Currat *et al.* 2008). Les résultats de ces simulations sont soutenus par des données empiriques sur des remplacements d'espèces par compétition. En effet, les lignées mitochondriales indigènes (donc capturées) sont généralement polyphylétiques et la lignée mitochondriale colonisatrice persiste dans son espèce d'origine, au moins à faible fréquence (Melo-Ferreira *et al.* 2005; Berthier *et al.* 2006; Melo-Ferreira *et al.* 2007). Roca *et al.* (2005) décrit un cas intéressant de remplacement de l'éléphant de forêt *Loxodonta cyclotis* par l'éléphant de savane *L. africana*. Dans les écotones forêt-savane, les deux espèces d'éléphants entrent en contact et peuvent ainsi s'hybrider. Toutefois, l'hybridation est unidirectionnelle, les femelles d'éléphants de forêt s'hybrident avec les mâles d'éléphants de savane qui sont plus robustes et plus dominants que les mâles d'éléphants de forêt. Des croisements réciproques entre les hybrides femelles et les mâles de

savane ont fait disparaître toute trace de génome nucléaire appartenant à l'éléphant de forêt dans les populations hybrides, mais pas les traces de génome mitochondrial. Ainsi, l'ADNmt de l'éléphant de forêt est aujourd'hui présent dans plusieurs populations d'éléphants de savane situées à l'ancien emplacement de l'écotone forêt-savane. Toutefois, l'ADNmt de l'éléphant de savane persiste à fréquence élevée au sein de telles populations et cohabite avec l'ADNmt de l'espèce colonisée : l'éléphant de forêt.

Dans le cas des lézards ibériques du sud-est de l'Espagne, aucun individu séquencé dans la région de Valence ne présente l'ADNmt de la lignée Liolepis. Par ailleurs, la présence de deux haplotypes, ne différant que par une seule mutation dans la lignée Valencia distribuée autour de Valence, est bien évidemment contradictoire avec une éventuelle polyphylie des haplotypes Valencia présents chez *P. liolepis*. Afin de valider cette observation, il conviendrait de séquencer un fragment d'ADNmt plus variable que le fragment de la région de contrôle et d'étudier les relations de parenté entre les haplotypes Valencia présents chez *P. liolepis* avec les haplotypes présents dans le taxon à l'origine de cette espèce. La découverte du taxon à l'origine de la lignée Valencia est donc critique à ce stade de la recherche.

Bacquet (2008) a séquencé un fragment du gène ND4 chez des individus de *P. liolepis* possédant la lignée Valencia et chez des individus possédant la lignée Liolepis. Le fragment est réputé plus variable que la région de contrôle de l'ADNmt. En utilisant le test D (Tajima 1989) et le test H (Fay & Wu 2003), l'auteur n'a pas pu mettre en évidence d'écart à l'équilibre mutation-dérive au sein des haplotypes Valencia, ce qui soutient une stabilité démographique des populations de *P. liolepis* hébergeant cette lignée. Cette observation renforce les précédents arguments selon lesquels l'hypothèse démographique est peu probable pour expliquer la distribution actuelle de la lignée Valencia chez *P. liolepis*.

Des tests multi-locus ont été développés pour tester l'hypothèse sélective face à l'hypothèse démographique. Sous l'hypothèse sélective, il est attendu qu'un seul locus, celui soumis à sélection, subisse une réduction de diversité allélique, tandis que sous l'hypothèse démographique, tous les loci devraient subir une telle réduction. Galtier *et al.* (2000) ont développé un test qui, par simulation, estime l'âge (A) et la force (F) de l'évènement responsable de la perte de diversité allélique. La vraisemblance de trois modèles – modèle neutre, modèle démographique et modèle sélectif – est comparée. Les modèles non-neutres diffèrent du modèle neutre par l'existence même d'un évènement de réduction de la diversité allélique. Les modèles non-neutres diffèrent entre eux par le fait que le couple A/F ne varie pas entre locus dans le modèle démographique, au contraire du modèle sélectif. Un autre test,

développé par Hudson (1987) puis amélioré par Wright & Charlesworth (2004), est basé sur une prédiction de la théorie neutraliste de l'évolution moléculaire, qui est que la quantité de diversité allélique au sein d'une espèce est corrélée avec le niveau de divergence entre espèces. Ceci provient du fait que les deux paramètres dépendent directement du taux de mutation supposé neutre (Kimura 1983). Si le rapport entre les deux paramètres diffère entre locus, alors une évolution non neutre peut être invoquée. Bacquet (2008) a tenté d'appliquer ces tests en utilisant les séquences ND4 ainsi que les séquences des deux introns nucléaires présentées dans l'Article 3. Le test de Wright & Charlesworth n'a pu rejeter l'hypothèse du modèle d'évolution neutre pour l'ADNmt présent chez les *P. liolepis* de la région de Valence. Toutefois, ce test souffre d'un excès de conservatisme, particulièrement lorsque que les processus non-neutres sont anciens. Le test de Galtier *et al.* n'a, quant à lui, malheureusement pas pu être appliqué à cause d'un écart trop important à l'hypothèse d'absence de recombinaison intra-locus. Une investigation plus approfondie de l'hypothèse sélective devrait être prochainement mise en place, en utilisant d'autres marqueurs, à taux d'évolution plus lent que le ND4 mais plus rapide que la région de contrôle de l'ADNmt.

Enfin, des facteurs responsables d'un flux de gènes biaisé en faveur des gènes présents chez les mâles, tels que la dispersion biaisée en faveur des mâles, viendraient par ailleurs soutenir l'hypothèse démographique (Chan & Levin 2005). Chez les lézards, des cas de dispersion biaisée en faveur des mâles (Ujvari *et al.* 2007), tout comme des cas de dispersion biaisée en faveur des femelles (Olsson & Shine 2003; Ryberg *et al.* 2004) ont été rapportées. Il n'est donc pas possible d'inférer un quelconque biais de dispersion chez les *Podarcis* avec les informations actuelles. Par ailleurs, l'hypothèse de dominance des mâles colonisateurs sur les mâles indigènes (Roca *et al.* 2005) ne pourra être discutée qu'une fois le taxon à l'origine de la lignée Valencia identifié.

V.A.2.b. Hypothèse sélective

La distribution passée du taxon Valencia aurait pu, selon l'hypothèse sélective et à l'encontre de l'hypothèse démographique, être beaucoup plus réduite que la distribution actuelle de sa lignée mitochondriale (Fig. 12). Suite à de l'hybridation locale avec *Podarcis liolepis*, la lignée Valencia aurait pénétré à l'intérieur de cette espèce puis aurait diffusé jusqu'à sa limite actuelle de distribution. Dans ce cas, la diffusion de la lignée Valencia sur une aussi grande distance et surtout le remplacement complet de la lignée mitochondriale d'origine de *P.*

liolepis est peu vraisemblable par dérive seulement et suggérerait l'intervention d'un balayage sélectif.

La sélection de l'ADNmt est une question majeure en phylogénie et en phylogéographie dans la mesure où les marqueurs mitochondriaux sont souvent choisis pour leur neutralité supposée (Ballard & Whitlock 2004). La sélection de l'ADNmt peut être indirecte, c'est-à-dire issue d'un déséquilibre avec d'autres loci soumis à sélection et transmis maternellement. Par exemple, chez les espèces à sexe femelle hétérogamétique, tel que chez les oiseaux ou les papillons, une distorsion de ségrégation méiotique impliquant l'entraînement du chromosome W, peut mener à un excès de descendants femelles et ainsi entraîner une augmentation de la fréquence des allèles mitochondriaux qui sont en déséquilibre de liaison avec ce chromosome W (Jaenike 2001; Hoekstra 2003). Une autre source de sélection indirecte de l'ADNmt implique des micro-organismes transmis maternellement chez les arthropodes (Hurst & Jiggins 2005). Ces parasites sont équipés de mécanismes (*e.g.* sex-ratio biaisé en faveur des femelles par féminisation, incompatibilité cytoplasmique) facilitant la prolifération des filles infectées au dépend des filles non-infectées et des fils. Les plus communs des endoparasites des arthropodes appartiennent au genre de bactérie *Wolbachia* qui infectent environ 20% des espèces d'insectes (Werren & Windsor 2000; Jiggins *et al.* 2001).

Les preuves de sélection directe de l'ADNmt sont nettement moins nombreuses et il n'existe aucune démonstration irréfutable. L'adaptation au climat semble toutefois influencer la variation mitochondriale chez l'homme (Mishmar *et al.* 2003; Ruiz-Pesini *et al.* 2004) et chez la musaraigne (Ehinger *et al.* 2002; Fontanillas *et al.* 2005). Comme signalé au Chapitre I, le génome mitochondrial des Vertébrés code pour 13 protéines impliquées dans la phosphorylation oxydative. Des différences entre haplotypes pourraient générer de la variabilité dans ce mécanisme biochimique, qui pourrait ainsi être soumise à une sélection intense dans les climats froids. En effet, le gradient de protons de la chaîne de transport d'électrons peut être dissipé sous forme de chaleur plutôt qu'être utilisé pour la production d'ATP dans la graisse brune de mammifères (Blank 1992). Ce mécanisme a également été proposé très récemment pour expliquer un flux de gènes mitochondriaux restreint comparé au flux de gènes nucléaires, le long d'un gradient altitudinal, chez une espèce d'oiseau de la cordillère des Andes (Cheviron & Brumfield 2009). Malheureusement, et contrairement à ce qui est affirmé par les auteurs, les oiseaux ne sont connus ni pour posséder de la graisse brune, ni pour détourner leur mécanisme de phosphorylation (Grigg *et al.* 2004). D'autres mécanismes faisant intervenir les mitochondries joueraient toutefois un rôle dans

l'homéothermie endothermique des oiseaux (Bicudo *et al.* 2002). Les reptiles ne possèdent pas non plus de graisse brune (Grigg *et al.* 2004) et le rôle joué par les mitochondries de ces animaux ectothermes dans la production de chaleur n'est pas clair. Chez d'autres animaux ectothermes comme les invertébrés de la zone intertidale, les contraintes thermiques sont responsables de la sélection de variants mitochondriaux sur la base d'une plus grande capacité à réparer les dommages membranaires engendrés par des extrêmes thermiques (Somero 2002). Il n'est pas improbable que des pressions de sélection similaires puissent s'opérer sur d'autres ectothermes comme les reptiles.

L'hypothèse de sélection de l'ADNmt causée par le climat a été initialement émise pour expliquer la diffusion de l'ADNmt du lièvre variable chez les espèces ibériques de lièvre, puisque le lièvre variable n'a vraisemblablement jamais été présent dans la Péninsule Ibérique (Melo-Ferreira *et al.* 2005). Cependant, le niveau élevé de polymorphisme de la lignée mitochondriale du lièvre variable à l'intérieur des populations ibériques des autres espèces, ainsi que l'origine polyphylétique des haplotypes introgressés, suggèrent des événements multiples d'introgession qui s'opposent à l'hypothèse sélective.

Dans le cas des lézards ibériques, la lignée Valencia correspondant dans la région de Valence n'est représentée que par deux haplotypes et présente une diversité réduite par rapport aux autres lignées mitochondriales de la région. Ceci suggère un événement unique d'introgession de la lignée Valencia vers *P. liolepis*, ce qui est en accord avec l'hypothèse sélective.

V.A.3. L'origine de la lignée Valencia

Au-delà de la zone étudiée dans l'Article 1, Pinho *et al.* (2006) ont identifié la lignée mitochondriale Valencia dans une population située au sud de la Péninsule Ibérique, dans la Sierra Nevada (Fig. 13). Malheureusement, les auteurs ne possèdent pas d'échantillon de référence pour les individus étudiés ce qui rend impossible l'identification précise de ces individus. Nous avons également trouvé cette lignée mitochondriale chez des individus localisés en altitude, au sommet de la Sierra de Cazorla (Fig. 13). Selon Philippe Geniez (com. pers.), les individus porteurs de cette lignée présentent une morphologie unique au sein du complexe d'espèces, avec un profil plus robuste chez les autres espèces (Fig. 13). Cette population de la Sierra de Cazorla pourrait représenter un taxon encore non-décrit, et pourrait être le taxon à l'origine de la lignée mitochondriale Valencia. Le même fragment de l'ADNmt

que celui utilisé dans l'Article 1 a été séquencé chez ces individus et a révélé un haplotype différent de celui présent chez *P. liolepis* et *P. hispanicus*. Toutefois, le taux d'évolution relativement lent de la région de contrôle ne permet pas d'établir un réseau d'haplotypes qui viserait à identifier si les haplotypes présents dans la Sierra de Cazorla sont à l'origine des haplotypes présents dans les autres populations. Le séquençage du fragment ND4 chez ces individus devrait permettre, dans un avenir proche, d'établir un tel réseau.

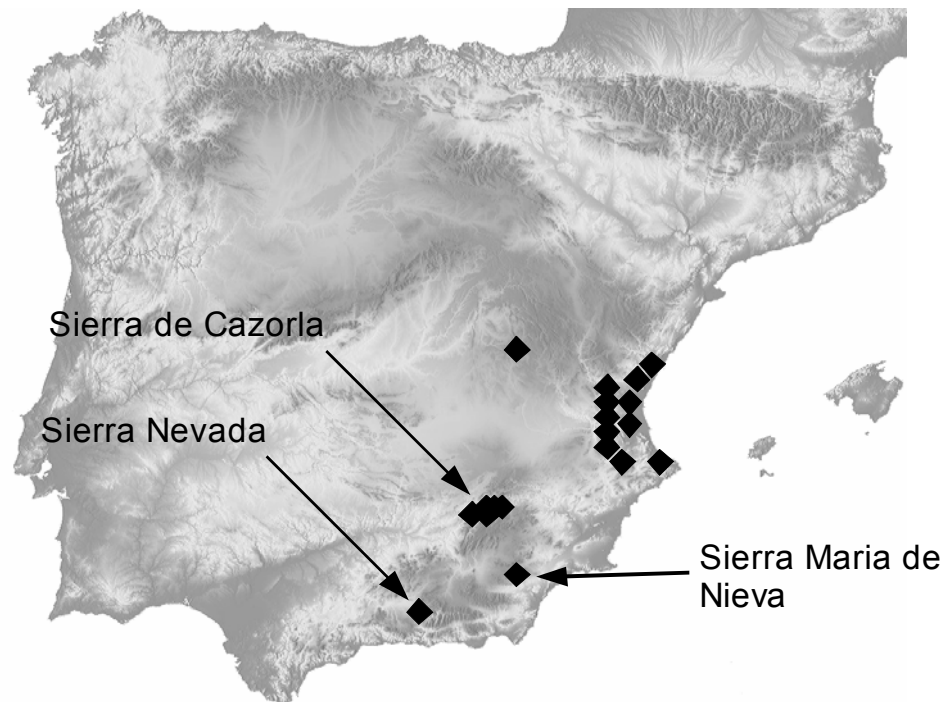


Figure 13. Distribution connue de la lignée mitochondriale Valencia. Quatre foyers de distribution sont actuellement connus : un en plaine dans la province de Valence et trois montagnards : une localité dans la Sierra Maria de Nieva, une localité dans la Sierra Nevada (Pinho *et al.* 2006) et une série de localités dans la Sierra de Cazorla (données non publiées).

Si les individus de la Sierra de Cazorla appartiennent effectivement au taxon à l'origine de la lignée Valencia, alors un scénario biogéographique faisant la synthèse des résultats présentés ci-dessus est relativement aisé à construire. En effet, le taxon putatif de Cazorla pourrait être adapté au froid et ainsi, aurait pu envahir les plaines et rentrer en contact avec les autres taxons de *Podarcis* lors des périodes froides du Quaternaire. C'est lors de ces contacts que des événements d'hybridation auraient pu avoir lieu et que l'introgression de la lignée mitochondriale Valencia à l'intérieur de *P. liolepis* et *P. hispanicus* aurait été initiée. Par la suite, le réchauffement climatique lié à la fin de la période glaciaire aurait contraint le

taxon de Cazorla à se réfugier dans une région montagneuse où il serait aujourd'hui endémique. Le scénario de refuge montagnard n'est pas inapproprié étant donnée la présence de nombreuses espèces de plantes endémiques à la chaîne de montagne de Cazorla (Gómez-Campo *et al.* 1984) ainsi que la présence d'une espèce endémique de lézard : *Algyroides marchi* Valverde 1958.

V.A.4. Conséquences taxonomiques de la discordance cyto-nucléaire

La biologie systématique, c'est-à-dire l'étude de la diversité biologique et de ses origines, est devenue un champ d'étude central dans la biologie moderne, trouvant des applications dans des champs aussi divers que l'épidémiologie médicale (Biek *et al.* 2006), l'écologie théorique (Anderson *et al.* 2004), la génétique de la conservation (Andreasen 2005) ou encore la biologie du développement (Minelli 2007). Une étape préliminaire, bien souvent indispensable à la réalisation de telles études, est l'identification et la délimitation des espèces. L'utilisation encore massive de marqueurs cytoplasmiques dans les études de systématique, ainsi que la fréquence des différents phénomènes pouvant mener à une discordance entre l'histoire des génomes cytoplasmiques et l'histoire des espèces, présentés au Chapitre IV, soulèvent la question de l'impact de ces discordances sur notre manière actuelle de caractériser la diversité du vivant.

Au sein du complexe d'espèces *P. hispanicus*, la population insulaire des îles Columbretes, situées au large de la côte est de la Péninsule Ibérique, a été la première à être élevée de manière formelle au rang d'espèce (Castilla *et al.* 1998a; Castilla *et al.* 1998b), et prit le nom de *Podarcis atratus* (Boscá, 1916). Par la suite, Busack *et al.* (2005) proposèrent de traiter la forme du nord-est de l'Espagne (= "*P. hispanica* morphotype 3" dans Pinho *et al.* 2006) comme une espèce distincte et lui donnèrent le nom de *P. atratus*. Les arguments avancés sont que la population des îles Columbretes est conspécifique des populations continentales de ce morphotype ; arguments qui confortent les propos de Geniez (2001).

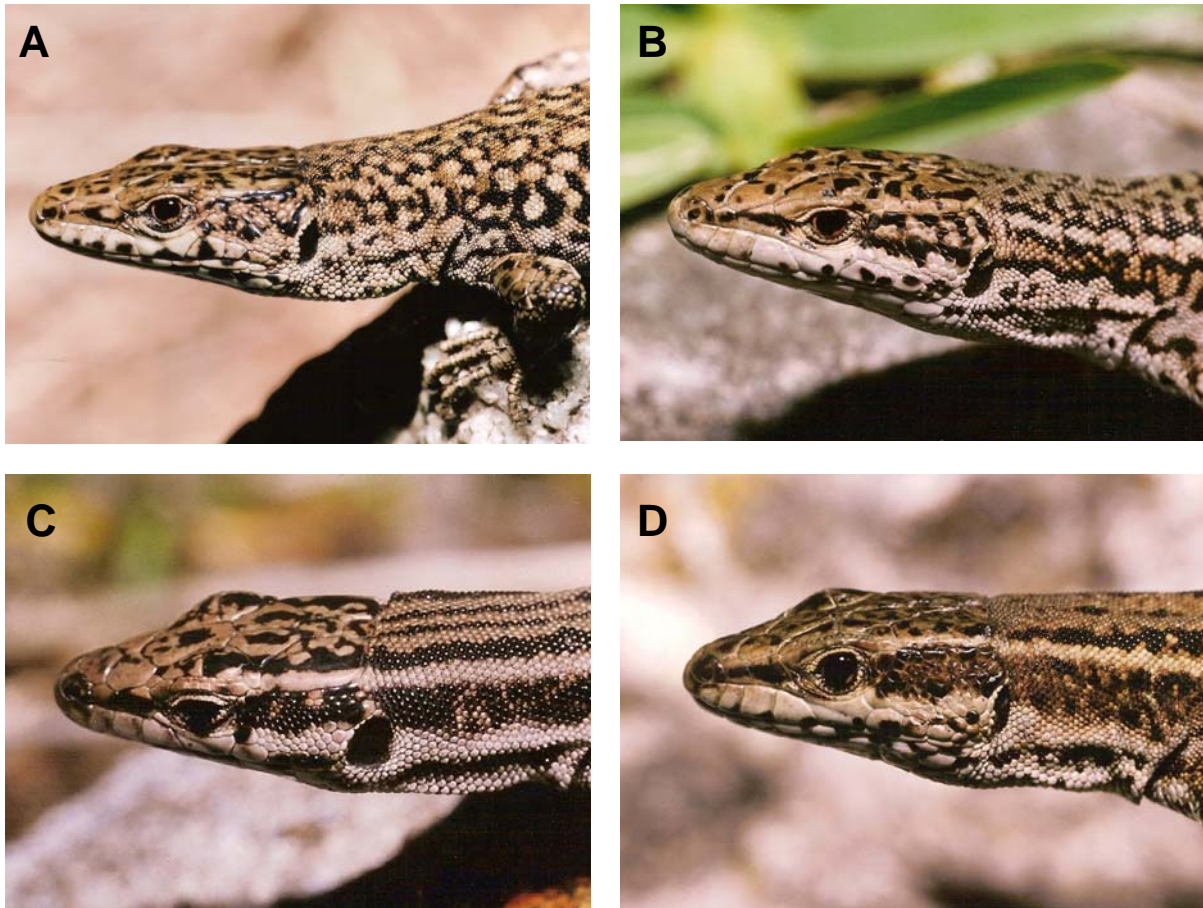


Figure 14. Vues dorso-latérales de têtes de *Podarcis* spp. du complexe d'espèces *hispanicus*. A. *P. liolepis*, Cap de Coste, France. B. *Podarcis* sp. « type 2 », Villanueva de los Escuderos, Espagne. C. *P. hispanicus*, Puebla de don Fabrique, Espagne. D. *Podarcis* sp., Puerto del Pinar, Sierra de Cazorla, Espagne. La photo D illustre le profil particulier de la tête (museau court et robuste) des individus de la Sierra de Cazorla, qui pourraient correspondre à un taxon non décrit.

La forme du nord-est de l'Espagne a été initialement décrite sous le nom *Lacerta muralis* var. *liolepis* Boulenger, 1905 comme rappelé par Geniez (2001) et Guillaume (1987). Le nom *liolepis* étant plus ancien que le nom *atratus*, il devrait avoir priorité sur ce dernier pour la dénomination des populations des îles Columbretes et de la forme du nord-est de l'Espagne. Toutefois, la forme du nord-est de l'Espagne a été génétiquement caractérisée à partir d'individus provenant de Barcelone (Harris & Sá Sousa 2002; Pinho *et al.* 2003), Burgos et Medinaceli (Castilla y León ; Pinho *et al.* 2006), Girona, Tarragona et les Pyrénées centrales (Harris & Sá Sousa 2002) et Andorre (Harris *et al.* 2002), et correspond à la lignée mitochondriale "*Podarcis hispanicus* type 3" (Pinho *et al.* 2006). Les syntypes de *Lacerta muralis* var. *liolepis* sont originaires principalement de la région de Valence et des alentours,

une région où les lézards *Podarcis* sont caractérisés par la lignée mitochondriale Valencia. Précédemment, nous avons démontré que les individus localisés autour de Valence et portant la lignée mitochondriale Valencia appartenaient à la même unité morphologique et nucléaire que la forme du nord-est de la Péninsule Ibérique. Il est donc vraisemblable qu'à l'époque à laquelle Boulenger décrit *Lacerta muralis* var. *liolepis*, les individus collectés et faisant office de syntypes, appartenaient également à la même unité évolutive que les individus du nord-est de l'Espagne.

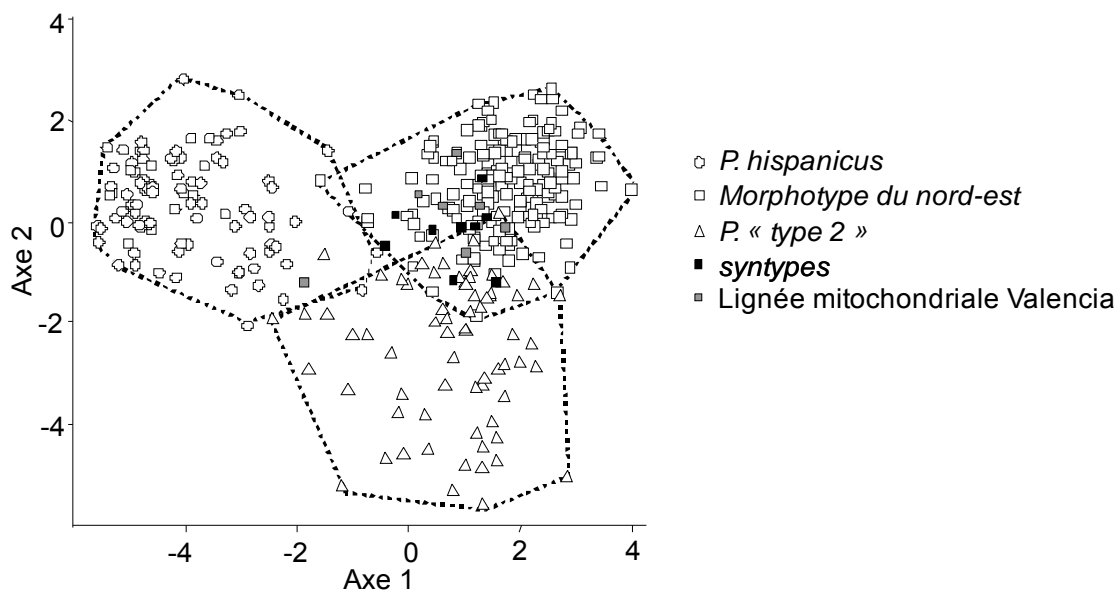


Figure 15. Première (Axe 1) et deuxième (Axe 2) composante d'une analyse discriminante effectuée à partir de 16 variables morphologiques sur des spécimens de *P. hispanicus* complexe spp. Les spécimens « Lignée mitochondriale Valencia » et « syntypes » ne sont pas actifs dans l'analyse discriminante, c'est-à-dire qu'ils ne sont pas pris en compte dans la construction des fonctions discriminantes. Les lignes pointillées relient les individus extrêmes pour chacun des trois morphotypes donnés en référence : *P. « type 2 »*, *P. hispanicus* et *Morphotype du nord-est*.

Afin de valider ce propos, nous avons étudié la morphologie des syntypes de *Lacerta muralis* var. *liolepis* en comparaison à des individus de référence appartenant au morphotype du nord-est de l'Espagne (et portant la lignée mitochondriale Liolepis), des individus collectés récemment dans et autour de la ville de Valence (et possédant la lignée mitochondriale Valencia), des *P. « type 2 »* et des *P. hispanicus* tels que redéfinis par Geniez (2001) (Fig. 15 ; des analyses morphologiques détaillées sont données dans l'Article 3).

Graphiquement, la morphologie des syntypes de *Lacerta muralis* var. *liolepis* est incluse dans la variation normale du morphotype du nord-est, tout comme les individus portant la lignée Valencia collectés récemment autour et dans la ville de Valence. Ce résultat soutient le fait que les syntypes appartiennent au taxon résidant dans le nord-est de l'Espagne. En conclusion, les populations localisées dans les îles Columbretes, dans le nord-est de l'Espagne et autour de Valence, appartiennent à une même espèce, qui devrait à l'avenir être dénommée *P. liolepis*. C'est pour cette raison que nous avons anticipé cette révision nomenclaturale et que nous avons utilisé ce nom d'espèce tout au long de ce document.

Au-delà des seules modifications nomenclaturales entraînées par la reconnaissance d'une discordance cyto-nucléaire au sein du complexe d'espèces de *Podarcis hispanicus*, il convient de souligner que c'est la reconnaissance même des unités évolutives qui aurait été altérée si elle avait été uniquement fondée sur des analyses mitochondriales. Une étude récente, basée sur des séquences d'introns nucléaires, a démontré l'absence de flux de gènes entre individus porteurs de la lignée mitochondriale Valencia et individus porteurs de la lignée mitochondriale Liolepis, soutenant ainsi l'idée de deux espèces distinctes pour ces deux lignées (Pinho *et al.* 2008). Cependant, la lignée Valencia est représentée dans cette étude par des individus appartenant à des populations distinctes : deux spécimens provenant de la distribution de *P. liolepis*, quatre spécimens provenant de la Sierra Nevada (et de morphologie inconnue), un individu de la Sierra Maria de Nieva (que nous avons montré être un *P. hispanicus*) et deux spécimens provenant de la Sierra de Cazorla appartenant probablement à une espèce non décrite. Les flux de gènes nucléaires ont donc été estimés en incluant dans une même population des individus appartenant à des espèces différentes, ce qui prévient toute interprétation ces résultats. Cette erreur provient du fait que la variation morphologique n'a jamais été utilisée pour construire des hypothèses de délimitation d'unités évolutives alternatives, et ceci bien que ces hypothèses soient disponibles depuis plus de vingt ans (Guillaume 1987). Le complexe d'espèces de *Podarcis hispanicus* est en cours de révision systématique, et cette révision aurait été erronée si elle ne s'était basée que sur les résultats des analyses mitochondriales et si elle avait nié les travaux historiques qui décrivent la variation morphologique au sein de ce complexe.

V.B. Exploitations des discordances cyto-nucléaires pour comprendre l'origine des incongruences entre la phylogénie des plantes et la phylogénie des pollinisateurs chez les figuiers *Galoglychia*.

V.B.1. Incongruences phylogénétiques, changement d'hôte et duplication/extinction

Les relations interspécifiques spécialisées sont très diversifiées, allant du mutualisme à l'association hôte-parasite en passant par la cohabitation neutre. Le degré de congruence entre les phylogénies des espèces en interaction, ou en d'autres termes le degré de spéciation parallèle, a été historiquement perçu comme un bon estimateur de la force d'une interaction (Hoberg *et al.* 1997). En effet, si chaque événement de spéciation chez un des partenaires est suivi par un événement de spéciation chez l'autre partenaire, alors les phylogénies des deux partenaires présenteront un fort degré de similitude et l'interaction sera qualifiée de forte (Hafner & Nadler 1988; Page & Hafner 1996; Chen *et al.* 1999; Thao *et al.* 2000). Cependant même dans le cas d'interactions où les espèces interagissent étroitement, comme dans le cas des systèmes hôte-parasite, des processus écologiques se produisant après les événements de spéciation peuvent entraver la congruence entre phylogénies. Les changements d'hôtes (Fig. 16) représentent certainement le processus le plus commun à l'origine d'incongruence phylogénétique dans ce type de système (Menken 1996; Schoonhoven *et al.* 1998; Charleston & Robertson 2002; Jackson 2004; Ricklefs *et al.* 2004; Huyse & Volckaert 2005). Les changements d'hôtes peuvent se produire dans diverses situations, comme lorsqu'un hôte, mais pas le parasite colonise, un nouvel habitat, ce qui laisse une niche vacante pour un nouveau parasite (Paterson *et al.* 1999), ou lorsque qu'un hôte est remplacé par un autre hôte par compétition. Un autre phénomène pouvant réduire la congruence phylogénétique est la duplication/extinction. Ce phénomène se produit lorsqu'une espèce de parasite se duplique en l'absence de spéciation chez l'hôte puis que des événements de cospéciation sont suivis par des événements d'extinctions asymétriques (Fig. 16) (Reed *et al.* 2007). Bien que différents par nature, ces deux phénomènes peuvent conduire à des patrons d'incongruence phylogénétique très similaires et leur distinction est souvent largement hypothétique (Jackson 1999). Une méthode permettant de distinguer les deux hypothèses est basée sur la comparaison de l'âge finement daté de nœuds homologues dans les phylogénies des deux partenaires. L'obtention de telles phylogénies est cependant très délicate à cause de l'incertitude parfois élevée associée aux points de calibration, du taux d'évolution parfois

hétérogène entre les branches, ou encore des explosions radiatives (Percy *et al.* 2004; Rønsted *et al.* 2005). Une méthode fiable permettant d'identifier la source d'une incongruence entre la phylogénie d'un hôte et celle de son partenaire reste donc à définir.

Parmi les interactions interspécifiques étroites, les interactions spécifiques plante-pollinisateur tiennent une place particulière dans la mesure où le pollinisateur, en transportant des gènes nucléaires du génome de la plante hôte, façonne directement la structure génétique des populations de l'espèce hôte (Wesselingh & Arnold 2000; Martin *et al.* 2008). En donnant lieu à de l'hybridation, un changement d'hôte suivi d'une utilisation transitoire de deux espèces d'hôte par une même espèce de pollinisateur peut donc laisser des traces génétiques dans le génome des hôtes. A la suite de tels événements d'hybridations, des croisements en retour entre les descendants et l'une des deux espèces parentales peuvent également permettre l'introduction du génome cytoplasmique d'une espèce parentale dans l'autre espèce parentale. Comme signalé au Chapitre II, ce phénomène est facilité par les caractéristiques architecturales et évolutives de l'ADNcp (Liston & Kadereit 1995; Van Droogenbroeck *et al.* 2006; Prentice *et al.* 2008). Ainsi, dans le cas d'un changement d'hôte accompagné d'une utilisation transitoire mais durable de deux hôtes par un pollinisateur, l'ADNcp pourrait être utilisé pour retracer des événements passés d'hybridation et donc de changement d'hôte.

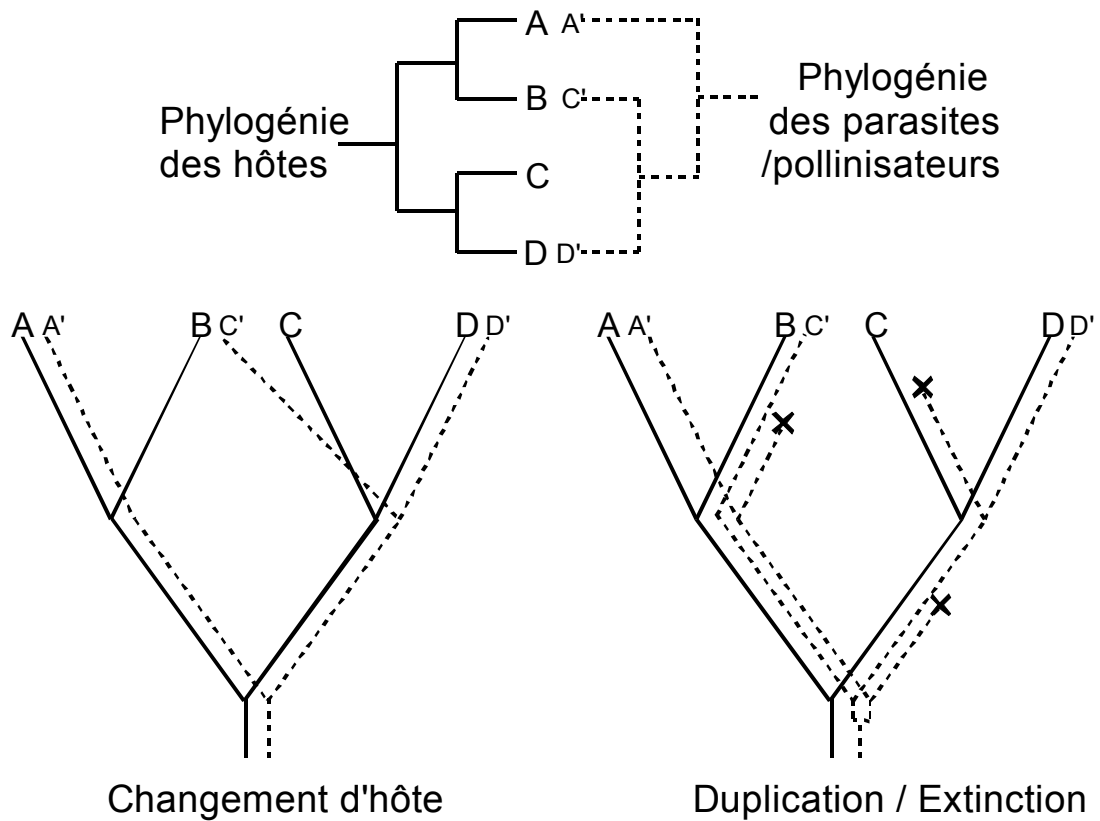


Figure 16. Hypothèses de changement d'hôte et de duplication/extinction expliquant une incongruence entre la phylogénie des hôtes et la phylogénie des parasites ou des pollinisateurs. Dans l'hypothèse de duplication/extinction, les croix représentent des extinctions de parasites qui suivent un évènement basal de spéciation sur l'espèce ancestrale aux hôtes ABCD.

A l'inverse d'un changement d'hôtes, la duplication/extinction n'est pas associée à une augmentation de la probabilité d'occurrence d'un évènement d'hybridation. Ainsi, l'apparition d'évènements d'hybridations dans une lignée de plantes en fréquence « anormale » pourrait constituer un argument en faveur de l'hypothèse de changement d'hôte et non en faveur de l'hypothèse de duplication/extinction. Comme nous l'avons montré au chapitre précédent, les introgressions peuvent être révélées au travers de discordances cyto-nucléaires. Ainsi, les discordances peuvent offrir un outil précieux pour définir quelle est l'hypothèse à l'origine d'une incongruence entre la phylogénie de l'hôte et la phylogénie du pollinisateur.

V.B.2. *Changements d'hôtes chez les figuiers Galoglychia mis en évidence par les discordances cyto-nucléaires*

Historiquement, les incongruences entre phylogénie des hôtes et phylogénie des pollinisateurs chez les figuiers ont été attribuées à des changements d'hôte (Cook & Rasplus 2003; Machado *et al.* 2005). Les nombreuses exceptions aux règles « une espèce d'hôte pour une espèce de pollinisateur » et « un genre de guêpe pollinise une sous-section ou une section » qui semblent caractériser l'association hôte-parasite chez les figuiers de la section *Galoglychia* (voir Chapitre III) ont soulevé, récemment encore, des débats quant à leur origine. Il a ainsi été proposé que des événements de duplication anciens suivis par des extinctions asymétriques aient pu être la cause de l'association relativement confuse qu'ont les pollinisateurs *Alfonsiella* et *Elisabethiella* avec leurs hôtes (Erasmus *et al.* 2007; Jousselin *et al.* 2008). Notamment, Jousselin *et al.* (2008) soutiennent cette hypothèse en démontrant une régression linéaire significative du taux de divergence des séquences dans des paires *Elisabethiella*/guêpes parasites qui suivent un patron de cospéciation. Les auteurs notent toutefois que ce résultat doit être interprété avec une extrême précaution puisqu'il est basé sur un seul fragment ITS pour la phylogénie des insectes et sur un faible nombre de nœuds. Ainsi, la question de savoir si *Elisabethiella* et *Alfonsiella* représentent deux radiations séparées sur des mêmes hôtes n'est toujours pas résolue.

Les deux discordances cyto-nucléaires significatives révélées par la comparaison de la phylogénie chloroplastique et de la phylogénie nucléaire, dont la cause a été attribuée à de l'hybridation, impliquent une espèce pollinisée par des *Alfonsiella* (*F. craterostoma*) et une espèce pollinisée par des *Elisabethiella* (*F. craterostoma*). En revanche, aucun cas d'hybridation n'a été observé dans les sous-sections ne présentant pas un patron d'incongruence plante-pollinisateur. Cette observation est en accord avec l'hypothèse que ces incongruences puissent avoir été causées par des changements d'hôtes. Au-delà de l'argument basé sur la discordance cyto-nucléaire, deux autres observations soutiennent l'hypothèse de changement d'hôte. Premièrement, le raisonnement tenu par les défenseurs de l'hypothèse de duplication/extinction devrait également s'appliquer à la sous-section *Crassicostae* qui est pollinisée par trois genres de guêpes (*Paragaon*, *Elisabethiella*, *Nigeriella*). Comme *Elisabethiella* et *Nigeriella* sont également pollinisées par la sous-section *Chlamydodora*, ces genres auraient dû se différencier très précocement dans l'histoire de la section *Galoglychia* et c'est seulement bien plus tard que l'un des genres se serait éteint chez ces espèces et aurait survécu chez d'autres espèces. Ce scénario apparaît assez improbable.

Deuxièmement, Erasmus *et al.* apportent eux-même des preuves directes d'échanges en cours de pollinisateurs entre espèces de *Ficus*. En effet, des échantillons d'*Alfonsiella binghami* en provenance de Nelspruit (Afrique du Sud) forment un unique ensemble phylogénétique, qu'ils aient été collectés sur *F. petersii* ou sur *F. sthulmanii*, différent des échantillons d'*A. binghami* collectés sur des *F. sthulmannii* en provenance de Lekgalameetse Afrique du Sud) et d'Amani (Tanzanie).

La comparaison de la phylogénie nucléaire des plantes et de la phylogénie des insectes pollinisateurs, aussi bien celle présentée dans l'Article 2 que celle précédemment publiée par Rønsted *et al.* (2007a), révèle un nouveau cas d'incongruence (Fig. 17). Selon une phylogénie récente des pollinisateurs (Erasmus *et al.* 2007), les genres pollinisant les sous-sections *Cyathistipulae* et *Crassicostae* forment un ensemble monophylétique avec les genres pollinisant les sous-sections *Chlamydodorae* et *Platyphyllae*. Cette association est donc incongruente avec la monophylie des sous-sections *Cyathistipulae*, *Crassicostae*, *Caulocarpae* et *Galoglychia* selon la phylogénie nucléaire.

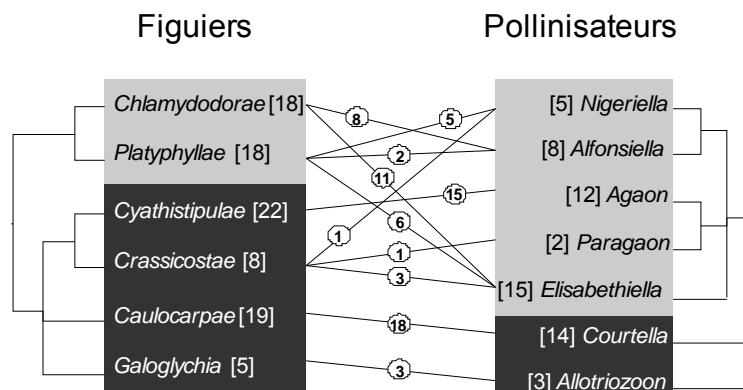


Figure 17. Associations entre les sous-sections de la section *Galoglychia* (à gauche) et les genres de pollinisateurs (à droite). Le nombre d'espèces pour chaque sous-section et chaque genre est donné entre crochets. Les connecteurs indiquent quel genre de pollinisateur est associé avec quelle sous-section, avec le nombre d'espèces de pollinisateurs impliqués dans chaque association genre/sous-section donné à l'intérieur des cercles. Le fond coloré surligne les principaux clades afin de montrer l'incongruence basale entre la phylogénie des figuiers de la section *Galoglychia* (d'après Rønsted *et al.* 2007a) et la phylogénie des insectes (d'après Erasmus *et al.* 2007) liée à la position des sous-sections *Cyathistipulae* et *Crassicostae* et de leurs pollinisateurs.

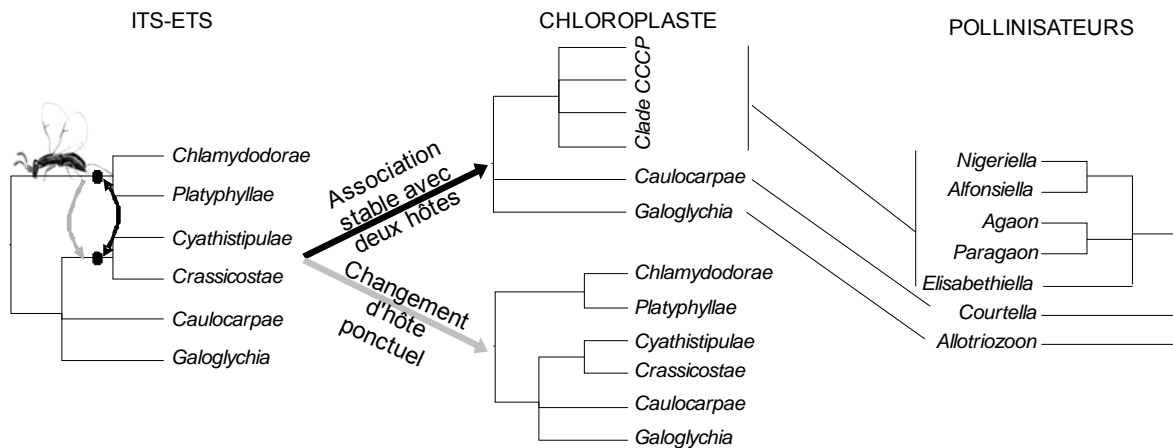


Figure 18. Hypothèses de changement d'hôte chez les figuiers de la section *Galoglychia*. Une guêpe pollinisatrice de l'ancêtre commun aux *Chlamydodora* et *Platyphyllae* aurait pollinisé un ancêtre commun aux *Cyathistipulae* et *Crassicostae*. Toutefois, selon l'hypothèse qu'un changement d'hôte unique et ponctuel se serait produit (flèches grises), il n'y aurait pas de transfert du génome chloroplastique et la lignée de *Ficus* résultant de l'hybridation porterait nécessairement un génome chloroplastique apparenté à celui des *Caulocarpae* et des *Galoglychia* ; ce qui est contraire à la phylogénie chloroplastique. A l'inverse, si les pollinisateurs utilisent les deux hôtes ancestraux de manière pérenne (flèches noires), des croisements en retours répétés peuvent permettre l'introduction du génome chloroplastique à l'origine des *Chlamydodora* et *Platyphyllae* à l'intérieur de l'ancêtre des *Cyathistipulae* et *Crassicostae*. La phylogénie chloroplastique soutient cette deuxième hypothèse, ainsi que l'observation dans la nature d'utilisation stable de deux hôtes par une même espèce de pollinisateur.

Le clade cytoplasmique CCCP est exclusivement pollinisé par le clade de pollinisateur comprenant les genres *Agaon*, *Paragaon*, *Alfonsiella*, *Nigeriella* et *Elisabethiella*. On peut alors supposer qu'un premier changement d'hôte par une guêpe pollinisant un ancêtre des *Chlamydodora* actuels, vers un ancêtre commun aux *Cyathistipulae* et *Crassicostae* aurait pu donner lieu à la lignée de pollinisateur *Paragaon-Agaon* (Fig. 18). Ensuite, un deuxième transfert de pollinisateur de *Crassicostae* vers *Chlamydodora* doit être invoqué afin de transporter le génome chloroplastique ancestral à *Cyathistipulae* et *Crassicostae* chez l'ancêtre des *Chlamydodora*. Une telle succession de deux événements de changements d'hôtes peut paraître improbable. Toutefois, un mécanisme simple pouvant être à l'origine de ce transfert de cytoplasme est l'utilisation stabilisée de deux espèces d'hôtes (l'ancêtre des *Cyathistipulae* et *Crassicostae* et l'ancêtre des *Chlamydodora*) par les pollinisateurs à

l'origine de la lignée *Agaon-Paragaon*. Cette double association stabilisée permettrait en effet des croisements en retour entre les hybrides et les deux espèces parentales. Des observations de terrain viennent soutenir la plausibilité de ce mécanisme, avec plusieurs cas rapportés d'utilisation stable de deux espèces de figuiers hôtes par une même espèce de guêpe (Erasmus *et al.* 2007). Par exemple, *Elisabethiella stuckenbergii* est connu pour polliniser *F. burkei*, *F. natalensis natalensis*, *F. lingua depauperata* et *F. petersii* (Underhill 2008), un patron d'association qui pourrait être responsable du transfert de cytoplasme dans n'importe quelle direction, indépendamment de quelle espèce était l'hôte original. Le scénario de changement d'hôte paraît donc parfaitement approprié pour expliquer l'incongruence entre la phylogénie des insectes pollinisateurs et la phylogénie de l'hôte.

En conclusion, nous avons observé trois cas de discordances cyto-nucléaires imputables à de l'hybridation introgressive dans la phylogénie des figuiers *Galoglychia*, qui affectent tous les trois des portions de la phylogénie qui sont incongruentes avec la phylogénie des insectes. En revanche, aucune introgression n'a été détectée dans des clades qui ne présentent pas d'incongruence entre hôtes et pollinisateur. Ce patron de distribution des introgressions soutient fortement l'hypothèse que le changement d'hôte puisse être à l'origine des incongruences entre hôtes et pollinisateurs, puisque des événements d'introgression ne sont pas attendus sous l'hypothèse de duplication/extinction. Ainsi, les discordances cyto-nucléaires offrent des indices précieux sur l'histoire des associations symbiotiques entre les *Ficus* et leurs guêpes pollinisatrices.

Le raisonnement présenté ici montre des limites liées au faible nombre de discordances cyto-nucléaires mis en évidence. L'utilisation d'outils développés pour étudier l'évolution de caractères dans une phylogénie pourrait permettre de démontrer statistiquement l'association non aléatoire entre la distribution des événements d'introgression et la phylogénie des insectes. Toutefois, de tels outils sont généralement basés sur la comparaison d'une distribution de caractères observés avec une distribution simulée selon un modèle d'évolution stochastique. Il est bien évident que la capacité d'un test à détecter un écart à un modèle nul est dépendante du nombre d'occurrences du caractère au sein de la phylogénie. Pour cette raison, nos trois événements d'introgressions ne suffiront pas à obtenir un résultat assez robuste pour qu'il puisse être interprété de manière fiable. Le faible nombre d'introgressions validées statistiquement dans notre jeu de données est à mettre en relation directe avec le manque de résolution de la phylogénie chloroplastique. Un affinement de la phase de développement des marqueurs chloroplastiques ainsi que l'ajout de nouveaux

marqueurs devraient dans un futur proche permettre une meilleure résolution des branchements à l'intérieur du clade CCCP. L'augmentation du nombre d'évènements d'introgessions statistiquement soutenus devrait permettre à terme d'utiliser des méthodes investiguant statistiquement la distribution des discordances cyto-nucléaires dans la phylogénie des plantes et des insectes.

CONCLUSION

Depuis leur association avec les cellules eucaryotes, les organites intracellulaires n'ont jamais tout à fait perdu leur liberté. Bien que le transfert d'une partie de leurs gènes vers le noyau les contraigne à baigner dans un cytoplasme, certains mécanismes évolutifs rendent possible une évolution des organites indépendante de celle des lignées cellulaires. L'hybridation introgressive est un de ces mécanismes. A l'aide d'un modèle animal et d'un modèle végétal, nous avons montré qu'étudier l'origine des lignées de génomes cytoplasmiques permettait d'apporter des lumières sur des processus démographiques (mise en évidence de déplacements de populations et de mise en contact avec l'exemple des lézards ibériques) et écologiques (mise en évidence de changements d'hôtes chez les pollinisateurs des figuiers). Tandis que les marqueurs cytoplasmiques ont été et continuent d'être des marqueurs largement utilisés pour décrire des patrons phylogéographiques et phylogénétiques (identification et distribution géographique des lignées évolutives, description de leurs liens de parenté), l'utilisation de ces marqueurs dans le but de comprendre des processus est dans une certaine mesure originale.

Dans un article de référence en phylogéographie de mammifères européens, Bilton et ses collaborateurs ont écrit : « La variation de l'ADNmt a été analysée sur l'ensemble de cette distribution [le Paléarctique] selon l'hypothèse largement acceptée qu'une telle variation reflète pleinement une histoire des populations non biaisée par la sélection ou les flux de gènes » (Bilton *et al.* 1998). Une telle assertion était relativement commune à la fin du siècle passé et s'inscrivait dans un contexte de découverte et d'essor des marqueurs moléculaires utilisés dans le but de retracer l'histoire du vivant. C'est toutefois dès l'aube de cet essor que des débats sur la pertinence de l'emploi des marqueurs cytoplasmiques en phylogénie et phylogéographie sont apparus (Avice 1994). En dépit des très nombreuses études qui ont, au cours de cette dernière décennie, illustré le désormais bien connu « gene tree and species tree are not the same », le débat sur l'utilisation des marqueurs mitochondriaux face aux

marqueurs nucléaires est toujours d'actualité (Zink & Barrowclough 2008; Barrowclough & Zink 2009; Edwards & Bensch 2009). Selon Zink & Barrowclough (2008), l'intérêt de l'ADNmt comme marqueur est lié à la fenêtre temporelle à explorer. Les auteurs distinguent ainsi l'utilité de l'ADNmt selon que les études sont d'ordre phylogéographique ou phylogénétique. Dans leur article (cité déjà plus de 35 fois un an et demi après sa parution, essentiellement par des articles voulant justifier l'emploi de l'ADNmt), Zink et Barrowclough rappellent que les allèles mitochondriaux sont réciproquement monophylétiques quatre fois plus rapidement que les allèles nucléaires, ce qui rend les premiers plus appropriés que les derniers pour reconnaître et définir des unités évolutives dans un contexte phylogéographique. De manière étonnante, les auteurs ne discutent que très partiellement les phénomènes d'introgession, comme si une mauvaise reconnaissance des unités évolutives n'était majoritairement le fait que d'un tri incomplet de polymorphisme ancestral. Par ailleurs, les marqueurs nucléaires seraient, selon ces auteurs, plus adéquats pour étudier les processus démographiques et évolutifs et les relations phylogénétiques au niveau supra-spécifique.

La validité de la conclusion de Zink et Barrowclough sur l'utilité des marqueurs cytoplasmiques en phylogéographie pourrait alors dépendre de la fréquence des événements d'introgession. La fréquence de ces événements est également étroitement associée à l'échelle temporelle : ces événements peuvent se produire avec une fréquence d'autant plus élevée que la divergence entre les populations ou espèces considérées est récente et que leur isolement reproducteur est incomplet (Coyne & Orr 1997; Price & Bouvier 2002). Toutefois, les conséquences de ces introgessions ne sont, elles, pas dépendantes de l'échelle temporelle. En effet, si les allèles d'un gène introgressé finiront tous par être monophylétiques à l'intérieur d'une espèce d'accueil, il n'en demeure pas moins que ces allèles seront toujours phylogénétiquement plus proches des allèles descendants de leur ancêtre commun, présent dans l'espèce d'origine, que des allèles de n'importe quelle autre espèce. C'est cette propriété que nous avons observée chez les figuiers *Galoglychia*, lorsque nous expliquons la position discordante de l'ancêtre commun aux *Cyathistipulae* et *Crassicostae*, entre les phylogénies chloroplastique et nucléaire. Ainsi, si les hybridations introgressives sont des événements relativement fréquents, la nature non-recombinante des génomes cytoplasmiques confèrerait alors un réel désavantage aux marqueurs cytoplasmiques par rapport aux marqueurs nucléaires pour les reconstructions de liens de parenté (en phylogénie et phylogéographie) et pour la délimitation d'unités évolutives (en phylogéographie).

Savoir si les événements d'introgession sont fréquents au point de contraindre l'emploi de marqueurs cytoplasmiques utilisés dans le but de décrire des patrons est une

question délicate. L'article de Mallet (2005) donne des fréquences d'hybridation rapportées dans différents groupes taxonomiques, végétaux et animaux. Toutefois, la fréquence des hybridations n'est pas exactement équivalente à la fréquence des introgressions, la différence étant liée à certains mécanismes d'isolement post-zygotique. Le message de Mallet est toutefois de souligner que les événements d'hybridation, et donc sans doute d'introgression, sont probablement assez communs, même dans des groupes taxonomiques où l'on croyait ces phénomènes marginaux (par exemple les oiseaux et les mammifères). Ce que rappelle par ailleurs Mallet, c'est la variabilité des fréquences d'hybridation et donc probablement d'introgression entre les différents groupes taxonomiques. Des facteurs biologiques propres à certaines lignées peuvent augmenter considérablement les fréquences d'introgression. C'est le cas par exemple des syndromes de pollinisation, où la stochasticité associée au mode de pollinisation anémophile est un facteur favorisant l'hybridation. Toutefois, les effets de tels facteurs sur les taux d'hybridation pourraient être beaucoup plus difficiles à prédire que suspecté jusqu'ici. Ainsi, nous avons démontré que même dans une association plante/pollinisateur de type spécifique, des événements d'introgression peuvent se produire.

Des facteurs géographiques pourraient par ailleurs jouer un rôle prépondérant dans la fréquence des événements d'introgression. Il n'existe pas, à notre connaissance, de méta-analyse traitant de cette question. Toutefois, il est bien connu que certains mécanismes démographiques (*e.g.* des déplacements de populations liés à des modifications de climat) sont des facteurs favorisant la rencontre entre espèces faiblement différenciées, ce qui favorise également leur hybridation. Dans ce contexte, les anciennes zones de refuge glaciaire ont vraisemblablement été le siège privilégié de tels événements d'introgression (Comes & Kadereit 1998), comme en témoigne nos travaux sur les lézards ibériques. De même, des travaux préliminaires sur les figuiers *Sycomorus* d'Afrique du sud et de Madagascar semblent souligner ce point. Nous avons en effet pu mettre en évidence une discordance flagrante entre la phylogénie chloroplastique et la phylogénie nucléaire des espèces de *Sycomorus* de cette région, se traduisant par l'occurrence de deux clades – un malgache et un sud Africain – dans la phylogénie chloroplastique tandis que les espèces malgaches sont morphologiquement et nucléairement polyphylétiques. L'exceptionnelle diversité de Madagascar est bien connue et trouverait en partie sa source dans la succession de cycles d'isolement et de remise en contact de taxons dont les conséquences furent prononcées sur cette île à cause d'une topographie particulière (Vences *et al.* 2009).

Finalement, nous pensons que les questions relatives à la pertinence de l'emploi de marqueurs cytoplasmiques dans le but de décrire des patrons phylogéographiques ou

phylogénétiques devraient davantage considérer des facteurs taxonomiques et biogéographiques. Ainsi, il nous apparaîtrait aujourd'hui peu approprié d'utiliser exclusivement des marqueurs cytoplasmiques chez des plantes endémiques d'anciens refuges glaciaires afin de décrire un patron phylogéographique.

Les marqueurs cytoplasmiques présentent indéniablement des avantages, pour partie d'ordre technique comme leur facilité d'amplification et la résolution automatique de l'état de caractère liée à leur nature haploïde. La facilité de leur utilisation semble placer ces outils à l'extrémité d'un continuum dont l'autre extrémité est représentée par l'alpha-taxonomie. En effet, l'incessante augmentation de la biodiversité décrite aboutit à une constante complexification des clefs d'identification des taxons (même si, en contre partie, les supports de diffusion d'information actuels rendent ces clefs plus facilement disponibles). D'un autre côté, l'émergence de l'écologie évolutive appelle à une relance de la taxonomie, dans la mesure où la taxonomie constitue un pilier de l'écologie évolutive. Une réponse à ce constat a été proposée par certains sous la forme d'un aboutement des extrémités du continuum. La taxonomie moléculaire est apparue, avec comme fer de lance le code-barres moléculaire. Par son approche simple-marqueur, le code-barres moléculaire marque une rupture avec des approches de plus en plus intégratives qui caractérisent la biologie moderne. Est-ce que l'outil code-barres moléculaire peut réellement offrir une alternative à l'identification par l'alpha-taxonomie ? La réponse est clairement non. Nous avons pu voir que l'utilisation seule des marqueurs cytoplasmiques ne pourrait en aucun cas révéler l'existence de nouvelles espèces (puisque une lignée mitochondriale peut-être présente sans que son espèce hôte ne le soit) ni même permettre d'attribuer avec certitude un nom d'espèce à un taxon. Ceci ne signifie pas pour autant que cet outil ne possède pas d'applications intéressantes. Un bon nombre de travaux en écologie peuvent se contenter d'une identification au niveau du genre et le code-barres moléculaire possède généralement un très bon taux d'identification à ce niveau taxonomique (Hollingsworth *et al.* 2009b). En outre, les techniques modernes de séquençage haut-débit permettront d'obtenir des informations sur la taxonomie, facilement et en grande quantité. Par ailleurs, l'emploi de cet outil peut permettre de cibler rapidement un taxoniste qui apportera par la suite son aide pour une identification plus fine. L'utilisation du code-barres moléculaire dans le but de découvrir de nouvelles espèces est également envisageable, s'il n'est pas utilisé seul. Par exemple, c'est parce que les herpétologues ne reconnaissaient l'existence que de trois morphotypes de Lézards ibériques dans le sud est de l'Espagne que nous avons recherché l'existence d'un taxon à l'origine d'une lignée mitochondriale surnuméraire. Comme cet exemple le montre, l'utilisation conjointe de connaissances

taxonomiques traditionnelles et de l'outil « marqueur cytoplasmique » a permis de décrire correctement le patron phylogéographique. Finalement, le code-barres moléculaire peut être un outil intéressant pourvu qu'on ne le considère que comme un outil et non pas comme une discipline. En effet, le code-barres moléculaire n'offre pas la possibilité, comme le fait l'alpha-taxonomie, d'émettre des hypothèses et de les tester. Dans les précédentes études portant sur le lézard ibérique, les hypothèses taxonomiques avaient été construites à partir des résultats d'analyses faites uniquement avec des marqueurs mitochondriaux. Les marqueurs nucléaires avaient été utilisés par la suite pour comprendre les processus démographiques et évolutifs expliquant les patrons décrits par les marqueurs cytoplasmiques.

Tout au long de cette thèse, nous nous sommes efforcés à démontrer que des patrons phylogénétiques et phylogéographiques fiables ne pouvaient émerger de l'utilisation seule de marqueurs cytoplasmiques. A l'inverse, nous avons pu démontrer que ces mêmes marqueurs étaient de formidables outils pour comprendre les processus démographiques, évolutifs et écologiques.



RÉFÉRENCES BIBLIOGRAPHIQUES

- Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, **29**, 380-395.
- Allen JF (1996) Separate sexes and the mitochondrial theory of ageing. *Journal of Theoretical Biology*, **180**, 135-140.
- Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417-434.
- Anderson TM, Lachance MA, Starmer WT (2004) The relationship of phylogeny to community structure: The cactus yeast community. *American Naturalist*, **164**, 709-721.
- Andreasen K (2005) Implications of molecular systematic analyses on the conservation of rare and threatened taxa: Contrasting examples from Malvaceae. *Conservation Genetics*, **6**, 399-412.
- Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from 7 individuals. *Genetics*, **103**, 287-312.
- Arnold ML (1997) *Natural hybridization and evolution* Oxford University Press, New York.
- Ashley MV, Laipis PJ, Hauswirth WW (1989) Rapid segregation of heteroplasmic bovine mitochondria. *Nucleic Acids Research*, **17**, 7325-7331.
- Avise JC (1994) *Molecular markers, natural history, and evolution* Chapman & Hall, New York.
- Avise JC (2000) *Phylogeography: the history and formation of species* Harvard Univ. Press, Cambridge, MA.
- Avise JC (2004) *Molecular markers, natural history and evolution* Sinauer Associates, Sunderland.
- Avise JC, Lansman RA, Shade RO (1979) Use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations .1. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**, 279-295.
- Bacquet P (2008) *Étude des causes évolutives d'une introgression mitochondriale chez le Lézard occitan Podarcis liolepis (Reptilia, Lacertidae)*, mémoire de diplôme de Master, Université Montpellier II Sciences et Techniques du Languedoc
- Baker RH, Wilkinson GS, DeSalle R (2001) Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). *Systematic Biology*, **50**, 87-105.
- Ballard JWO (2000) When one is not enough: Introgression of mitochondrial DNA in *Drosophila*. *Molecular Biology and Evolution*, **17**, 1126-1130.

- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729-744.
- Barja G, Herrero A (2000) Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. *Faseb Journal*, **14**, 312-318.
- Barker FK, Lutzoni FM (2002) The utility of the incongruence length difference test. *Systematic Biology*, **51**, 625-637.
- Barrowclough GF, Zink RM (2009) Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. *Molecular Ecology*, **18**, 2934-2936.
- Barton N, Jones JS (1983) Mitochondrial DNA: New clues about evolution. *Nature*, **306**, 317-318.
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology Evolution and Systematics*, **16**, 113-148.
- Barton NH, Hewitt GM (1989) Adaptation, speciation, and hybrid zones. *Nature*, **341**, 497-503.
- Baum DA, Small RL, Wendel JF (1998) Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Systematic Biology*, **47**, 181-207.
- Baur E (1909) Das Wesen und die Erblichkeitsverhältnisse der Varietates albomarginatae hort von *Pelargonium zonale*. *Zschr. f. indukt. Abst.- und Vererbungslehre*, **1**, 330-351.
- Berli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763-773.
- Bell CD, Patterson RW (2000) Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). *American Journal of Botany*, **87**, 1857-1870.
- Bellstedt DU, Linder HP, Harley EH (2001) Phylogenetic relationships in *Disa* based on non-coding trnL-trnF chloroplast sequences: Evidence of numerous repeat regions. *American Journal of Botany*, **88**, 2088-2100.
- Bendich AJ (2004) Circular chloroplast chromosomes: The grand illusion. *Plant Cell*, **16**, 1661-1666.
- Berg CC (1986) Subdivision of *Ficus* Subg *Urostigma* Sect *Galoglychia* (Moraceae). *Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen Series C-Biological and Medical Sciences*, **89**, 121-127.
- Berg CC, Wiebes JT (1992) *African fig trees and fig wasps*, Amsterdam: North Holland.
- Bergsten J (2005) A review of long-branch attraction. *Cladistics*, **21**, 163-193.

- Bergsten J, Miller BK (2004) *Acilius* phylogeny (Coleoptera: Dytiscidae), problems with long-branch attraction and morphological intersexual coevolution. *Cladistics*, **20**, 76-77.
- Bergstrom CT, Pritchard J (1998) Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics*, **149**, 2135-2146.
- Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, **424**, 197-201.
- Berlin S, Ellegren H (2001) Evolutionary genetics: Clonal inheritance of avian mitochondrial DNA. *Nature*, **413**, 37-38.
- Bernatchez L, Glemet H, Wilson CC, Danzmann RG (1995) Introgression and fixation of Arctic Char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of Brook Trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 179-185.
- Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 3101-3109.
- Bicudo J, Bianco AC, Vianna CR (2002) Adaptive thermogenesis in hummingbirds. *Journal of Experimental Biology*, **205**, 2267-2273.
- Biek R, Walsh PD, Leroy EM, Real LA (2006) Recent common ancestry of Ebola Zaire virus found in a bat reservoir. *Plos Pathogens*, **2**, 885-886.
- Bilton DT, Mirol PM, Mascheretti S, *et al.* (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **265**, 1219-1226.
- Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 11331-11338.
- Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics*, **35**, 125-148.
- Birky CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513-527.

- Blank JL (1992) Phenotypic variation in physiological response to seasonal environments. In: *Mammalian energetics: interdisciplinary views of metabolism and reproduction* (eds. Tomasi TE, Horton TH), pp. 186–212. Comstock Publishing,, Ithaca, NY.
- Bogehagen DF (1999) Repair of mtDNA in vertebrates. *American Journal of Human Genetics*, **64**, 1276-1281.
- Bortiri E, Heuvel BV, Potter D (2006) Phylogenetic analysis of morphology in *Prunus* reveals extensive homoplasy. *Plant Systematics and Evolution*, **259**, 53-71.
- Bousquet J, Strauss SH, Li P (1992) Complete congruence between morphological and *Rbcl*-based molecular phylogenies in Birches and related species (Betulaceae). *Molecular Biology and Evolution*, **9**, 1076-1088.
- Brandley MC, Schmitz A, Reeder TW (2005) Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic Biology*, **54**, 373-390.
- Bromham L, Eyre-Walker A, Smith NH, Smith JM (2003) Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology & Evolution*, **18**, 2-4.
- Brower AVZ, DeSalle R, Vogler A (1996) Gene trees, species trees, and systematics: A cladistic perspective. *Annual Review of Ecology and Systematics*, **27**, 423-450.
- Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 1967-1971.
- Buckler ES, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics*, **145**, 821-832.
- Buckley TR, Arensburger P, Simon C, Chambers GK (2002) Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Systematic Biology*, **51**, 4-18.
- Buckley TR, Cordeiro M, Marshall DC, Simon C (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology*, **55**, 411-425.
- Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, **42**, 384-397.
- Burger G, Saint-Louis D, Gray MW, Lang BF (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell*, **11**, 1675-1694.

- Busack SD, Lawson R, Arjo WM (2005) Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphibia-Reptilia*, **26**, 239-256.
- Calder WAI (1984) *Size, function, and life history* Harvard University Press, Cambridge, MA.
- Castilla AM, Fernandez-Pedrosa V, Backeljau T, *et al.* (1998a) Conservation genetics of insular *Podarcis* lizards using partial cytochrome b sequences. *Molecular Ecology*, **7**, 1407-1411.
- Castilla AM, Fernandez-Pedrosa V, Harris DJ, *et al.* (1998b) Mitochondrial DNA divergence suggests that *Podarcis hispanica atrata* (Squamata : Lacertidae) from the Columbretes Islands merits specific distinction. *Copeia*, 1037-1040.
- Cavalier T (1970) Electron microscopic evidence for chloroplast fusion in zygotes of *Chlamydomonas reinhardtii*. *Nature*, **228**, 333-&.
- Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720-729.
- Chang JT (1996) Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Mathematical Biosciences*, **134**, 189-215.
- Chapman JW, Carlton JT, Bellinger MR, Blakeslee AMH (2007) Premature refutation of a human-mediated marine species introduction: the case history of the marine snail *Littorina littorea* in the Northwestern Atlantic. *Biological Invasions*, **9**, 995-1008.
- Charleston MA, Robertson DL (2002) Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Systematic Biology*, **51**, 528-535.
- Chase MW, Cowan RS, Hollingsworth PM, *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295-299.
- Chase MW, Soltis DE, Olmstead RG, *et al.* (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Annals of the Missouri Botanical Garden*, **80**, 528-580.
- Chen XA, Li S, Aksoy S (1999) Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journal of Molecular Evolution*, **48**, 49-58.
- Cheviron ZA, Brumfield RT (2009) Migration-selection balance and local adaptation of mitochondrial haplotypes in Rufous collared Sparrows (*Zonotrichia Capensis*) along an elevational gradient. *Evolution*, **63**, 1593-1605.

- Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17741-17746.
- Cho Y, Qiu YL, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14244-14249.
- Ciborowski KL, Consuegra S, de Leaniz CG, *et al.* (2007) Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA. *Biology Letters*, **3**, 554-557.
- Colless DH (1995) Relative symmetry of cladograms and phenograms: an experimental study. *Systematic Biology*, **44**, 102-108.
- Collura RV, Auerbach MR, Stewart CB (1996) A quick, direct method that can differentiate expressed mitochondrial genes from their nuclear pseudogenes. *Current Biology*, **6**, 1337-1339.
- Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432-438.
- Cook JM, Rasplus JY (2003) Mutualists with attitude: coevolving fig wasps and figs. *Trends in Ecology & Evolution*, **18**, 241-248.
- Correns C (1909) Zur Kenntniss der Roue von Kem und Plasma bei der Vererbung. *Zschr. f. indukt. Abst. und Vererbungslehre*, **1**, 291-329.
- Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443-1458.
- Coyne JA, Orr HA (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution*, **51**, 295-303.
- Crochet PA, Chaline O, Surget-Groba Y, Debain C, Cheylan M (2004) Speciation in mountains: phylogeography and phylogeny of the rock lizards genus *Iberolacerta* (Reptilia : Lacertidae). *Molecular Phylogenetics and Evolution*, **30**, 860-866.
- Cruzan MB, Arnold ML, Carney SE, Wollenberg KR (1993) cpDNA Inheritance in interspecific crosses and evolutionary inference in Louisiana Irises. *American Journal of Botany*, **80**, 344-350.
- Cunningham CW (1997a) Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution*, **14**, 733-740.

- Cunningham CW (1997b) Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology*, **46**, 464-478.
- Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *Plos Biology*, **2**, 2264-2274.
- Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution*, **62**, 1908-1920.
- Curtis SE, Clegg MT (1984) Molecular evolution of chloroplast DNA sequences. *Molecular Biology and Evolution*, **1**, 291-301.
- Darlu P, Lecomte G (2002) When does the incongruence length difference test fail? *Molecular Biology and Evolution*, **19**, 432-437.
- Dawid IB, Blackler AW (1972) Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. *Developmental Biology*, **29**, 152-&.
- Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *Plos Genetics*, **2**, 762-768.
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332-340.
- Dellaporta SL, Xu A, Sagasser S, *et al.* (2006) Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 8751-8756.
- Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK (2000) High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*, **289**, 2342-2344.
- Dowling TE, Moritz C, Palmer JD (1990) Nucleic acids II: restriction site analysis. In: *Molecular Systematics* (eds. Hillis DM, Moritz C), pp. 250-317. Sinauer Associates, Sunderland, MA.
- Doyle JJ (1992) Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144-163.
- Dyall SD, Brown MT, Johnson PJ (2004) Ancient invasions: From endosymbionts to organelles. *Science*, **304**, 253-257.
- Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, **18**, 2930-2933.

- Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839-1854.
- Ehinger M, Fontanillas P, Petit E, Perrin N (2002) Mitochondrial DNA variation along an altitudinal gradient in the greater white-toothed shrew, *Crocidura russula*. *Molecular Ecology*, **11**, 939-945.
- Eigen M, Schuster P (1977) Hypercycle: Principle of natural self organization .a. Emergence of hypercycle. *Naturwissenschaften*, **64**, 541-565.
- Embley TM, van der Giezen M, Horner DS, *et al.* (2003) Hydrogenosomes, mitochondria and early eukaryotic evolution. *Iubmb Life*, **55**, 387-395.
- Ephrussi B (1953) *in Nucleo-cytoplasmic relations in microorganisms: Their bearing on cell heredity and differentiation* Clarendon, Oxford, UK.
- Erasmus JC, van Noort S, Jousselin E, Greeff JM (2007) Molecular phylogeny of fig wasp pollinators (Agaonidae, Hymenoptera) of *Ficus* section *Galoglychia*. *Zoologica Scripta*, **36**, 61-78.
- Erickson DL, Spouge J, Resch A, Weigt LA, Kress WJ (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon*, **57**, 1304-1316.
- Escobar-Paramo P, Sabbagh A, Darlu P, *et al.* (2004) Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Molecular Phylogenetics and Evolution*, **30**, 243-250.
- EyreWalker A, Gaut BS (1997) Correlated rates of synonymous site evolution across plant genomes. *Molecular Biology and Evolution*, **14**, 455-460.
- Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics*, **10**, 315-319.
- Farris JS, Kallersjo M, Kluge AG, Bult C (1995) Constructing a significance test for incongruence. *Systematic Biology*, **44**, 570-572.
- Fay JC, Wu CI (2003) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.
- Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, **27**, 401-410.
- Fontaine KM, Cooley JR, Simon C (2007) Evidence for paternal leakage in hybrid periodical cicadas (Hemiptera: *Magicicada* spp.). *PLoS ONE*, **2**, e892.
- Fontanillas P, Depraz A, Giorgi MS, Perrin N (2005) Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. *Molecular Ecology*, **14**, 661-670.

- Ford CS, Ayres KL, Toomey N, *et al.* (2009) Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society*, **159**, 1-11.
- Forterre P (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie*, **87**, 793-803.
- Fow CW, Roff DA, Fairbairn DJ (2001) *Evolutionary ecology: concepts and case studies* Oxford University Press, USA.
- Francisco JF, Brown GG, Simpson MV (1979) Further studies on Type A and Type B Rat mtDNA: Cleavage maps and evidence for cytoplasmic inheritance in mammals. *Plasmid*, **2**, 426-436.
- Freudenstein JV, Chase MW (2001) Analysis of mitochondrial nad1b-c intron sequences in Orchidaceae: Utility and coding of length-change characters. *Systematic Botany*, **26**, 643-657.
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **34**, 397-423.
- Gabalton T, Huynen MA (2003) Reconstruction of the proto-mitochondrial metabolism. *Science*, **301**, 609-609.
- Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, **155**, 981-987.
- Garcia-Moreno J (2004) Is there a universal mtDNA clock for birds? *Journal of Avian Biology*, **35**, 465-468.
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 10274-10279.
- Geniez P (2001) *Variation géographique des lézards du genre Podarcis (Reptilia, Sauria, Lacertidae) dans la péninsule Ibérique, l'Afrique du nord et le sud de la France*, Université Montpellier II & Ecole Pratique des Hautes Etudes.
- Gillespie JH (2000) Genetic drift in an infinite population: The pseudohitchhiking model. *Genetics*, **155**, 909-919.
- Gillooly JF, Allen AP, Brown JH, *et al.* (2005) The metabolic basis of whole-organism RNA and phosphorus content. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 11923-11927.

- Glockner G, Rosenthal A, Valentin K (2000) The structure and gene repertoire of an ancient red algal plastid genome. *Journal of Molecular Evolution*, **51**, 382-390.
- Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, **49**, 652-670.
- Gómez-Campo C, Bermudez-de-Castro L, Cagiga MJ, Sanchez-Yelamo MD (1984) Endemism in the iberian peninsula. *Webbia*, **38**, 709-714.
- Govindaraju DR, Dancik BP, Wagner DB (1989) Novel chloroplast DNA polymorphism in a sympatric region of two pines. *Journal of Evolutionary Biology*, **2**, 49-59.
- Grant PR, Grant BR (1992) Hybridization in bird species. *Science*, **256**, 193-197.
- Gray MW, Burger G, Cedergren R, *et al.* (1999) A genomics approach to mitochondrial evolution. *Biological Bulletin*, **196**, 400-402.
- Gray MW, Lang BF, Cedergren R, *et al.* (1998) Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Research*, **26**, 865-878.
- Grigg GC, Beard LA, Augee ML (2004) The evolution of endothermy and its diversity in mammals and birds. *Physiological and Biochemical Zoology*, **77**, 982-997.
- Guillaume CP (1987) *Les petits Lacertidés du bassin méditerranéen occidental (Genera Podarcis et Archaeolacerta essentiellement). Sur quelques problèmes d'ordre systématique et biogéographique*, diplôme de Doctorat, Université des Sciences et Techniques du Languedoc.
- Guillaumin M, Descimon H (1976) La notion d'espèce chez les lépidoptères. In: *Les Problèmes de l'Espèce dans le Règne Animal (Volume 1)* (ed. Bosquet C), pp. 129-201. Société zoologique de France.
- Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.
- Gyllensten U, Wilson AC (1987) Interspecific mitochondrial DNA transfer and the colonization of scandinavia by mice. *Genetical Research*, **49**, 25-29.
- Hafner MS, Nadler SA (1988) Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*, **332**, 258-259.
- Hampl V, Cepicka I, Flegr J, Tachezy J, Kulda J (2004) Critical analysis of the topology and rooting of the parabasalian 16S rRNA tree. *Molecular Phylogenetics and Evolution*, **32**, 711-723.
- Hansen AK, Escobar LK, Gilbert LE, Jansen RK (2007) Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implications for phylogenetic studies. *American Journal of Botany*, **94**, 42-46.

- Hardy OJ, Maggia L, Bandou E, *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15**, 559-571.
- Harris DJ, Carranza S, Arnold EN, Pinho C, Ferrand N (2002) Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, **29**, 1257-1262.
- Harris DJ, Sá Sousa P (2001) Species distinction and relationships of the western Iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetological Journal*, **11**, 129-136.
- Harris DJ, Sá Sousa P (2002) Molecular phylogenetics of Iberian wall lizards (*Podarcis*): Is *Podarcis hispanica* a species complex? *Molecular Phylogenetics and Evolution*, **23**, 75-81.
- Hartmann S, Nason JD, Bhattacharya D (2001) Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. *Journal of Molecular Evolution*, **53**, 124-134.
- Hastings IM (1992) Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. *Genetical Research*, **59**, 215-225.
- Havey MJ (1997) Predominant paternal transmission of the mitochondrial genome in cucumber. *Journal of Heredity*, **88**, 232-235.
- Hayashi JI, Yonekawa H, Gotoh O, Watanabe J, Tagashira Y (1978) Strictly maternal inheritance of rat mitochondrial DNA. *Biochemical and Biophysical Research Communications*, **83**, 1032-1038.
- Heard SB (1992) Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution*, **46**, 1818-1826.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 313-321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.
- Herre EA, Machado CA, Bermingham E, *et al.* (1996) Molecular phylogenies of figs and their pollinator wasps. *Journal of Biogeography*, **23**, 521-530.
- Heuertz M, Koffi GK, Hardy OJ (*in press*) Evidence for chloroplast DNA recombination in a tropical tree and implications for phylogeographic inference. *Molecular Biology and Evolution*.

- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.
- Hewitt GM (2001) Speciation, hybrid zones and phylogeography: or seeing genes in space and time. *Molecular Ecology*, **10**, 537-549.
- Hipp AL, Hall JC, Sytsma KJ (2004) Congruence versus phylogenetic accuracy: Revisiting the incongruence length difference test. *Systematic Biology*, **53**, 81-89.
- Hiratsuka J, Shimada H, Whittier R, *et al.* (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct transfer RNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular & General Genetics*, **217**, 185-194.
- Hoberg EP, Brooks DR, Siegel-Causey D (1997) Host-parasite cospeciation: history, principles, and prospects. In: *Host-parasite evolution: general principles and avian models* (eds. Clayton DH, Moore J), pp. 212-235. Oxford Univ. Press, Oxford.
- Hoekstra HE (2003) Unequal transmission of mitochondrial haplotypes in natural populations of field mice with XY females (genus *Akodon*). *American Naturalist*, **161**, 29-39.
- Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics*, **131**, 397-412.
- Holder K, Montgomerie R, Friesen VL (1999) A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear dna sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution*, **53**, 1936-1950.
- Holland BR, Penny D, Hendy MD (2003) Outgroup misplacement and phylogenetic inaccuracy under a molecular clock: A simulation study. *Systematic Biology*, **52**, 229-238.
- Hollingsworth ML, Clark AA, Forrest LL, *et al.* (2009a) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, **9**, 439-457.
- Hollingsworth PM, Forrest LL, Spouge JL, *et al.* (2009b) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794-12797.
- Howell N, Oostra RJ, Bolhuis PA, *et al.* (2003) Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. *American Journal of Human Genetics*, **72**, 1460-1469.
- Hudson RR (1987) Estimating the recombination parameter of a finite population model without Selection. *Genetical Research*, **50**, 245-250.

- Hudson RR (1992) Gene trees, species trees and the segregation of ancestral alleles. *Genetics*, **131**, 509-512.
- Hudson RR, Turelli M (2003) Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182-190.
- Huelsenbeck JP, Hillis DM, Jones R (1996) Parametric bootstrapping in molecular phylogenetics: application and performance. In: *Parametric bootstrapping in molecular phylogenetics: application and performance* (eds. Ferraris JD, Palumbi SR). John Wiley & Sons, New York.
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 1525-1534.
- Hurst LD, Hamilton WD (1992) Cytoplasmic fusion and the nature of sexes. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **247**, 189-194.
- Hurst LD, Peck JR (1996) Recent advances in understanding of the evolution and maintenance of sex. *Trends in Ecology & Evolution*, **11**, A46-A52.
- Hutchison C, Newbold JE, Potter SS, Edgell MH (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature*, **251**, 536-538.
- Huyse T, Volckaert FAM (2005) Comparing host and parasite phylogenies: *Gyrodactylus* flatworms jumping from goby to goby. *Systematic Biology*, **54**, 710-718.
- Innan H, Nordborg M (2002) Recombination or mutational hot spots in human mtDNA? *Molecular Biology and Evolution*, **19**, 1122-1127.
- Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome b gene of mammals. *Journal of Molecular Evolution*, **32**, 128-144.
- Jackson AP (2004) A reconciliation analysis of host switching in plant-fungal symbioses. *Evolution*, **58**, 1909-1923.
- Jackson JA (1999) Analysis of parasite host-switching: limitations on the use of phylogenies. *Parasitology*, **119**, S111-S123.
- Jaenike J (2001) Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics*, **32**, 25-49.
- Jiggins FM (2003) Male-killing *Wolbachia* and mitochondrial DNA: Selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, **164**, 5-12.

- Jiggins FM, Bentley JK, Majerus MEN, Hurst GDD (2001) How many species are infected with *Wolbachia*? Cryptic sex ratio distorters revealed to be common by intensive sampling. *Proceedings of the Royal Society B-Biological Sciences*, **268**, 1123-1126.
- Johnson AA, Johnson KA (2001) Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase. *Journal of Biological Chemistry*, **276**, 38090-38096.
- Johnson LA, Soltis DE (1994) Matk DNA sequences and phylogenetic reconstruction in Saxifragaceae. *Systematic Botany*, **19**, 143-156.
- Jousselin E, Rasplus JY, Kjellberg F (2003) Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus*. *Evolution*, **57**, 1255-1269.
- Jousselin E, van Noort S, Berry V, *et al.* (2008) One fig to bind them all: Host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. *Evolution*, **62**, 1777-1797.
- Kimura M (1983) *The neutral theory of molecular evolution* Cambridge University Press, Cambridge, UK.
- King DA (1996) Allometry and life history of tropical trees. *Journal of Tropical Ecology*, **12**, 25-44.
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution*, **29**, 170-179.
- Kivisild T, Shen PD, Wall DP, *et al.* (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373-387.
- Klak C, Khunou A, Reeves G, Hedderson T (2003) A phylogenetic hypothesis for the Aizoaceae (Caryophyllales) based on four plastid DNA regions. *American Journal of Botany*, **90**, 1433-1445.
- Klein NK, Payne RB (1998) Evolutionary associations of brood parasitic finches (*Vidua*) and their host species: Analyses of mitochondrial DNA restriction sites. *Evolution*, **52**, 566-582.
- Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics*, **50**, 149-159.
- Ko KS, Jung HS (2002) Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum*. *Molecular Phylogenetics and Evolution*, **23**, 112-122.
- Kolodner RD, Tewari KK (1975) Chloroplast DNA from higher plants replicates by both Cairns and rolling circle mechanism. *Nature*, **256**, 708-711.

- Kondo R, Satta Y, Matsuura *et al.* (1990) Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics*, **126**, 657-663.
- Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. *Trends in Genetics*, **21**, 647-654.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369-8374.
- Kroon AM, Vos WMD, Bakker H (1978) Heterogeneity of rat liver mitochondrial DNA. *Biochimica et Biophysica Acta*, **519**, 269-273.
- Kujoth GC, Bradshaw PC, Haroon S, Prolla TA (2007) The role of mitochondrial DNA mutations in mammalian aging. *Plos Genetics*, **3**, 161-173.
- Lahaye R, Van der Bank M, Bogarin D, *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923-2928.
- Lakshmipathy U, Campbell C (1999) Double strand break rejoining by mammalian mitochondrial extracts. *Nucleic Acids Research*, **27**, 1198-1204.
- Lecointre G, Rachdi L, Darlu P, Denamur E (1998) *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Molecular Biology and Evolution*, **15**, 1685-1695.
- Lee WJ, Kocher TD (1995) Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics*, **139**, 873-887.
- Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends in Genetics*, **21**, 655-663.
- Lester L, Meade A, Pagel M (2006) The slow road to the eukaryotic genome. *Bioessays*, **28**, 57-64.
- Li WH, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution*, **25**, 330-342.
- Liston A, Kadereit JW (1995) Chloroplast DNA evidence for introgression and long distance dispersal in the desert annual *Senecio flavus* (Asteraceae). *Plant Systematics and Evolution*, **197**, 33-41.
- Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D (1996) Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis.

- Proceedings of the National Academy of Sciences of the United States of America*, **93**, 1930-1934.
- Loeb LA (2001) A mutator phenotype in cancer. *Cancer Research*, **61**, 3230-3239.
- Lonsdale DM, Brears T, Hodge TP, Melville SE, Rottmann WH (1988) The plant mitochondrial genome: Homologous recombination as a mechanism for generating heterogeneity. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **319**, 149-163.
- Lopez-Vaamonde C, Rasplus JY, Weiblen GD, Cook JM (2001) Molecular phylogenies of fig wasps: Partial cocladogenesis of pollinators and parasites. *Molecular Phylogenetics and Evolution*, **21**, 55-71.
- Losos JB, Jackman TR, Larson A, de Queiroz K, Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115-2118.
- Lunt DH, Hyman BC (1997) Animal mitochondrial DNA recombination. *Nature*, **387**, 247-247.
- Lynch M (2002) Intron evolution as a population genetic process. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6118-6123.
- Lynch M (2006) The origins of eukaryotic gene structure. *Molecular Biology and Evolution*, **23**, 450-468.
- Lynch M, Conery JS (2003) The origins of genome complexity. *Science*, **302**, 1401-1404.
- Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. *Science*, **311**, 1727-1730.
- MacDonald D, Barrett P (1995) *European Mammals. Evolution and Behaviour* Harper Collins.
- Macey JR, Larson A, Ananjeva NB, Papenfuss TJ (1997) Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. *Molecular Biology and Evolution*, **14**, 30-39.
- Machado CA, Jouselin E, Kjellberg F, Compton SG, Herre EA (2001) Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **268**, 685-694.
- Machado CA, Robbins N, Gilbert MTP, Herre EA (2005) Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6558-6565.

- Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.
- Maddison WP, Maddison DR (2004) Mesquite: A modular system for evolutionary analysis. <http://mesquiteproject.org>.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229-237.
- Martin AP, Naylor GJP, Palumbi SR (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**, 153-155.
- Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4087-4091.
- Martin NH, Sapir Y, Arnold ML (2008) The genetic architecture of reproductive isolation in Louisiana irises: Pollination syndromes and pollinator preferences. *Evolution*, **62**, 740-752.
- Martin W, Rujan T, Richly E, *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12246-12251.
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **358**, 59-83.
- Martinez-Solano I, Teixeira J, Buckley D, Garcia-Paris M (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, **15**, 3375-3388.
- Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325-1335.
- Maruyama T, Birky CW (1991) Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. *Genetics*, **127**, 449-451.
- Mateo JA, LopezJurado LF, Guillaume CP (1996) Proteic and morphological variations in ocellated lizards (Lacertidae): A complex of species across the Strait of Gibraltar. *Comptes Rendus de l'Academie des Sciences Serie III-Sciences de la vie - Life Sciences*, **319**, 737-746.

- Melo-Ferreira J, Boursot P, Randi E, *et al.* (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, **16**, 605-618.
- Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459-2464.
- Menken SBJ (1996) Pattern and process in the evolution of insect-plant associations: *Yponomeuta* as an example. *Entomologia Experimentalis Et Applicata*, **80**, 297-305.
- Michaloud G, Michaloudpelletier S, Wiebes JT, Berg CC (1985) The Co-occurrence of two pollinating species of fig wasp and one species of fig. *Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen Series C-Biological and Medical Sciences*, **88**, 93-119.
- Mildbraed J, Burret M (1911) Die Afrikanischen Arten der Gattung *Ficus*. *Linn. Bot. Jarhb.*, **46**, 163-269.
- Millen RS, Olmstead RG, Adams KL, *et al.* (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*, **13**, 645-658.
- Milligan BG (1992) Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. *American Journal of Botany*, **79**, 1325-1328.
- Mindell DP, Knight A, Baer C, Huddleston CJ (1996) Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. *Molecular Biology and Evolution*, **13**, 422-426.
- Minelli A (2007) Invertebrate taxonomy and evolutionary developmental biology. *Zootaxa*, 55-60.
- Mishmar D, Ruiz-Pesini E, Golik P, *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 171-176.
- Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*, **83**, 383-404.
- Molbo D, Machado CA, Sevenster JG, Keller L, Herre EA (2003) Cryptic species of fig-pollinating wasps: Implications for the evolution of the fig-wasp mutualism, sex allocation, and precision of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5867-5872.

- Moore WS (1995) Inferring phylogenies from mtDNA variation: Mitochondrial gene trees versus nuclear gene trees. *Evolution*, **49**, 718-726.
- Moran P, Kornfield I (1993) Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei, Cichlidae) of lake Malawi. *Molecular Biology and Evolution*, **10**, 1015-1029.
- Moran P, Kornfield I (1995) Were population bottlenecks associated with the radiation of the Mbuna species flock (Teleostei, Cichlidae) of lake Malawi. *Molecular Biology and Evolution*, **12**, 1085-1093.
- Mose MJ (1961) Spermiogenesis in the crayfish (*Procambarus clarkii*). I. Structural characterization of the mature sperm. *Journal of Biophysical and Biochemical Cytology*, **9**, 222-228.
- Mounolou JC, Lacroute F (2005) Mitochondrial DNA: an advance in eukaryotic cell biology in the 1960s. *Biology of the Cell*, **97**, 743-748.
- Muller M (1998). In: *Evolutionary relationships among Protozoa* (eds. Coombs GH, Vickerman MA, Sleigh MA, Warren A), pp. 109-111, Kluwer, Dordrecht, Netherlands.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853-858.
- Nabholz B, Glemin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals: The longevity hypothesis. *Molecular Biology and Evolution*, **25**, 120-130.
- Nicholls H (2003) DNA: the barcode of life? Originally published on behalf of Elsevier by BioMedNet. Disponible sur <http://www.uoguelph.ca/~phebert/media/BioMedNet%20News%20article.pdf>.
- Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, **16**, 358-364.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics*, **158**, 885-896.
- Ogihara Y, Terachi T, Sasakuma T (1992) Structural analysis of length mutations in a hot spot region of wheat chloroplast DNAs. *Current Genetics*, **22**, 251-258.
- Ohta T (1993) An examination of the generation time effect on molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 10676-10680.

- Oldenburg DJ, Bendich AJ (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *Journal of Molecular Biology*, **335**, 953-970.
- Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: A review of methods and data Analysis. *American Journal of Botany*, **81**, 1205-1224.
- Olmstead RG, Reeves PA (1995) Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcl* and *ndhf* sequences. *Annals of the Missouri Botanical Garden*, **82**, 176-193.
- Olsson M, Shine R (2003) Female-biased natal and breeding dispersal in an alpine lizard, *Niveoscincus microlepidotus*. *Biological Journal of the Linnean Society*, **79**, 277-283.
- Page RDM (1998) GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics*, **14**, 819-820.
- Page RDM, Charleston MA (1997) From gene to organismal phylogeny: Reconciled trees and the gene tree species tree problem. *Molecular Phylogenetics and Evolution*, **7**, 231-240.
- Page RDM, Hafner MS (1996) Molecular phylogenies of host-parasite cospeciation: Gophers and lice as a model system. In: *New uses for new phylogenies* (eds. Harvey PH, Brown AJL, Maynard Smith J, Nee S), pp. 255-270. Oxford Univ. Press, Oxford.
- Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure but slowly in sequence. *Journal of Molecular Evolution*, **28**, 87-97.
- Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution*, **55**, 859-868.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568-583.
- Parra-Olea G, Wake DB (2001) Extreme morphological and ecological homoplasy in tropical salamanders. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 7888-7891.
- Paterson AM, Palma RL, Gray RD (1999) How frequently do avian lice miss the boat? Implications for coevolutionary studies. *Systematic Biology*, **48**, 214-223.
- Patton JL, Smith MF (1994) Paraphyly, polyphyly, and the nature of species boundaries in Pocket gophers (genus *Thomomys*). *Systematic Biology*, **43**, 11-26.
- Percy DM, Page RDM, Cronk QCB (2004) Plant-insect interactions: Double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology*, **53**, 120-127.

- Pérez Cueva A (1994) *Atlas climático de la comunidad Valenciana* Comunidad Valenciana, Valencia.
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution*, **55**, 501-508.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, **37**, 187-214.
- Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, **21**, 2319-2325.
- Pinho C, Ferrand N, Harris DJ (2006) Reexamination of the Iberian and North African *Podarcis* (Squamata : Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, **38**, 266-273.
- Pinho C, Harris DJ, Ferrand N (2003) Genetic polymorphism of 11 allozyme loci in populations of wall lizards (*Podarcis sp.*) from the Iberian Peninsula and North Africa. *Biochemical Genetics*, **41**, 343-359.
- Pinho C, Harris DJ, Ferrand N (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis spp.*) are an assemblage of incipient species. *Bmc Evolutionary Biology*, **8**.
- Pol D, Siddall ME (2001) Biases in maximum likelihood and parsimony: A simulation approach to a 10-taxon case. *Cladistics*, **17**, 266-281.
- Pont-Kingdon G, Okada NA, Macfarlane JL, *et al.* (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial *MutS*: A possible case of gene transfer from the nucleus to the mitochondrion. *Journal of Molecular Evolution*, **46**, 419-431.
- Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution*, **54**, 396-402.
- Prentice HC, Malm JU, Hathaway L (2008) Chloroplast DNA variation in the European herb *Silene dioica* (red campion): postglacial migration and interspecific introgression. *Plant Systematics and Evolution*, **272**, 23-37.
- Price T (1997) Correlated evolution and independent contrasts. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **352**, 519-529.
- Price TD, Bouvier MM (2002) The evolution of F-1 postzygotic incompatibilities in birds. *Evolution*, **56**, 2083-2089.
- Ramirez MJ (2006) Further problems with the incongruence length difference test: "hypercongruence" effect and multiple comparisons. *Cladistics*, **22**, 289-295.

- Ramírez WB (1970) Host specificity of fig wasps (Agaonidae). *Evolution*, **24**, 680-691.
- Rand DM (1994) Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology & Evolution*, **9**, 125-131.
- Rannala B, Bertorelle G (2001) Using linked markers to infer the age of a mutation. *Human Mutation*, **18**, 87-100.
- Rasplus JY (1996) The one-to-one species specificity of the *Ficus*-Agaoninae mutualism: How casual? *Biodiversity of African Plants*, 639-649.
- Raufaste N, Orth A, Belkhir K, *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593-616.
- Reed DL, Light JE, Allen JM, Kirchman JJ (2007) Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. *Bmc Biology*, **5**.
- Reilly JG, Thomas CA (1980) Length polymorphisms, restriction site variation, and maternal inheritance of mitochondrial DNA of *Drosophila melanogaster*. *Plasmid*, **3**, 109-115.
- Richly E, Chinnery PF, Leister D (2003) Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends in Genetics*, **19**, 356-362.
- Ricklefs RE, Fallon SM, Bermingham E (2004) Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. *Systematic Biology*, **53**, 111-119.
- Rieseberg LH, Beckstromsternberg S, Doan K (1990) *Helianthus annuus ssp texanus* has chloroplast DNA and nuclear ribosomal-RNA genes of *Helianthus debilis ssp cucumerifolius*. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 593-597.
- Rieseberg LH, Choi HC, Ham D (1991) Differential cytoplasmic *versus* nuclear introgression in *Helianthus*. *Journal of Heredity*, **82**, 489-493.
- Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature Genetics*, **37**, 96-100.
- Rogers JS (1993) Response of colless tree imbalance to number of terminal taxa. *Systematic Biology*, **42**, 102-105.
- Rønsted N, Salvo G, Savolainen V (2007a) Biogeographical and phylogenetic origins of African fig species (*Ficus* section Galoglychia). *Molecular Phylogenetics and Evolution*, **43**, 190-201.
- Rønsted N, Weiblen GD, Cook JM, *et al.* (2005) 60 million years of co-divergence in the fig-wasp symbiosis. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 2593-2599.

- Rønsted N, Yektaei-Karin E, Turk K, Clarkson JM, Chase MW (2007b) Species-level phylogenetics of large genera: prospects of studying co-evolution and polyploidy. In: *Reconstructing the tree of life: taxonomy and systematics of species rich taxa* (eds. Hodkinson T, Parnell J, Waldren S). Systematic Association Series. CRC Press, Boca Raton, Florida.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223-226.
- Ryberg K, Olsson M, Wapstra E, *et al.* (2004) Offspring-driven local dispersal in female sand lizards (*Lacerta agilis*). *Journal of Evolutionary Biology*, **17**, 1215-1220.
- Sá Sousa P (2000) A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetological Journal*, **10**, 1-11.
- Sagan L (1967) On Origin of Mitosing Cells. *Journal of Theoretical Biology*, **14**, 225-274.
- Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, **78**, 629-644.
- Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics*, **156**, 879-891.
- Schoonhoven LM, Jermy T, van Loon JJA (1998) *Insect–plant biology. From physiology to evolution* Chapman and Hall, London.
- Schwartz M, Vissing J (2002) Paternal inheritance of mtDNA in a patient with mitochondrial myopathy. *European Journal of Human Genetics*, **10**, 239-239.
- Shaw J, Lickey EB, Beck JT, *et al.* (2005) The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142-166.
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, **94**, 275-288.
- Shaw KL (1999) A nested analysis of song groups and species boundaries in the Hawaiian cricket genus *Laupala*. *Molecular Phylogenetics and Evolution*, **11**, 332-341.
- Shearer TL, Van Oppen MJH, Romano SL, Worheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, **11**, 2475-2487.
- Sherengul W, Kondo R, Matsuura ET (2006) Analysis of paternal transmission of mitochondrial DNA in *Drosophila*. *Genes & Genetic Systems*, **81**, 399-404.

- Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, **51**, 492-508.
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114-1116.
- Shoubridge EA, Wai T (2007) Mitochondrial DNA and the mammalian oocyte. In: *Mitochondrion in the Germline and Early Development*, pp. 87-111.
- Siddall ME, Whiting MF (1999) Long-branch abstractions. *Cladistics*, **15**, 9-24.
- Smith GR (1992) Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. *Systematic Biology*, **41**, 41-57.
- Sogin ML (1991) Early evolution and the origin of eukaryotes. *Current Opinion in Genetics and Development*, **1**, 457-463.
- Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353-373.
- Soltis DE, Soltis PS, Thompson JN, Pellmyr O (1992) Chloroplast DNA variation in *Lithophragma* (Saxifragaceae). *Systematic Botany*, **17**, 607-619.
- Somero GN (2002) Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. *Integrative and Comparative Biology*, **42**, 780-789.
- Sota T, Vogler AP (2001) Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Systematic Biology*, **50**, 39-59.
- Stace CA (1975) *Hybridization and the Flora of the British Isles* Academic Press.
- Stace CA (1997) *The New Flora of the British Isles, (2nd edn)* Cambridge University Press.
- Stadler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11730-11735.
- Steele KP, Vilgalys R (1994) Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany*, **19**, 126-142.
- Stevens PF (1984) Homology and phylogeny: Morphology and systematics. *Systematic Botany*, **9**, 395-409.
- Stoeckle M (2003) Taxonomy, DNA, and the bar code of life. *Bioscience*, **53**, 796-797.
- Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence specific oligonucleotide probes. *American Journal of Human Genetics*, **48**, 370-382.

- Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **269**, 137-142.
- Sullivan J, Holsinger KE, Simon C (1995) Among site rate variation and phylogenetic analysis of 12s ribosomal-RNA in Sigmodontine rodents. *Molecular Biology and Evolution*, **12**, 988-1001.
- Swofford DL, Waddell PJ, Huelsenbeck JP, *et al.* (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology*, **50**, 525-539.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DAA polymorphism. *Genetics*, **123**, 585-595.
- Takahata N, Slatkin M (1984) Mitochondrial Gene Flow. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **81**, 1764-1767.
- Tang YY, Manfredi G, Hirano M, Schon EA (2000) Maintenance of human rearranged mitochondrial DNAs in long-term cultured transmitochondrial cell lines. *Molecular Biology of the Cell*, **11**, 2349-2358.
- Tateno Y, Nei M, Tajima F (1982) Accuracy of estimated phylogenetic trees from molecular data .1. Distantly related species. *Journal of Molecular Evolution*, **18**, 387-404.
- Tegelstrom H (1987) Transfer of mitochondrial DNA from the Northern red-backed Vole (*Clethrionomys rutilus*) to the Bank Vole (*Clethrionomys glareolus*). *Journal of Molecular Evolution*, **24**, 218-227.
- Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution*, **37**, 221-244.
- Thao ML, Moran NA, Abbot P, *et al.* (2000) Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Applied and Environmental Microbiology*, **66**, 2898-2905.
- Thyagarajan B, Padua RA, Campbell C (1996) Mammalian mitochondria possess homologous DNA recombination activity. *Journal of Biological Chemistry*, **271**, 27536-27543.
- Tilney-Bassett RAE (1970) The control of plastid inheritance in *Pelargonium*. *Genetic Research, Cambridge*, **16**, 49-61.

- Trabelsi S, Gargouri-Bouزيد R, Vedel F, *et al.* (2005) Somatic hybrids between potato *Solanum tuberosum* and wild species *Solanum vernei* exhibit a recombination in the plastome. *Plant Cell Tissue and Organ Culture*, **83**, 1-11.
- Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, **22**, 925-933.
- Ujvari B, Dowton M, Madsen T (2007) Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biology Letters*, **3**, 189-192.
- Underhill JG (2008) *Phylogeny and codivergence in the fig-fig wasp mutualism: Sycoecine and Agaonid fig wasps (Chalcidoidea, Hymenoptera) associated with Ficus section Galoglychia (Moraceae)*. Diplôme de Master -Faculté des Science de Cape Town.
- Ursprung H, Schabtac (1965) Fertilization in Tunicates: Loss of paternal mitochondrion prior to sperm entry. *Journal of Experimental Zoology*, **159**, 379-&.
- Vaconcelos R, Carretero MA, Harris DJ (2006) Phylogeography of the genus *Blanus* (worm lizards) in Iberia and Morocco based on mitochondrial and nuclear markers preliminary analysis. *Amphibia-Reptilia*, **27**, 339-346.
- Vaillancourt RE, Petty A, McKinnon GE (2004) Maternal inheritance of mitochondria in *Eucalyptus globulus*. *Journal of Heredity*, **95**, 353-355.
- van der Niet T, Linder HP (2008) Dealing with incongruence in the quest for the species tree: A case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution*, **47**, 154-174.
- Van Droogenbroeck B, Kyndt T, Romeijn-Peeters E, *et al.* (2006) Evidence of natural hybridization and introgression between *Vasconcellea* species (Caricaceae) from southern Ecuador revealed by chloroplast, mitochondrial and nuclear DNA markers. *Annals of Botany*, **97**, 793-805.
- van Oppen MJH, Catmull J, McDonald BJ, *et al.* (2002) The mitochondrial genome of *Acropora tenuis* (Cnidaria : Scleractinia) contains a large group I intron and a candidate control region. *Journal of Molecular Evolution*, **55**, 1-13.
- Vences M, Wollenberg KC, Vieites DR, Lees DC (2009) Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, **24**, 456-465.
- Wang JL (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1395-1409.

- Wang JL (2009) A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148-2164.
- Wang JL, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429-446.
- Ward BL, Anderson RS, Bendich AJ (1981) The Mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, **25**, 793-803.
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1847-1857.
- Weiblen GD (2001) Phylogenetic relationships of fig wasps pollinating functionally dioecious *Ficus* based on mitochondrial DNA sequences and morphology. *Systematic Biology*, **50**, 243-267.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.
- Weiss S, Ferrand N (2006) *Phylogeography of Southern European Refugia: Evolutionary perspectives on the origins and conservation of European biodiversity* Springer.
- Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: *Molecular Systematics of Plants II* (eds. Soltis DE, Soltis PS, Doyle JJ), pp. 297-349. Kluwer Academic Publishers, Boston.
- Werren JH, Windsor DM (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 1277-1285.
- Wesselingh RA, Arnold ML (2000) Pollinator behaviour and the evolution of Louisiana iris hybrid zones. *Journal of Evolutionary Biology*, **13**, 171-180.
- Westerberhard MJ (1989) Phenotypic plasticity and the origins of diversity. *Annual Review of Ecology and Systematics*, **20**, 249-278.
- Wetzel DM, Hakansson S, Hu K, Roos D, Sibley LD (2003) Actin filament polymerization regulates gliding motility by apicomplexan parasites. *Molecular Biology of the Cell*, **14**, 396-406.
- Wheeler WC (1990) Nucleic acid sequence phylogeny and random outgroups. *Cladistics*, **6**, 363-367.
- White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**, 4925-4942.

- Wilcox TP, de Leon FJG, Hendrickson DA, Hillis DM (2004) Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. *Molecular Phylogenetics and Evolution*, **31**, 1101-1113.
- Wolfe AD, Randle CP (2004) Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: Implications for plant molecular systematics. *Systematic Botany*, **29**, 1011-1020.
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054-9058.
- Wolff JN, Gemell NJ (2008) Lost in the zygote: the dilution of paternal mtDNA upon fertilization. *Heredity*, **101**, 429-434.
- Wolstenholme DR (1992) Animal mitochondrial DNA structure and evolution. *International Review of Cytology*, **141**, 173-216.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 0097-0159.
- Wright SI, Charlesworth B (2004) The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics*, **168**, 1071-1076.
- Yaffe MP (1999) The machinery of mitochondrial inheritance and behavior. *Science*, **283**, 1493-1497.
- Yoder AD, Irwin JA, Payseur BA (2001) Failure of the ILD to determine data combinability for slow loris phylogeny. *Systematic Biology*, **50**, 408-424.
- Zhang DX, Hewitt GM (1996) Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247-251.
- Zillig W (1991) Comparative biochemistry of Archaea and Bacteria. *Current Opinion in Genetics & Development*, **1**, 544-551.
- Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107-2121.

ARTICLES

Article 1

Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex **p.131**

Julien P. Renoult, Philippe Geniez, Paul Bacquet, Laure Benoit et Pierre-André Crochet.

Molecular Ecology, 2009, doi: 10.1111/j.1365-294X.2009.04351.x

Article 2

*Cyto-nuclear discordance in the phylogeny of *Ficus* section Galoglychia and host shifts in plant-pollinator associations* **p.149**

Julien P. Renoult, Finn Kjellberg, Cinderella Grout, Sylvain Santoni, Bouchaïb Khadari

BMC Evolutionary Biology, in press.

Article 3

*Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) II: the valid name of the north-eastern Spanish form* **p.186**

Julien P. Renoult, Philippe Geniez, Paul Bacquet, Claude P. Guillaume, Pierre-André Crochet

Submitted to Zootaxa

Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex

JULIEN P. RENOULT,* PHILIPPE GENIEZ,+ PAUL BACQUET,‡ LAURE BENOIT§ and PIERRE-ANDRÉ CROCHET*

*CEFE-CNRS, UMR 5175, 1919 route de Mende, F-34293 Montpellier cedex 5, France, †CEFE-EPHE, UMR 5175, 1919 route de Mende, F-34293 Montpellier cedex 5, France, ‡Centre de Recherche BDIV, groupe Evolutionary Biology, Université Catholique de Louvain, Croix du sud, 4-5, Carnoy, b.108, B-1348, Louvain-la-Neuve, Belgium, §CEFE-CIRAD, UMR 5175, 1919 route de Mende, F-34293 Montpellier cedex 5, France

Abstract

Mitochondrial markers are still often used alone to identify evolutionary units, despite widespread evidence for processes such as incomplete lineage sorting or introgressive hybridization that may blur past population history. The combination of mitochondrial DNA data with other sources of information (morphology, nuclear genes) is a powerful tool to reveal when and why mitochondrial markers are potentially misleading. In this study, we evaluate the performance of mtDNA markers to unravel the evolutionary history of Spanish lizards from the *Podarcis hispanicus* species complex. We first uncover several cases of discordance between morphological and mitochondrial data in delimitation of taxa. To assess the origin of these discordances, we analysed the same populations using several independent nuclear loci. Both morphological and nuclear markers identified the same three evolutionary units in the region, while mitochondrial data revealed four deeply divergent lineages. We suggest here that the most likely scenario to explain this discordance is ancient mitochondrial introgression originating from a fourth evolutionary unit presently absent from the study area. Notably, this resulted in a complete replacement of the original lineage in a large part of the distribution of one of the taxa investigated. We discuss the potential evolutionary scenarios leading to this complete mitochondrial replacement and suggest why the previous studies have failed to recover the correct history of this species complex.

Keywords: discordance, introgression, morphology, phylogeography, *Podarcis hispanicus*

Received 10 March 2009; revision received 2 August 2009; accepted 11 August 2009

Introduction

An accurate description of biological diversity is a prerequisite for understanding its origin. Systematic studies thus lead to direct applications in fields as diverse as medical epidemiology (Biek *et al.* 2006), theoretical ecology (Anderson *et al.* 2004) and developmental biology (Minelli 2007) and provides the basic information for many conservation programmes (Vazquez & Gittle-

man 1998; Bininda-Emonds *et al.* 2000; Andreasen 2005).

A necessary step in such studies is to identify and delimit species. This task is sometimes wearisome in practice, mainly because markers used to identify evolutionary units may also be subject to selection, and therefore are intrinsically inadequate for inferring population histories (Avice 2004). For example, use of morphological variation for inferring relationships among populations often leads to errors because phenotypic variation reflects both population history and local adaptation or phenotypic plasticity (Avice 1994; Parra-Olea & Wake 2001). The use of neutral molecular

Correspondence: Julien P. Renoult and Pierre-André Crochet, Fax: +33 4672138; E-mails: julien.renoult@cefe.cnrs.fr and pierre-andre.crochet@cefe.cnrs.fr

markers should thus generally provide more accurate descriptions of population history than morphological characters alone (Avice 1994), hence molecular markers are increasingly used to identify and delimit species.

Until recently, most studies in molecular systematics and phylogenetics have used mitochondrial DNA (mtDNA; see Sunnucks 2000 for a review of the advantages of mtDNA). However, because gene trees do not necessarily reflect population history (Nichols 2001), the use of mtDNA alone in evolutionary studies has been criticized (Ballard & Whitlock 2004; Edwards *et al.* 2005). Well-known causes of discordance between mitochondrial and population histories include incomplete lineage sorting (Pamilo & Nei 1988; Jennings & Edwards 2005), which is especially problematic when multiple nodes in the tree are generated over a short period of time (Moran & Kornfield 1993; Klein & Payne 1998), and introgressive hybridization, i.e. exchange of genes between two evolutionary lineages (Seehausen 2004). Introgressive hybridization of mtDNA may have long-lasting consequences, particularly as the non-recombining mtDNA genome never becomes 'diluted' in the introgressed genetic background (Funk & Omland 2003; Morando *et al.* 2004). Moreover, introgression might be facilitated because mtDNA loci are less often linked to loci under negative selective pressure than nuclear loci (Harrison 1990) and can even undergo positive selection (i.e. Ruiz-Pesini *et al.* 2004). Intraspecific polymorphism in mtDNA lineages as a result of partial replacement of the original mtDNA by allospecific mtDNA has been documented in several organisms (e.g. Sullivan *et al.* 2004; Roca *et al.* 2005; Schelly *et al.* 2006). The most extreme outcome of complete replacement, in which the original mtDNA of a species has totally disappeared in part of its range, is much rarer (but see Bernatchez *et al.* 1995).

A classical procedure to detect misleading mtDNA relationships is to check for discordance with independent sets of markers. Agreement between results obtained with multiple types of markers provides strong support for the inferred evolutionary relationships between populations or taxa (Cummings *et al.* 1995). Conflicting results are also helpful, as they improve our understanding of the historical events that have led to the extant distribution of evolutionary lineages (Shaw 2002; Morando *et al.* 2004). This is particularly true in hotspots of biodiversity where evolutionary histories of populations are often difficult to unravel (Gillespie & Roderick 2002; Ennos *et al.* 2005).

The Iberian Peninsula was a major glacial refugium during the climatic oscillations of the late Pleistocene (Hewitt 1996; Taberlet *et al.* 1998) and it is part of one of the world's biodiversity hotspots (Myers *et al.* 2000). The Iberian reptile fauna is representative of this

biodiversity: its 45 reptile species include eleven (24%) endemic species and several species include species-level genetic lineages which are still not formally recognized (e.g. Mateo *et al.* 1996; Martínez-Solano *et al.* 2006; Vaconcelos *et al.* 2006). One of the most obvious cases of such unrecognized biodiversity is the lizard genus *Podarcis* Wagler, 1830 (Reptilia, Lacertidae). Five species are currently recognized in continental Iberian Peninsula: *Podarcis muralis* (Laurenti, 1768), *P. hispanicus* (Steindachner, 1870) sensu lato (see Geniez *et al.* 2007), *P. vaucheri* (Boulenger, 1905), *Podarcis bocagei* (Seoane, 1884) and *P. carbonelli* (Pérez-Mellado, 1981). The last two species occur in most places in sympatry with *P. hispanicus*, from which they differ in morphology, coloration and ecology, as well as genetically. They are thus clearly valid species. Within the traditional *P. hispanicus*, nuclear, mitochondrial and morphological markers delimit several distinct lineages that do not form a monophyletic unit when compared with *P. carbonelli* and *P. bocagei* (Sá Sousa 2000; Harris & Sá Sousa 2001, 2002; Harris *et al.* 2002; Pinho *et al.* 2003, 2006, 2007) suggesting that *P. hispanicus* as currently understood corresponds to a species complex (Pinho *et al.* 2008).

Three distinct morphotypes of this complex occur in the southeastern part of the Iberian Peninsula. The *liolepis* morphotype ('Type 3' in Pinho *et al.* 2006; 'northeastern form' in Geniez *et al.* 2007) is found in the province of Valencia along the Mediterranean coast north of Alicante, in addition to a large region of northern Spain and southern France. The 'type 2' morphotype (Guillaume 1987; Geniez 2001; Harris & Sá Sousa 2001, 2002) occurs in the central region of the peninsula, whereas the *hispanicus* morphotype (as redefined in Geniez *et al.* 2007) is limited to the Spanish Levant in Andalusia and in the province of Murcia. Distribution ranges of these three morphotypes are contiguous, and no locality is known to accommodate more than one morphotype.

In the area occupied by these three morphotypes, four distinct mitochondrial lineages have been identified (Harris & Sá Sousa 2002; Pinho *et al.* 2006). The '*P. hispanica* Type 2' lineage only occurs in the range of the 'type 2' morphotype, but two lineages occur in the range of the *hispanicus* morphotype: '*Podarcis hispanica* sensu stricto' (called 'Valencia lineage' hereafter, Table 1) and '*P. hispanica* Galera' (called 'Hispanicus lineage' hereafter). Similarly, two mitochondrial lineages, namely the '*P. hispanica* Type 3' (called 'Liolepis lineage' hereafter) and the Valencia mitochondrial lineages, have been identified within the range of the *liolepis* morphotype.

Nuclear patterns of differentiation among these mtDNA lineages have provided conflicting results:

Table 1 Correspondence between names employed in this study to describe mitochondrial and morphological units with the names provided in the previous studies (Pinho *et al.* 2006, 2007, 2008). Corresponding symbols used in figures are indicated between parentheses. 'Reference' indicates the samples retrieved from Pinho *et al.* (2006), and used for mitochondrial lineage identification

Present study		Previous studies	
Mitochondrial lineage	Morphotype	Lineage name	Reference
—	—	<i>Podarcis muralis</i>	<i>Gua1</i>
—	—	<i>Podarcis bocagei</i>	<i>BTA1</i>
—	—	<i>Podarcis hispanica</i> type 1A/type 1B	<i>Mon8/Trj1</i>
—	—	<i>Podarcis carbonelli</i>	<i>PR1</i>
Type 2 (open triangle)	'type 2' (filled triangle)	<i>Podarcis hispanica</i> type 2	<i>Phv4/Tie1</i>
Valencia (filled diamond)	—	<i>Podarcis hispanica sensu stricto</i>	<i>And8, Pod12, Mot1</i>
—	—	<i>Podarcis (hispanica) vaucheri</i>	<i>Cin1/Ouk7</i>
—	—	<i>Podarcis hispanica</i> Tunisia/Sirwah	<i>LK6/Js6</i>
Lioplepis (open square)	<i>lioplepis</i> (cross)	<i>Podarcis hispanica</i> type 3 or NE type	<i>Bur2</i>
Hispanicus (open circle)	<i>hispanicus</i> (filled circle)	<i>Podarcis hispanica</i> Galera	<i>Gal3</i>

based on ten allozyme loci, Pinho *et al.* (2007) revealed a lack of nuclear genetic differentiation between individuals bearing the Valencia and the Lioplepis mitochondrial lineages, but Pinho *et al.* (2008) found no gene flow between populations with Valencia haplotypes and populations with Lioplepis haplotypes using nuclear introns sequence variation. They also found significant gene flow between Valencia populations and populations with Hispanicus haplotypes.

It is thus still unclear whether the limits of taxonomic units based on mitochondrial markers are discordant with those based on nuclear markers. Besides, the correspondence between evolutionary units delimited by nuclear markers and those delimited by morphological data has never been investigated in this region. Currently, all that we have is an apparent discordance between mitochondria- and morphology-based taxonomy as suggested by comparison of published maps produced by separate teams working on different specimens.

The aims of this study were thus to investigate rigorously possible discrepancies between markers in the southeastern Iberian lineages of the *P. hispanicus* complex and to elucidate their origin. We first studied the concordance between delimitations of taxa based on morphological and mitochondrial data. This revealed several cases of discordance between these two kinds of markers. Next, we assessed the origin of these discordances with nuclear DNA (nDNA): if the geographical patterns of variation in nuclear and mitochondrial DNA were concordant, then the different morphotypes would correspond to instances of local adaptation or phenotypic plasticity rather than to true evolutionary lineages. On the contrary, if the distribution of mtDNA variation did not match with morphotypes and nuclear genotypes, introgressive

hybridization or incomplete lineage sorting of mtDNA must have been involved.

Material and methods

Sample origin and DNA extraction

Specimens used for morphological analyses originated from various European collections (see Table S1, Supporting information). Specimens used for molecular analyses all originated from the 'Biogéographie et Ecologie des Vertébrés' collection (Montpellier, France). Sampling locations of the analysed specimens are shown in Figure 1 (see also Table S2, Supporting information). We collected all the genetic data presented here except for the sequences of the mitochondrial control region of 16 individuals (see Table 1), which were published previously (Pinho *et al.* 2006).

Total genomic DNA was extracted from ~40 mg of muscle using the Qiagen DNeasy[®] Tissue kit, following the manufacturer's protocol. For several poorly preserved samples, we performed the final elution with only 100 µL buffer (half the recommended volume) to concentrate the DNA. The resulting DNA was visually quantified after migration on agarose gels and diluted for use in PCR reactions.

Morphological analyses

Twenty-four quantitative and semi-quantitative variables describing size, pigmentation and pholidosis were measured (by PG; see Table S3, Supporting information) on a total of 607 lizards. Adult individuals of both sexes were pooled because they share the main traits that separate morphotypes (results not shown). The lizards originated from a total of 70 localities represented by at

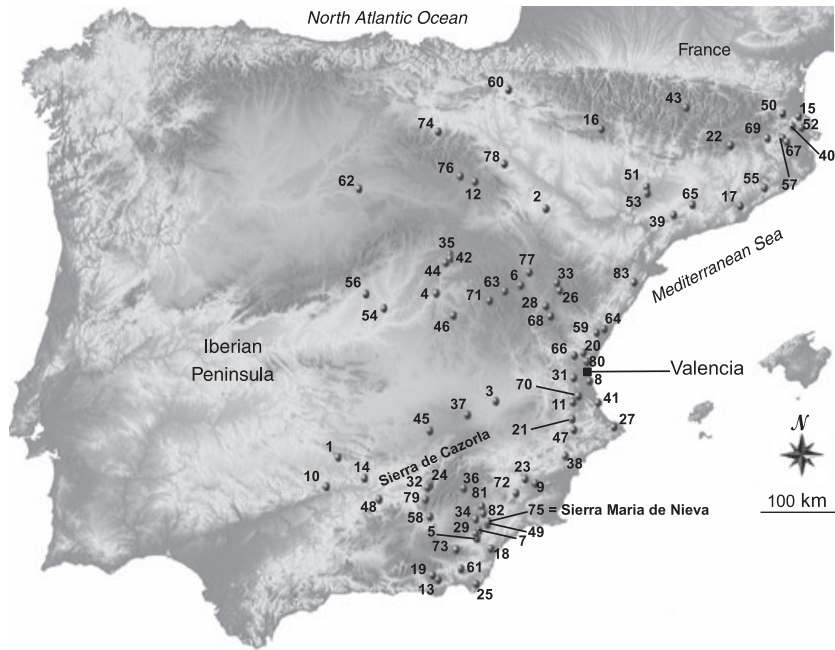


Fig. 1 The sampling sites. Map illustrating the sampling sites for morphological, mitochondrial and nuclear analyses (see also Table S2), as well as the main locations cited in this study.

least two individuals (other localities were discarded). First, a Discriminant Analysis was performed using the STATISTICA 6.1 software (StatSoft), including all of the measured specimens, to better define the limits between distribution ranges of morphotype. We defined a priori three groups of reference specimens (one per morphotype) using individuals located far from the suggested contact zones (approximately located in Geniez 2001) and that PG visually determined to be characteristic of each morphotype (localities outside the dashed square in Fig. 2). These reference specimens were used to construct two classification functions that classified individuals into the three morphological groups. Each locality was classified based on the average Mahalanobis distance from each group centroid (mean of the distances of all individuals in this locality). Each locality was attributed to the morphotype corresponding to the smallest average Mahalanobis distance. This method assumed that each locality did not accommodate more than one morphotype, in accordance with what is currently known in these lizards (see above) and with our visual examination of field or museum specimens. Finally, each lizard sample that needed to be identified (localities inside the dashed square in Fig. 2) was introduced as a supplementary item into the Discriminant Analysis.

The technique described above is suited to locate contact zones among morphotypes. However, it cannot identify more morphological groups than were defined a priori. To check whether the additional mitochondrial clade present in the study area (see results) corresponded to an overlooked fourth morphotype, we

performed a Principal Component Analysis (PCA) using the same set of variables. This analysis included 298 *iolepis* and 190 *hispanicus* individuals located in an 80 km-wide and 800 km-long band running from Almería (southern Andalusia, Spain) to Banyuls (south of France). Each locality was projected on a line running through the middle of the band to obtain a distance that was along the transect. This distance was then divided into 50 km-length sections and the first two individual principal component (PC) scores were averaged for each section. Mean PC scores were plotted against distance. We eventually compared averaged PC1 and PC2 scores between groups based on mitochondrial clade membership, and between groups based on the morphological classifications resulting from the Discriminant Analysis.

Mitochondrial DNA analyses

A 460 bp fragment of the mtDNA control region (CR) was amplified in a total of 69 specimens using the DL3F and DLA primers (Crochet *et al.* 2004) and a protocol adjusted from Pinho *et al.* (2006). Sequencing reactions were conducted using ABI Dye Terminator kits (Applied Biosystems) following the manufacturer's recommendations. Sequencing samples were electrophoresed on an ABI Prism 310 Genetic Analyzer using DL3F as the sequencing primer. Table S4 (Supporting information) provides GenBank accession numbers of the 69 sequences.

Sequences were manually compiled and aligned using MEGA v.3.1 (Kumar *et al.* 2004). A phylogenetic tree was reconstructed with 87 sequences, including 16

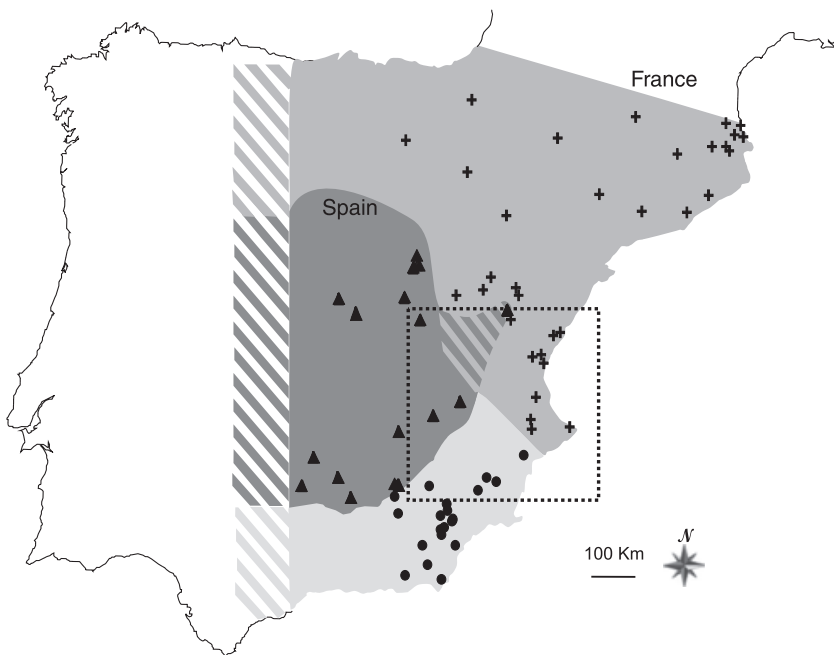


Fig. 2 Geographic distribution of the three morphotypes. Populations were assigned to a morphotype using a Discriminant Analysis. 'type 2' = filled triangles, *liolepis* = crosses and *hispanicus* = filled circles. Locations outside the dotted square were used as reference samples to compute the classification function (see text). Locations inside the dotted square were included in the analysis as supplementary items. Grey shadings are our interpretations of distributions of morphotypes derived from these results and are only used for illustrative purposes in Figs 2, 3 and 4. Grey hatching around locality 28 illustrates uncertainties about the distribution of 'type 2' in this area (see results section). Hatching on the left corresponds to the lack of precise information on distribution of morphotypes outside the study area.

sequences from Pinho *et al.* (2006), which allows us to represent the main clades known in the *Podarcis* genus and occurring in Iberia and North Africa (Table 1). Thus, we could assign each of our individuals to one of the mitochondrial lineages previously described in the *P. hispanicus* species complex. The tree was constructed using the maximum likelihood approach (ML, Felsenstein 1981). We used the heuristic search algorithm PHYL 2.4.4 (Guindon & Gascuel 2003) to estimate the ML tree. We used the ModelTest 3.06 software suite (Posada & Crandall 1998) and hierarchical likelihood ratio tests to select an appropriate model of sequence evolution. The best-fit model was the HKY, with an estimate of invariable sites (0.75) and a discrete approximation of the gamma distribution ($\alpha = 1.27$).

The nucleotide diversity (π), $\theta(s)$ computed from the number of segregating sites and haplotype diversity (h) were determined using ARLEQUIN 3.0 (Excoffier *et al.* 2005). We estimated the divergence time between Valencia haplotypes present in the *hispanicus* and in the *liolepis* morphotypes (see results) with a modified version of the coalescent-based program MDIV (Nielsen & Wakeley 2001). This program simultaneously estimates several parameters including theta ($\theta = 4 N\mu$), migration rate ($M = 2 Nm$) and the relative time of population divergence ($T = t/2N$) from the same data set. In this analysis, N is the gene effective population size (for mtDNA, N is equal to the effective population size of the female fraction of the population), t is the divergence time in years, and μ is the mutation rate per sequence per year. We used a modified version of MDIV (distributed by R. Nielsen) that provides the posterior

distribution of θ^*T , allowing the calculation of the 90% credible interval (CI90) for this parameter. The simulations were conducted in triplicate, with runs of 5×10^6 generations and a 10% burn-in. An infinite-sites model with no migrant per generation was set to accommodate for the lack of shared haplotypes between the *hispanicus* and *liolepis* populations. The evolutionary rate of the mtDNA control region in reptiles is not well known. The CR fragment we used evolves at the same rate as cytochrome *b* in a closely related lizard genus (Crochet *et al.* 2004), but in the genus *Podarcis* it could evolve at a slower rate (C. Pinho, personal communication). On the other hand, the estimated population divergence time is almost certainly inflated because the populations we are studying violate the MDIV assumption of equal population sizes (Chapman *et al.* 2007). We thus computed an interval of T using two evolutionary rates, which probably encompass the real rate based on previously published estimates in reptiles (Crochet *et al.* 2004). This level is approximately 1% and 2% sequence divergence per million years.

Nuclear DNA analyses

Beta-fibrinogen intron 7 amplification and analysis. Pinho *et al.* (2008) amplified the β -*fibint7* intron in different *Podarcis* mitochondrial lineages. They observed an easy to detect length polymorphism between various lineages as a result of a ~ 350 bp insertion. The length varies thus between ~ 600 bp and ~ 950 bp. For these reasons, we amplified DNA from 57 individuals using primers BF8 and Bfib (Pinho *et al.* 2008) and the PCR

conditions provided by these authors. The length polymorphism was evaluated by visualizing PCR products with UV light after electrophoresis on a 2% agarose gel.

Microsatellite amplification. A total of 104 individual samples were successfully genotyped for eight (Pb10, Pb11, Pb20, Pb37, Pb47, Pb50, Pb55 and Pb73) of the nine polymorphic microsatellite loci developed by Pinho *et al.* (2004). For multiplexing convenience, we excluded Pb66 from our analysis. Conditions of amplification are provided in Data S5 (Supporting information).

Delimitation of nuclear genetic units. The genetic structure of our samples of Iberian wall lizards was examined using the software GENELAND (Guillot *et al.* 2005). GENELAND is a spatially explicit clustering method based on a Bayesian model which assigns individuals probabilistically to clusters without a priori knowledge of population units and limits. It has been demonstrated to perform better than other spatially nonexplicit Bayesian clustering methods when genetic differentiation is weak (Fontaine *et al.* 2007) and when Wahlund effects may operate in one cluster (Coulon *et al.* 2006). In our case, samples of each lineage were collected over a wide geographic range and thus are expected to show gene-flow restrictions among them, at least because of isolation by distance, resulting in Wahlund effects within lineages. In addition, given the age of the different lineages in this complex (Pinho *et al.* 2006) and the mutation rates of microsatellites, we expected investigated taxa to be weakly differentiated because of homoplasy. Thus, the GENELAND software appears particularly appropriate for the study of the *Podarcis hispanicus* species complex. We used both microsatellites and nuclear-intron data together, with all individuals. We coded the intron as missing data for the samples in which it was not amplified. The maximum of convergence across runs was obtained from a minimum of 1.5×10^6 iterations. We performed a first series of 30 runs, allowing K to vary from 1 to 6 and using the following run parameters: 1.5×10^6 MCMC iterations sampled each 300 iterations, maximum rate of Poisson process fixed to 100, maximum number of nuclei in the Poisson-Voronoi tessellation fixed to 300, a full certainty associated with the spatial coordinates and the Dirichlet model of allelic frequencies. Moreover, we activated the newly included option that allows the algorithm to deal with null alleles (Guillot *et al.* 2008). We inferred the number of clusters from the modal value of K for these 30 runs, and then ran the MCMC 30 additional times with K fixed to this value, with the other parameters unchanged. We calculated the mean logarithm of posterior probability for

each of the 30 runs. The five runs with the lowest value were discarded to avoid spurious clustering resulting from a MCMC scheme with no real statistical significance, which is always possible with Bayesian methods. The posterior probability of population membership for each pixel of the spatial domain was then computed for each of these 25 remaining runs, using a burn-in of 1.5×10^5 iterations and 125 pixels along the X axis and 95 along the Y axis. We then visually inspected the consistency of the results across the 25 runs.

Descriptive statistics within and among clusters. For microsatellite data, the mean number of alleles per locus (N_a) was calculated as well as Nei's unbiased expected heterozygosity (H_E ; Nei 1978), observed heterozygosity (H_O) and Wright's F_{IS} statistic according to Weir & Cockerham (1984). The tests for departure from the Hardy-Weinberg equilibrium and for linkage disequilibrium between loci were conducted using 1000 permutations in each population. Differentiation indexes (F_{ST}) were calculated in accordance with the Weir & Cockerham (1984) procedure and the significance tested using 1000 permutations of individuals among groups. All calculations and tests were performed with GENETIX v.4.05 (Belkhir *et al.* 2001).

Results

Morphology-based clustering

All of the 51 reference localities were correctly classified using the Discriminant Analysis. Hence, we accepted the reference assignation and the classification functions as valid. One hundred and seven individuals originating from 19 localities within the contact zone (Fig. 2, inside the dashed square) were included as supplementary individuals in the classification functions to be assigned to one of the three morphotypes (*hispanicus*, *liolepis* and 'type 2').

The spatial distributions of the three morphotypes, inferred from the distribution of the classification scores averaged per locality, were nonoverlapping (Fig. 2). The *hispanicus* morphotype was restricted to the southern part of the study area, type 2 morphotype to the western part and the *liolepis* morphotype to the northern part. Individuals from Location 28, assigned on morphological grounds to 'type 2', were physically located far from the other 'type 2' populations and very close to locations attributed to *liolepis*. Because of the limited sample size for this location ($n = 2$), it is difficult to determine if the 'type 2' morphotype actually extends to this locality, or whether it should be considered as a misclassified sample.

In the *hispanicus*–*liolepis* contact zone, all the specimens belonging to the northernmost *hispanicus* location ($n = 6$) were individually classified as *hispanicus* (results not shown). In the two southernmost *liolepis* locations (21 and 27) located just north of the *hispanicus*–*liolepis* limit, 5 lizards of 21 were individually classified as *hispanicus*, although they have quite similar *hispanicus* and *liolepis* scores. Globally, in the southernmost *liolepis* samples, many individuals were separated by almost identical Mahalanobis distances from *hispanicus* and *liolepis* centroids. These results indicate that many *liolepis* specimens from the southernmost populations (close to the range limit of *hispanicus*) are morphologically close to *hispanicus*, while *hispanicus* specimens from the north of the distribution (close to *liolepis* range) do not differ in morphology from other *hispanicus* specimens. In addition, *hispanicus* populations that presented the Valencia mitochondrial haplotype (Vc) were unambiguously attributed to the *hispanicus* morphotype (see below, Fig. 3). Finally, populations 21, 27 and 47 were plotted at about equal distance from the *liolepis* centroid and the 'type 2' centroid (see Table S2 for details).

Mitochondrial sequence diversity and clustering

Our 69 specimens of *Podarcis hispanicus* from the southeastern Iberian Peninsula fell into four highly divergent mitochondrial lineages. Based on our results, we named these haplogroups the Hispanicus, Type 2, Liolepis and Valencia clades (Table 1). The overall sequence diversity was high ($h = 0.851 \pm 0.028$; $\pi = 0.018 \pm 1.1$; Table 2). Each of the 39 locations (including samples from Pinho *et al.* 2006) sampled was found to contain haplotypes of only one haplogroup. The Hispanicus, Type 2 and Liolepis haplotype groups present separate, nonoverlapping geographic distributions in the studied zone (Fig. 3a). Haplotype diversity and nucleotide diversity for these three groups varied from moderately low values in Type 2 ($h = 0.298 \pm 0.133$; $\pi = 0.9 \pm 0.7$) to very high in Liolepis ($h = 0.864 \pm 0.079$; $\pi = 3.5 \pm 1.9$). By contrast, the Valencia clade is found in several areas. This clade constitutes the only haplotype group found within a large part of southeastern Spain around the town of Valencia, while it is also found in a population (in the Sierra Maria de Nieva) surrounded by localities where we found the Hispanicus clade. In the main distribution range of the Valencia clade, only two different haplotypes were observed. One was present in most populations and individuals (23 individuals, haplotype Va in Fig. 3a, i.e. the same haplotype as Mot1 in Pinho *et al.* 2006) and the other one was only found in the two individuals from population 59 (Vb). Specimens presenting a haplotype belonging to the Valencia group sampled inside the Hispanicus range (the Sierra Maria

de Nieva population) all shared a unique haplotype (Vc) distinct from both Va and Vb haplotypes by two substitutions. Haplotype and sequence diversity of the Valencia group were the lowest of all mitochondrial groups ($h = 0.458 \pm 0.094$; $\pi = 0.8 \pm 0.5$) and these values were even lower when only Valencia individuals having a *liolepis* morphotype were analysed ($h = 0.153 \pm 0.092$; $\pi = 0.2 \pm 0.2$).

The MDIV program was used to estimate the divergence time between the Va and Vc haplotypes. Maximum posterior probabilities were reached for values of $\theta = 0.18$ (CI90: 0.09–0.51), $T = 5.9$ (2.86–9.3) and $\theta^*T = 1.06$ (0.5–3), corresponding to a divergence time between the Va and Vc haplotypes of 120 000–240 000 years based on an estimated rate of sequence divergence of 2% and 1% per million years respectively, for the 431 bp sequence.

Morphological variation within and between morphotypes

The first two components of the PCA explained 44% of the total variance (Table 3). We found a significant morphological differentiation along PC1 ($F_{(1; 263)} = 4.39$; $P < 10^{-3}$) but not PC2 ($F_{(1; 263)} = 0.27$; $P = 0.79$) between *liolepis* individuals from the Valencia clade ($m_{PC1} = -0.416$ and $m_{PC2} = 0.323$) and from the Liolepis clade ($m_{PC1} = -0.815$ and $m_{PC2} = 0.356$). However, the plot with the mean PC scores (m_{PC1} and m_{PC2}) per 50 km section (Fig. 3b) illustrates that the southernmost section of the distribution range of the Valencia clade contains individuals presenting a divergent morphology. And indeed, the differentiation found on PC1 vanished ($F_{(1; 237)} = 1.91$; $P = 0.08$) when the individuals from the southernmost section were removed from the analysis (Valencia: $m_{PC1} = -0.61$).

Within the *hispanicus* morphotype, samples presenting the Valencia haplotype (Sierra Maria de Nieva population: $m_{PC1} = 0.75$ and $m_{PC2} = -0.33$) were not differentiated morphologically (PC1: $F_{(1; 61)} = -0.86$; $P = 0.38$ and PC2: $F_{(1; 61)} = 1.51$; $P = 0.14$) from the *hispanicus* specimens presenting the typical Hispanicus haplotypes ($m_{PC1} = 0.86$ and $m_{PC2} = -0.66$). However, they were clearly differentiated (PC1: $F_{(1; 43)} = 4.98$; $P < 10^{-3}$) from the *liolepis* morphotype/Valencia clade from the closest geographic section ($m_{PC1} = -0.01$).

Within the *liolepis*–*hispanicus* contact zone, a gap in the score distribution of PC1 was visible along the transect, and corresponded to the morphotype limits shown by the Discriminant Analysis. All size variables were strongly correlated with PC1 (Table 3). The gap between PC1 scores across the contact zone is therefore mainly due to a discontinuous transition from the small *hispanicus* to the large *liolepis* morphotype. By

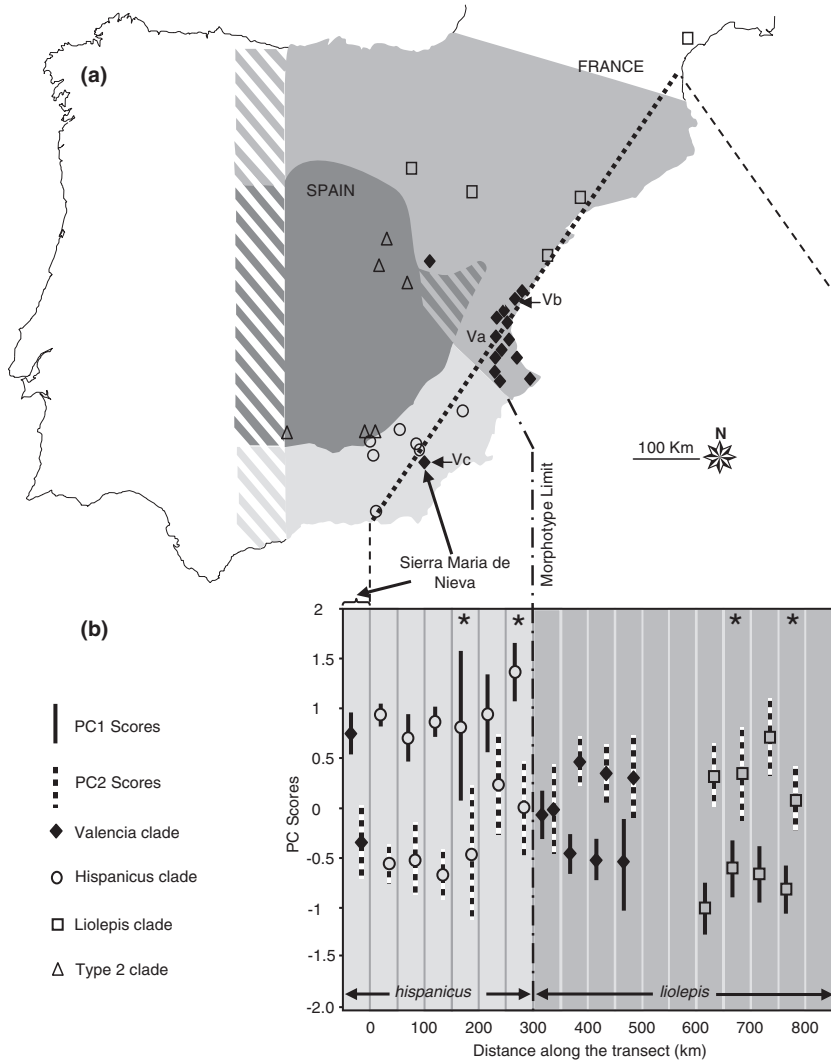


Fig. 3 Distribution of mitochondrial clades compared with the distribution of morphotypes. (a) The symbols indicate sampled locations for which a fragment of the mitochondrial Control Region was sequenced for one or several individuals. Open triangle = Type 2 mitochondrial clade, filled diamond = Valencia mitochondrial clade, open square = Liolepis mitochondrial clade, open circle = Hispanicus mitochondrial clade. No more than one mitochondrial clade was found in each location. Va, Vb and Vc indicate the distribution of the three different haplotypes occurring in the Valencia mitochondrial clade. The distribution of mitochondrial clades has to be compared with the distribution of morphotypes indicated by shading: light, medium and dark grey represent the distributions of *hispanicus*, *liolepis* and 'type 2' morphotypes respectively. Method used to construct shading is detailed in the legend of Fig. 2. The dotted line crossing Spain represents the transect along which morphological variation in *liolepis* and *hispanicus* is represented. (b) Geographical variation in PC scores extracted from a PCA performed with 24 morphological variables. Each column represents a 50-km-length section along the transect. Mean PC1 (plain line) and PC2 (dotted line) scores with 95% confidence intervals are given for each section. The symbols indicate the means and correspond to the mitochondrial clade occurring in a given section. Asterisks indicate areas where mitochondrial clade membership is inferred from haplotype distribution but not confirmed by specimen sequencing. The location 'Sierra Maria de Nieva' is represented separately, in the leftmost column. The limit between *liolepis* (medium shading) and *hispanicus* (light shading) morphotypes has been placed according to the Discriminant Analysis (Fig. 2).

contrast, PC2 scores show a gradual transition from one morphotype to another. Variables strongly correlated with PC2 concern the intensity of the ventral black pigmentation. The northernmost *liolepis* are thus less pigmented ventrally than the southernmost *hispanicus*.

Clustering based on nuclear markers

Beta-fibrinogen intron 7. A diagnostic allele presenting a 300 bp insert was observed at a homozygous state in all individuals classified morphologically as *hispanicus* by the Discriminant Analysis ($n = 18$), including the six

Table 2 Estimates of sequence diversity in southeastern Iberian mitochondrial lineages of the *Podarcis hispanicus* species complex. Hisp, Hispanicus; Lio, Lioplepis; Val, Valencia; Val(*P. lioplepis*), Valencia mtDNA found in *P. lioplepis*. n_s , number of sampled locations including locations retrieved from Pinho *et al.* (2006); n_i , number of analysed individuals; n_h , number of observed mtDNA haplotypes; π , haplotype diversity; $\theta(s)$, computed from the number of segregating sites. Standard deviations are shown in brackets

Group	n_s	n_i	n_h	h	π (%)	$\theta(s)$ per site (%)
Hisp, Type 2, Lio,Val	39	78	21	0.851 (0.028)	1.8 (1.1)	2.0 (1.0)
Hisp	8	12	5	0.758 (0.093)	2.1 (1.3)	2.0 (1.1)
Type 2	7	19	5	0.298 (0.133)	0.9 (0.7)	1.7 (0.9)
Lio	6	13	7	0.864 (0.079)	3.5 (1.9)	3.6 (1.7)
Val	18	34	4	0.458 (0.094)	0.8 (0.5)	0.7 (0.5)
Val(<i>P. lioplepis</i>)	15	26	2	0.153 (0.092)	0.2 (0.2)	0.3 (0.3)

individuals from the Sierra Maria de Nieva carrying the Valencia mitochondrial lineage. This insertion was not found anywhere else, except for the two individuals (from locations 21 and 27, Table S2) sampled within the *lioplepis* range, less than 20 km north of the *hispanicus*–*lioplepis* limit. These two individuals were heterozygous, presenting a copy of both long and short alleles.

Microsatellite variability and differentiation between morphological and mitochondrial groups. Allele numbers varied from 11 for pb11 in 'type 2' to 37 for pb55 in *lioplepis* (Table 4). No linkage disequilibrium was detected between our microsatellite loci (Table S6, Supporting information). In both types of grouping (morphology vs. mitochondria-based), each group had an overall observed heterozygosity (H_O) lower than the expected heterozygosity (H_E), and presented a significant deficit in heterozygotes (Table 4). Pairwise F_{ST} estimates showed significant differentiation between morphotypes, even between the *lioplepis* and the 'type 2' morphotypes that are the least differentiated. The highest F_{ST} values were observed between groups belonging to the same mitochondrial clade, but to different morphological groups (Valencia clade/*hispanicus* morphotype vs. Valencia clade/*lioplepis* morphotype; $F_{ST} = 5.2\%$; $P < 10^{-2}$). On the contrary, specimens presenting the same morphotype but different mtDNA lineages were not significantly differentiated (Valencia clade/*hispanicus* morphotype vs. Hispanicus clade/*hispanicus* morphotype; $F_{ST} = 0.003$; $P < 10^{-2}$) or they were poorly differentiated (Valencia clade/*lioplepis* morphotype vs. *Lioplepis* clade/*lioplepis* morphotype; $F_{ST} = 0.019$; $P < 5.10^{-2}$).

Delimitation of nuclear units using the Bayesian clustering method. Posterior distributions of the estimated number of populations (K) across the 30 runs presented a clear mode at $K = 3$ for each of the 30 runs (Fig. 4d). Twenty-four of the 25 selected runs were consistent in placing all of the *hispanicus* specimens (whatever their

mitochondrial genome) into a single cluster. Indeed, individuals from the Sierra Maria de Nieva, presenting a Valencia mitochondrial haplotype, clustered with other *hispanicus* specimens, presenting a Hispanicus mitochondrial lineage. The remaining individuals were classified into two clusters following somewhat different patterns across these 24 runs. In 11 runs, individuals from the Valencia and the *Lioplepis* clades were classified into the same cluster (Fig. 4a–c), except for

Table 3 Factor coordinates of 24 morphological variables used in the principal component analysis. * and ** indicate moderate (0.5–0.7) and strong (>0.7) correlation between the variable and the factor respectively

Variables		PC1	PC2
Scalation	Dors	−0.42	−0.23
	VenL	0.16	0.41
	Guls	0.03	−0.40
	Fpor	−0.47	−0.15
	Lame	−0.12	−0.33
	Temp	0.24	−0.46
	TeMa	−0.68*	0.33
Size	Mass	−0.74**	0.41
	ID10	−0.75**	0.34
	SVL	−0.80**	−0.10
	HeNe	−0.80**	−0.39
	PilL	−0.84**	−0.39
Pigmentation	PilW	−0.89**	−0.26
	HeH	−0.87**	−0.29
	Vert	0.15	−0.43
	Bif	0.62*	−0.33
	DoLa	0.43	−0.20
	Frag	−0.07	0.07
	SDLa	0.05	−0.49
Eigenvalue % Variance	Pari	0.69*	−0.47
	Ponc	−0.12	−0.71**
	Rond	−0.02	−0.65*
	Tria	−0.19	−0.10
	PilP	−0.00	−0.67*
		6.72	3.77
		29	16

Table 4 Genetic diversity and pairwise differentiations between clusters defined according to morphological and/or mitochondrial analyses. Mean number of overall alleles (N_a), Nei's diversity (H_E), observed heterozygosity (H_O) and fixation indices (F_{IS}) of Weir & Cockerham (1984) are given by the software GENETIX. F_{IS} values refers to Hardy–Weinberg equilibrium; F_{ST} values indicate genetic homogeneity between populations. * $P < 0.05$; ** $P < 0.01$

No.	Group	Sample size	N_a	H_E	H_O	F_{IS}	Pairwise F_{ST}			
							No. 1	2	4	6
1	<i>hispanicus</i> morphotype	33	20.63	0.88	0.75	0.17**	—	—	—	—
2	<i>liolepis</i> morphotype	48	27.50	0.94	0.70	0.26**	0.040**	—	—	—
3	'type 2' morphotype	23	16.00	0.90	0.71	0.24**	0.048**	0.022**	—	—
4	Hispanicus clade/ <i>hispanicus</i> morphotype	23	17.75	0.88	0.77	0.14**	—	—	—	—
5	Valencia clade/ <i>hispanicus</i> morphotype	10	10.25	0.83	0.70	0.21**	—	—	0.003	0.052**
6	Valencia clade/ <i>liolepis</i> morphotype	27	20.12	0.91	0.71	0.24**	0.050**	—	—	—
7	Lirolepis clade/ <i>liolepis</i> morphotype	21	18.37	0.92	0.68	0.28**	0.040**	—	—	0.019*

the two individuals from location 27, which were systematically attributed to the other cluster with Type 2 clade specimens. In six runs, GENELAND assigned all remaining individuals to a single cluster and left the remaining cluster without any assigned individual. In three runs, the Valencia clade specimens clustered with all the Type 2 specimens and with individuals of the southernmost Lirolepis localities, except for the westernmost Valencia specimens that clustered with the rest of the Lirolepis specimens. A similar pattern was found in two runs except that southern Type 2 specimens clustered with most Valencia specimens and northern Type 2 specimens clustered with most Lirolepis specimens. Two runs clustered Valencia specimens with a portion of the Type 2 and all the Lirolepis specimens. Finally, a 25th run showed an aberrant clustering with no coherence with regard to either mitochondrial or morphological clustering. Most importantly, none of the 25 runs placed individuals of the Valencia clade into their own cluster. The occurrence of a 'ghost' cluster, that is to say without individuals assigned to it (because centred on the Mediterranean Sea), has already been reported by the authors of GENELAND (Guillot *et al.* 2005) and in other studies (Coulon *et al.* 2006; Fontaine *et al.* 2007) and remains a poorly understood problem.

Discussion

Mitochondria-based taxonomy revisited: cyto-nuclear discordances

In accordance with previous results (e.g. Pinho *et al.* 2006), we identified four mitochondrial lineages in south-eastern Iberia, the amounts of sequence divergence among which are typical of interspecific comparisons. As we suspected, this is incongruent with our identification of only three morphotypes in the same area, as found by Geniez (2001). The distributions of

the Hispanicus, Lirolepis and Type 2 mtDNA lineages closely match morphotype distributions, as they have been sampled only in populations assigned to the *hispanicus*, *liolepis* and 'type 2' morphotypes respectively. On the contrary, the Valencia clade was found in populations morphologically assigned to *liolepis* (with haplotype Va and Vb) and *hispanicus* (with haplotype Vc). It is thus the occurrence of this additional Valencia lineage in two of the three morphotypes that creates discordance between morphology and mitochondria. Moreover, GENELAND identified three genetic clusters in the study area that generally correspond to the morphological clusters (but see Discussion below). The occurrence of three nuclear clusters, concordant with morphotypes but not with the four mitochondrial lineages identified in the study area, implies that there is extensive cyto-nuclear discordance in these populations.

The first instance of discordance is the occurrence of the Valencia haplotype Vc in individuals from the Sierra Maria de Nieva population. GENELAND never associated these individuals with the other Valencia individuals but always with the specimens of the Hispanicus clade. This result is unambiguously supported by the length polymorphism of the beta-fibrinogen intron – the Sierra Maria de Nieva population shares the same alleles as the Hispanicus clade populations – as well as by the absence of significant differentiation at microsatellite loci between individuals of the Valencia and Hispanicus mtDNA clades.

The second case is the Valencia individuals presenting haplotypes Va and Vb that do not cluster into a separate group in any of the GENELAND runs. However, it should be noted that not every GENELAND run converged towards the same results (see below for a complete discussion of this issue). However, despite these evidences of occasional flaws in the nuclear clustering procedure, the fact that the most frequently obtained

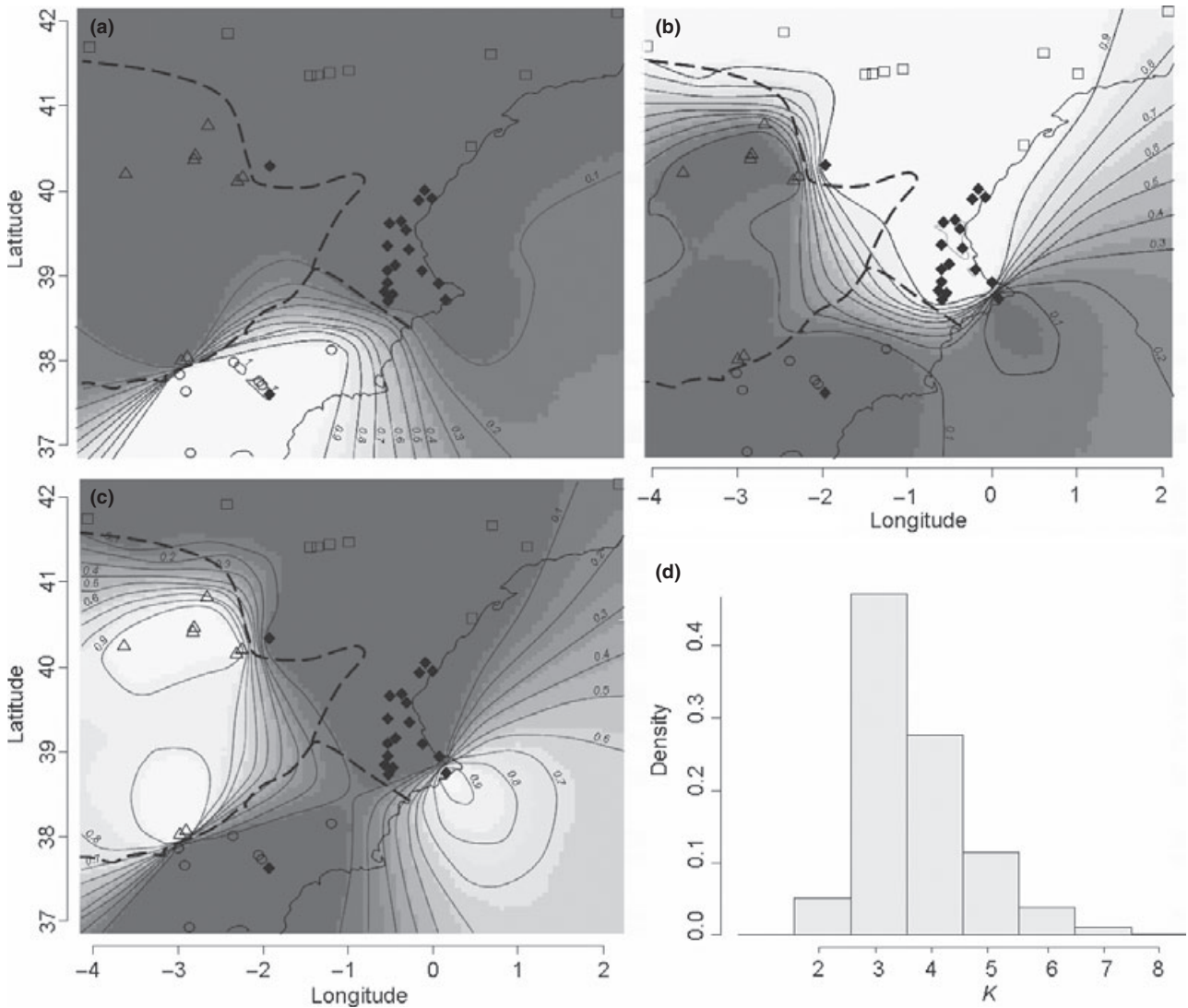


Fig. 4 Nuclear clustering performed with the GENELAND Bayesian algorithm and comparison with mitochondrial and morphological assignment. Plots (a) to (c) represent the assignment of pixels to each cluster, based on the clustering pattern showing the highest modal value among the 25 selected runs: assignment to (a) *hispanicus*; (b) *liolepis*; (c) 'type 2'. Highest membership values correspond to the palest shade and contour lines illustrate the spatial change in assignment values. Symbols for each locality correspond to their mitochondrial clade (open triangle, filled diamond, open square and open circle refer to Type 2, Valencia, Liolepis and Hispanicus mitochondrial clade respectively). Bold dashed lines represent the limits between morphological clusters according to Fig. 2. The scales indicate the geo-referenced coordinates in decimal units of degrees (WGS 84 projection system). (d). Posterior density distribution of the number of clusters along the MCMC chain after burn-in as estimated from the 30 replicate runs.

clustering pattern was congruent with morphological clustering strongly suggests that this general pattern reflects real genetic clustering.

Taken together, morphological and nuclear results thus suggest the occurrence of only three evolutionary units in the southeastern part of the Iberian Peninsula, corresponding to the previously recognized morphotypes. The three congruent morphological, mitochondrial and nuclear clusters correspond to three species: *Podarcis hispanicus*, *P. liolepis* and *P. sp.* 'type 2'. On the contrary, the occurrence of two highly divergent

mtDNA lineages in two of these species (*Hispanicus* and *Valencia* mitochondrial clades in *P. hispanicus*; *Lirolepis* and *Valencia* clades in *P. liolepis*) constitutes two clear cases of cyto-nuclear discordance.

The conclusions drawn here seem to contradict the results of Pinho *et al.* (2008), as they found virtually no nuclear gene flow between the *Valencia* mitochondrial clade and the *Lirolepis* mitochondrial clade. However, the specimens of the *Valencia* clade they used originated from four distinct populations: two specimens belong to the *P. liolepis/Valencia* mitochondrial clade

populations, four specimens originate from the Sierra Nevada in southern Spain (an area isolated from *P. liolepis* populations by the distribution range of *P. hispanicus*) and belonged to an unknown taxon, one specimen originated from the Sierra Maria de Nieva (this specimen was also analysed in our study) and belonged to *P. hispanicus*, and two specimens originated from the Sierra de Cazorla (located around the *P. hispanicus*-*P. sp.* 'type 2' contact zone) and were also impossible to assign to any taxon. The estimate of nuclear gene flow between the Valencia and *Liolepis* lineages proposed by Pinho *et al.* was therefore based on the use of at least two taxa (*P. hispanicus* and *P. liolepis*) to represent the Valencia lineage. These gene flow estimates are thus difficult to compare with our findings of limited nuclear differentiation within *P. liolepis*.

Inconsistent nuclear clusters and variation within Podarcis liolepis

Our interpretation of the occurrence of the Valencia mitochondrial lineage in southern populations of *liolepis* as a case of cyto-nuclear discordance relies on our interpretation of the origin of the occasionally inconsistent patterns of nuclear clustering, and the weak but significant genetic and morphological differentiation observed between the northern and southern *liolepis* populations.

First, the difficulties of the software to consistently recover the same clusters may be caused by shared alleles because of incomplete lineage sorting or introgression, as evidenced by Pinho *et al.* (2008), by a high level of homoplasy in hypervariable microsatellite markers, or by an underlying genetic structure within *liolepis* and 'type 2' morphotypes (see below). A limited resolution in the GENELAND clustering procedure is indeed suggested by the systematic nuclear assignment of location 27 to *P. sp.* 'type 2'. Morphology, geography and distribution of mitochondrial lineages all indicate that these individuals are *P. liolepis* with a Valencia lineage. Nuclear intron evidences past hybridization with *P. hispanicus*, which is not unexpected given its location. However, genetic influences from *P. sp.* 'type 2' is highly unlikely as location 27 lies far from the 'type 2' – *liolepis* contact zone. Such assignment could therefore be caused by homoplasy or by incomplete lineage sorting.

Second, genetic variation within *liolepis* is likely to explain results of GENELAND runs where Va and Vb haplotypes did not cluster with all *Liolepis* haplotypes but where the southernmost location of *Liolepis* clustered with Valencia and the northernmost and westernmost Valencia location clustered with the rest of the *Liolepis* locations. This explanation is supported by the stronger values of heterozygote deficiency within the *liolepis* (and 'type 2') morphotypes than within the *hispanicus*

morphotype. This suggests that genetic structuring among populations is more pronounced in the first two species than in the last one, in accordance with their wider distribution range. Such spatial genetic structure is expected to generate genetic differentiation between populations of *liolepis* presenting the Valencia haplotype and the *Liolepis* haplotype because they occupy different parts of the *liolepis* range.

In addition, the change in morphology of *liolepis* along the transect does not correspond to the change in mtDNA lineages: northern specimens carrying Valencia haplotypes are morphologically identical to typical *liolepis* and not to southern Valencia specimens. This morphological variation is thus better explained by introgression of morphological characters from *hispanicus* or by morphological convergence with this taxon as a result of adaptation to similar environments. Several arguments favour the introgression hypothesis. First, the presence of two lizards heterozygous for the beta-fibrinogen intron in the southern part of the *liolepis* range evidences past nuclear introgression between *liolepis* and *hispanicus* in their contact zone. Second, past or present gene flow may explain the occasional occurrence of typical *hispanicus* characters such as a split dark vertebral line, white belly, bluish tail in juveniles, lack of a distinct masseteric plate and shape of temporal scales in southernmost *liolepis* (see Geniez *et al.* 2007). Third, the contrasting pattern of the variation between PC1 and PC2 scores across the contact zone is also in accordance with introgression between *liolepis* and *hispanicus*. The gradual variation of ventral black punctuations revealed by PC2 is hardly explicable as an adaptation to a gradually varying environment while the size variation revealed by PC1 could be linked to climate variation, as small size is considered to be an adaptation to dry and hot habitats in lizards (Horton 1972). It should be noted that the location of the *liolepis*-*hispanicus* contact zone matches an abrupt transition in mean annual precipitation and temperature (Pérez Cueva 1994).

Origin of the discordances: introgressive hybridization of mitochondrial DNA

The presence of the Valencia mitochondrial lineage in *P. liolepis* and *P. hispanicus* may be due to either incomplete lineage sorting of ancestral polymorphism or introgressive hybridization. Incomplete lineage sorting has previously been suggested to explain shared polymorphism of nuclear introns in the same lizards (Pinho *et al.* 2008). However, the time needed to completely sort the ancestral polymorphism of mtDNA is on average four times smaller than for nuclear genes (Pamilo & Nei 1988), so mitochondrial gene genealogies are predicted to reach reciprocal monophyly before nuclear gene genealogies.

In our case, two points suggest that the cyto-nuclear discordances are not resulting from incomplete lineage sorting. First, the Valencia lineage is embedded within a larger cytoplasmic clade with the North African and South Iberian lineage of *P. vaucheri* (Pinho *et al.* 2006). The *Liolepis* and *Hispanicus* mitochondrial lineages are grouped into another monophyletic clade. The common ancestor to all of these clades is older than 7 Myr (Pinho *et al.* 2006). This age estimate is old enough to suggest that complete lineage sorting of the ancestral polymorphism should have been achieved under the 'no gene flow' scenario, unless the effective population size of these lizards is unrealistically large (between 2 and 7 million individuals for generation times between 1 and 3 years). Second, with the incomplete sorting hypothesis, it would be highly unlikely to find Valencia haplotypes in both *P. liolepis* and *P. hispanicus*, while finding no trace of the Type 2, *bocagei* and 'Type 1' mtDNA lineages, even though these are more closely related to Valencia than are *Liolepis* and *Hispanicus*.

On the contrary, the introgressive hybridization model is quite compatible with our results. The Iberian Peninsula was a major ice age refugium during the Pleistocene and introgressive hybridization has been predicted to occur in such locations (Avice 2000). By driving successive cycles of species range expansion followed by retraction in many taxa, the Quaternary cyclic climatic oscillations repeatedly separated, then put back into contact, many evolutionary units (Hewitt 1996; Taberlet *et al.* 1998). If we hypothesize that the haplotypes of the Valencia lineage present in *P. hispanicus* and *P. liolepis* have been isolated from each other after they 'invaded' these taxa (no current mitochondrial gene flow between *P. liolepis* and *P. hispanicus*), the age of their separation is the maximal possible age for both introgression events. The divergence time between the Valencia haplotypes found in *hispanicus* and *liolepis* was estimated to be approximately 120 000/240 000 years ago. This time indicates that their introgression indeed occurred during the last Quaternary climatic oscillations, even if there is some uncertainty over the estimates (see Methods).

The calculation of divergence time presented above relies on the hypothesis that hybridizations have occurred following contacts between the unknown Valencia taxon and both *P. hispanicus* and *P. liolepis* independently. Alternatively, hybridization could have occurred following contact with one of these taxa only. Then, hybridization between *P. hispanicus* and *P. liolepis* would have allowed the transfer of the Valencia lineage from one taxon to the other. This hypothesis of indirect transmission has been proposed in Iberian hares (Melo-Ferreira *et al.* 2007, 2009) in which the mtDNA of the arctic species *Lepus timidus* would have been captured by *L. europaeus* through hybridization with *L. granaten-*

sis. Disentangling the two hypotheses will not be an easy task as long as we have no knowledge of the mtDNA present in the original Valencia-carrying taxon. However, if the Valencia mtDNA lineage had been passed from one taxon to the other (either from *liolepis* to *hispanicus* or the opposite), we would have expected the mtDNA polymorphism of the 'receiving' taxon to be embedded within the polymorphism of the 'donor' taxon. This is contrary to what we observe: the Valencia haplotypes that are present in *liolepis* and *hispanicus* form reciprocally monophyletic groups. We thus favour the hypothesis of two independent introgression events from the original Valencia-carrying taxon into *hispanicus* and *liolepis*. This hypothesis should nevertheless be tested by sequencing of mtDNA segments displaying more variability (such as ND4, see Pinho *et al.* 2008) to allow better examination of the relationships between the Valencia haplotypes present in *hispanicus* and *liolepis*, and by a more complete sampling of the region located south of the contact zone. Ultimately, an accurate test of the 'two introgression events' hypothesis would require locating remnant populations of the original Valencia-carrying taxon, provided they still exist. Currently, the taxon at the origin of the Valencia lineage has still not been identified.

The scenarios of mitochondrial replacement

The persistence of an alien mitochondrial lineage in a related taxon is not an uncommon phenomenon (reviewed in Mallet 2005). However, the extent of the mitochondrial introgression of the Valencia lineage within *P. liolepis* and the virtual lack of allelic polymorphism of this lineage deserve consideration. Two evolutionary scenarios can explain the current distribution of the Valencia mitochondrial lineage.

First, the distribution range of the Valencia-carrying taxon may have originally been more reduced than inferred from the current distribution of the Valencia mitochondrial lineage. After local hybridization with *P. liolepis*, this lineage would have secondarily replaced the *Liolepis* lineage in parts of its original range and would have spread until its current distribution in the region. In that case, the spread of the Valencia lineage over such a large area and the exclusion of the *Liolepis* lineage from all populations within this range are extremely unlikely by drift alone, and would rather suggest a selective sweep. Possible selection pressures on mtDNA include adaptation to climate, which appears to account for mitochondrial variation in humans (Mishmar *et al.* 2003; Ruiz-Pesini *et al.* 2004). The climate-driven selective hypothesis was originally emitted to explain the spread of the arctic *Lepus timidus* mitochondria within Iberian species of hares (Melo-Ferreira *et al.* 2005).

However, the high level of polymorphism of *timidus* mtDNA within Iberian populations of the other species and the polyphyletic origin of the introgressed haplotypes suggest multiple introgression events that do not strengthen the selective hypothesis (Melo-Ferreira *et al.* 2007). In the Valencia clade, only two haplotypes (one of which is rare and is derived from the common one by a single substitution) have been found in the southern *P. liolepis* populations, resulting in a reduced mtDNA diversity compared with the other mtDNA clades and taxa (Lirolepis clade in *P. liolepis*, Hispanicus clade in *P. hispanicus*). This supports the idea of a unique introgression event from the Valencia lineage towards *P. liolepis*, which fits well with the selective hypothesis. Besides, ectotherm organisms such as reptiles are good candidates for such selection scenarios, as mitochondrial variants may exhibit marked fitness differentials (Somero 2002). Last, the areas currently occupied by the Lirolepis and the Valencia clades incidentally experience contrasted climatic conditions (Pérez Cueva 1994).

In the second evolutionary scenario, the present distribution of the Valencia lineage would correspond to the former distribution range of the Valencia-carrying taxon. Direct competition with admixture in the colonization front between the invading *P. liolepis* and the invaded Valencia taxon could then have swamped most of the original nuclear genome but not the mtDNA. Recent theoretical models strongly suggest that when the range of a native species is invaded by another species, introgression of neutral genes occurs almost exclusively from the displaced species to the spreading species (Currat & Excoffier 2004; Currat *et al.* 2008). Moreover, the phenomenon would amplify as the invading species expands, leading to very high frequency or even fixation of foreign alleles in the invading species. This demographic hypothesis of competitive replacement would thus appear to fit our data well. However, simulation studies showed that complete mitochondrial replacement is reached only with a very high number of repeated hybridizations in the replacement front (Currat *et al.* 2008). The simulations are supported by empirical data on supposed competitive replacement: the foreign (captured) mitochondrial lineages are polyphyletic and the spreading lineage almost always persists in its own species, at least at low frequency (e.g. García-París *et al.* 2003; Roca *et al.* 2005; Berthier *et al.* 2006; Melo-Ferreira *et al.* 2007, 2009). None of these phenomena were observed with the Valencia mitochondrial lineage. In addition, in the case of neutral introgression, mtDNA would probably not be the only molecule to show signs of introgression. For example, in the case of the Iberian hares (Melo-Ferreira *et al.* 2009), the hypothesis of competitive replacement is favoured by signals of wide-range nuclear

introgression. In the case of Iberian Wall Lizards, we would expect that populations of the invading *P. liolepis* carrying the alien mtDNA haplotypes would also exhibit signs of nuclear introgression, and thus differ significantly in nuclear genes from nonintrogressed populations (here carrying the Lirolepis lineage). This was not the case in our study.

Factors responsible for reduced female-mediated gene flow compared with male-mediated gene flow, such as male-biased dispersal, would also favour the demographic scenario. In lizards, both male-biased (Ujvari *et al.* 2008) and female-biased (Olsson & Shine 2003; Ryberg *et al.* 2004) dispersal have been reported. It is therefore not possible to infer any gender-asymmetric dispersal in the genus *Podarcis* with current data. Moreover, dominance of the invading males over the invaded males could also account for differential introgression between maternally and biparentally inherited genes (Roca *et al.* 2005). However, such reproductive behaviour cannot be discussed as long as the Valencia-carrying taxon remains unidentified.

Further understanding of the processes that led to the extensive introgression of Valencia mtDNA into *P. liolepis* and *P. hispanicus* would clearly require the identification of the species at the origin of this lineage. On the one hand, the Valencia-carrying taxon could have totally disappeared. Mitochondrial introgression by a ghost lineage has been recently suspected in a fish genus (Bossu & Near 2009). On the other hand, the Valencia-carrying taxon may be extant but still undiscovered yet. To date, two mountain areas not sampled in this study are known to harbour populations with the Valencia lineage: the Sierra de Cazorla and the Sierra Nevada (Harris & Sá Sousa 2002; Pinho *et al.* 2006, 2007). One of these mountains, the Sierra de Cazorla, is already known to harbour a relict endemic lizard species, *Algyroides marchi*. During cooler glacial periods, a taxon inhabiting these mountain ranges today would probably have had a wider distribution, potentially contacting other *Podarcis* species. This putative scenario is nevertheless complicated by the result of Pinho *et al.* (2007) demonstrating the lack of nuclear genetic differentiation between *P. liolepis* and the Valencia populations of the Sierra Nevada. This finding is hardly explicable with our current knowledge of the species complex. Unravelling the origin of the Valencia taxon should be the goal of future works.

Conclusion: what confidence is there in single-marker based studies?

With the integrative approach in biology characterizing the new century (Pigliucci 2003), the drive to build taxonomic expertise is getting new impetus (Boero 2001).

The constantly growing complexity of morphology-based identification systems, coupled with the ceaseless decrease in the number of taxonomists, leads to the search for new approaches to taxon recognition. For example, the Barcode project proposes to use a single mitochondrial gene to identify and to discover new species (Hebert *et al.* 2003). In addition, some authors have explicitly assumed that for particular regions including refugia, single mtDNA sequences are good representatives of the genetic variability of such regions (e.g. Bilton *et al.* 1998). Recently, Barrowclough & Zink (2009) declared that 'if one's goal is to detect recently isolated groups of individuals or populations, i.e. discover taxa, the first step in a study ought to be the construction of an mtDNA gene tree'. One of the general messages of our study is to illustrate how misleading taxonomic conclusions would have been if based on mitochondrial markers, a message that refutes Barrowclough and Zink's statement. Indeed, the Valencia lineage has been assumed to correspond to a distinct taxon of the *P. hispanicus* complex (namely populations of *hispanicus sensu stricto*) in previous studies (Harris & Sá Sousa 2002; Harris *et al.* 2002; Pinho *et al.* 2003, 2006). None of these studies objectively analysed the morphological variation among and between lineages and therefore did not use it as a marker for systematics. The taxonomy of the *P. hispanicus* species complex is currently in revision and this revision would be inaccurate if based on mitochondrial data alone, particularly in southeastern Spain.

In addition to extensive mitochondrial introgressions shown in this study, incomplete lineage sorting is already known to affect our current view of individual nuclear gene genealogies in the *Podarcis hispanicus* species complex (Pinho *et al.* 2008). These two causes of discordance between gene trees and species trees are a major risk for single-marker taxonomy, particularly in regions where biodiversity is high (as in biodiversity hotspots or in past glacial refugia). These same areas are also those where the need to develop taxonomic work is most urgent. In other words, single-marker studies will be more prone to errors precisely in places where accurate characterization of biodiversity is most needed. In hotspots of diversity, only integrative studies will allow for a full understanding of singular events at the origin of each clade, and only an 'integrative taxonomy' (Will *et al.* 2005) will accurately recognize the specific richness in diversity of these regions.

Acknowledgements

We thank F. Kjellberg, D. McKey, D. Renoult, A. Baker and C. Pinho for constructive comments on an earlier version of this manuscript. We are also thankful for help and comments from

M.-P. Dubois regarding the molecular methods. J.P. Renoult was supported by a grant from the French Ministry of Research.

References

- Anderson TM, Lachance MA, Starmer WT (2004) The relationship of phylogeny to community structure: The cactus yeast community. *American Naturalist*, **164**, 709–721.
- Andreasen K (2005) Implications of molecular systematic analyses on the conservation of rare and threatened taxa: Contrasting examples from Malvaceae. *Conservation Genetics*, **6**, 399–412.
- Avis JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.
- Avis JC (2000) *Phylogeography: the History and Formation of Species*. Harvard Univ. Press, Cambridge, MA.
- Avis JC (2004) What is the field of biogeography, and where is it going? *Taxon*, **53**, 893–898.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.
- Barrowclough GF, Zink RM (2009) Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. *Molecular Ecology*, **18**, 2934–2936.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier, Montpellier.
- Bernatchez L, Glemet H, Wilson CC, Danzmann RG (1995) Introgression and fixation of arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 179–185.
- Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 3101–3109.
- Biek R, Walsh PD, Leroy EM, Real LA (2006) Recent common ancestry of Ebola Zaire virus found in a bat reservoir. *PloS Pathogens*, **2**, 885–886.
- Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 1219–1226.
- Bininda-Emonds ORP, Vazquez DP, Manne LL (2000) The calculus of biodiversity: integrating phylogeny and conservation. *Trends in Ecology & Evolution*, **15**, 92–94.
- Boero F. (2001) Light after dark: the partnership for enhancing expertise in taxonomy. *Trends in Ecology & Evolution*, **16**, 266–266.
- Bossu CM, Near TJ (2009) Gene trees reveal repeated instances of mitochondrial DNA introgression in orange-throat Darters (Percidae: *Etheostoma*). *Systematics Biology*, **58**, 114–129.
- Chapman JW, Carlton JT, Bellinger MR, Blakeslee AMH (2007) Premature refutation of a human-mediated marine species introduction: the case history of the marine snail *Littorina littorea* in the Northwestern Atlantic. *Biological Invasions*, **9**, 995–1008.

- Coullot A, Guillot G, Cosson JF *et al.* (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669–1679.
- Crochet PA, Chaline O, Surget-Groba Y, Debain C, Cheylan M (2004) Speciation in mountains: phylogeography and phylogeny of the rock lizards genus *Iberolacerta* (Reptilia: Lacertidae). *Molecular Phylogenetics and Evolution*, **30**, 860–866.
- Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA-sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, **12**, 814–822.
- Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *Plos Biology*, **2**, 2264–2274.
- Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution*, **62**, 1908–1920.
- Edwards SV, Kingan SB, Calkins JD *et al.* (2005) Speciation in birds: Genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6550–6557.
- Ennos RA, French GC, Hollingsworth PM (2005) Conserving taxonomic complexity. *Trends in Ecology & Evolution*, **20**, 164–168.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0, an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Felsenstein J (1981) Evolutionary trees from DNA sequences – a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368–376.
- Fontaine MC, Baird SJE, Piry S *et al.* (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biology*, **5**, 30.
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology and Systematics*, **34**, 397–423.
- García-París M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal of viviparity across contact zones in Iberian populations of Fire Salamanders (*Salamandra*) inferred from discordance of genetic and morphological traits. *Evolution*, **57**, 129–143.
- Geniez P (2001) *Variation géographique des lézards du genre Podarcis (Reptilia, Sauria, Lacertidae) dans la péninsule Ibérique, l'Afrique du nord et le sud de la France*. Université Montpellier II & Ecole Pratique des Hautes Etudes, Montpellier.
- Geniez P, Cluchier A, Sá Sousa P, Guillaume CP, Crochet PA (2007) Systematics of the *Podarcis hispanicus* complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetological Journal*, **17**, 69–80.
- Gillespie RG, Roderick GK (2002) Arthropods on islands: Colonization, speciation, and conservation. *Annual Review of Entomology*, **47**, 595–632.
- Guillaume CP (1987) *Les petits Lacertidés du bassin méditerranéen occidental (Genera Podarcis et Archaeolacerta essentiellement). Sur quelques problèmes d'ordre systématique et biogéographique*. Université des Sciences et Techniques du Languedoc, Montpellier.
- Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.
- Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with GENELAND: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, **24**, 1406–1407.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.
- Harris DJ, Sá Sousa P (2001) Species distinction and relationships of the western Iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetological Journal*, **11**, 129–136.
- Harris DJ, Sá Sousa P (2002) Molecular phylogenetics of Iberian wall lizards (*Podarcis*): Is *Podarcis hispanica* a species complex? *Molecular Phylogenetics and Evolution*, **23**, 75–81.
- Harris DJ, Carranza S, Arnold EN, Pinho C, Ferrand N (2002) Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, **29**, 1257–1262.
- Harrison RG (1990) Hybrid zones: windows on evolutionary process. *Oxford Survey in Evolutionary Biology*, **7**, 69–128.
- Hebert PDN, Cywinska A, Ball SL, Dewaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 313–321.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
- Horton DR (1972) Lizard scales and adaptation. *Systematic Zoology*, **21**, 441–443.
- Jennings WB, Edwards SV (2005) Speciation history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution*, **59**, 2033–2047.
- Klein NK, Payne RB (1998) Evolutionary associations of brood parasitic finches (*Vidua*) and their host species: Analyses of mitochondrial DNA restriction sites. *Evolution*, **52**, 566–582.
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.
- Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**, 229–237.
- Martinez-Solano I, Teixeira J, Buckley D, Garcia-Paris M (2006) Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, **15**, 3375–3388.
- Mateo JA, LopezJurado LF, Guillaume CP (1996) Proteic and morphological variations in ocellated lizards (Lacertidae): A complex of species across the Strait of Gibraltar. *Comptes Rendus De L'Academie Des Sciences Serie III-Sciences De La Vie-Life Sciences*, **319**, 737–746.
- Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC (2005) Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, **14**, 2459–2464.
- Melo-Ferreira J, Boursot P, Randi E *et al.* (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, **16**, 605–618.
- Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genome legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, **18**, 2643–2658.

- Minelli A (2007) Invertebrate taxonomy and evolutionary developmental biology. *Zootaxa*, **1668**, 55–60.
- Mishmar D, Ruiz-Pesini E, Golik P *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 171–176.
- Moran P, Kornfield I (1993) Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei, Cichlidae) of lake Malawi. *Molecular Biology and Evolution*, **10**, 1015–1029.
- Morando M, Avila LJ, Baker J, Sites JW (2004) Phylogeny and phylogeography of the *Liolaemus darwini* complex (Squamata: Liolaemidae): Evidence for introgression and incomplete lineage sorting. *Evolution*, **58**, 842–861.
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, **16**, 358–364.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Olsson M, Shine R (2003) Female-biased natal and breeding dispersal in an alpine lizard, *Niveoscincus microlepidotus*. *Biological Journal of the Linnean Society*, **79**, 277–283.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.
- Parra-Olea G, Wake DB (2001) Extreme morphological and ecological homoplasy in tropical salamanders. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 7888–7891.
- Pérez Cueva A (1994) *Atlas climático de la comunidad Valenciana*. Comunidad Valenciana, Valencia.
- Pigliucci M (2003) From molecules to phenotypes? The promise and limits of integrative biology. *Basic and Applied Ecology*, **4**, 297–306.
- Pinho C, Harris DJ, Ferrand N (2003) Genetic polymorphism of 11 allozyme loci in populations of wall lizards (*Podarcis* sp.) from the Iberian Peninsula and North Africa. *Biochemical Genetics*, **41**, 343–359.
- Pinho C, Sequeira F, Godinho R, Harris DJ, Ferrand N (2004) Isolation and characterization of nine microsatellite loci in *Podarcis bocagei* (Squamata: Lacertidae). *Molecular Ecology Notes*, **4**, 286–288.
- Pinho C, Ferrand N, Harris DJ (2006) Reexamination of the Iberian and North African *Podarcis* (Squamata : Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, **38**, 266–273.
- Pinho C, Harris DJ, Ferrand N (2007) Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (*Podarcis*, Lacertidae). *Biological Journal of the Linnean Society*, **91**, 121–133.
- Pinho C, Harris DJ, Ferrand N (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. *BMC Evolutionary Biology*, **8**, 63.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature Genetics*, **37**, 96–100.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223–226.
- Ryberg K, Olsson M, Wapstra E, Madsen T, Anderholm S, Ujvari S (2004) Offspring-driven local dispersal in female sand lizards (*Lacerta agilis*). *Journal of Evolutionary Biology*, **17**, 1215–1220.
- Sá Sousa P (2000) A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetological Journal*, **10**, 1–11.
- Schelly R, Salzburger W, Koblmüller S, Duftner N, Sturmbauer C (2006) Phylogenetic relationships of the lamprologine cichlid genus *Lepidolamprologus* (Teleostei : Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *Molecular Phylogenetics and Evolution*, **38**, 426–438.
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.
- Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16122–16127.
- Somero GN (2002) Thermal physiology and vertical zonation of intertidal animals: Optima, limits and cost for living. *Integrative and Comparative Biology*, **42**, 780–789.
- Sullivan JP, Lavoue S, Arnegard ME, Hopkins CD (2004) AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea : Teleostei). *Evolution*, **58**, 825–841.
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, **15**, 199–203.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Ujvari B, Dowton M, Madsen T (2008) Population genetic structure, gene flow and sex-biased dispersal in frillneck lizards (*Chlamydosaurus kingii*). *Molecular Ecology*, **17**, 3557–3564.
- Vaconcelos R, Carretero MA, Harris DJ (2006) Phylogeography of the genus *Blanus* (worm lizards) in Iberia and Morocco based on mitochondrial and nuclear markers preliminary analysis. *Amphibia-Reptilia*, **27**, 339–346.
- Vazquez DP, Gittleman JL (1998) Biodiversity conservation: Does phylogeny matter? *Current Biology*, **8**, R379–R381.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Julien P. Renoult is a PhD student in evolutionary biology, working on phylogeny and phylogeography of several animal and plant models, and on the evolution of visual signals in animal communication. Philippe Geniez is a permanent researcher in herpetology specialized in the Western Palaearctic fauna. Laure Benoit is a DNA technician and Paul Bacquet is a PhD student interested in phylogeography, chemical ecology and speciation. Pierre André Crochet is a permanent researcher working on phylogeography, phylogeny, evolutionary ecology and taxonomy of birds, reptiles and amphibians of the Western Palaearctic.

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 List of collections that provided specimens analysed in this study

Table S2 Morphological, mitochondrial and nuclear classifications for each location

Table S3 List of morphological variables used in multivariate analyses with details of the semi-quantitative transformation for qualitative variables

Table S4 List of individuals sequenced for the CR mitochondrial gene with mitochondrial lineage and corresponding GenBank accession numbers

Data S5 Protocol of microsatellite amplification.

Table S6 *P*-values of the linkage disequilibrium between the eight microsatellite loci calculated with 1000 permutations for the three morphotypes: *hispanicus*, *liolepis* and 'type 2'. NS, nonsignificant; **P* < 0.05; ***P* < 0.01

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Cyto-nuclear discordance in the phylogeny of *Ficus* section *Galoglychia* and host shifts in plant-pollinator associations

Julien P. Renoult¹, Finn Kjellberg^{1§}, Cinderella Grout², Sylvain Santoni³, Bouchaïb Khadari^{4,5}

¹CNRS, UMR 5175 Centre d'Ecologie Fonctionnelle et Evolutive, Equipe Interactions Biotiques, F-34293 Montpellier cedex 5, France

²Montpellier SupAgro, UMR 1098, Développement et Amélioration des Plantes, Bat. 2, Campus CIRAD TA A 96/03, Av. Agropolis, 34398 Montpellier cedex 5, France

³INRA, UMR 1097, Diversité et Adaptation des Plantes Cultivées, Bat. 33, 2 place Viala, 34060 Montpellier Cedex 2, France

⁴INRA, UMR 1098, Développement et Amélioration des Plantes, Bat. 2, Campus CIRAD TA A 96/03, Av. Agropolis, 34398 Montpellier cedex 5, France

⁵Conservatoire Botanique National Méditerranéen de Porquerolles, UMR 1098, 76 Av. Gambetta, 83400 Hyères, France.

[§]Corresponding author

Email addresses:

JPR: julien.renoult@cefe.cnrs.fr

FK: finn.kjellberg@cefe.cnrs.fr

BK: khadari@supagro.inra.fr

Running title: Cyto-nuclear discordances in *Ficus*

Abstract

Background

Hybridization events are relatively common in vascular plants. However, the frequency of these events is unevenly distributed across the plant phylogeny. Plant families in which individual species are pollinated by specific pollinator species are predicted to be less prone to hybridization than other families. However, exceptions may occur within these families, when pollinators shift host-plant species. Indeed, host shifts are expected to increase the rate of hybridization events. Pollinators of *Ficus* section *Galoglychia* are suspected to have changed host repeatedly, based on several cases of incongruence between plant phylogeny and taxonomy, and insect phylogeny and taxonomy. We tracked cyto-nuclear discordance across section *Galoglychia* as evidence for hybridization. To achieve a proper global view, we first clarified the monophyly of section *Galoglychia* as it had been questioned by recent phylogenetic studies. Moreover, we investigated if fig size could be a factor facilitating host shifts.

Results

Phylogenetic chloroplast and nuclear results demonstrated the monophyly of section *Galoglychia*. Within section *Galoglychia*, we detected several cases of statistically significant cyto-nuclear discordance. Discordances concern both terminal nodes of the phylogenetic trees and one deep node defining relationships between subsections. Because nuclear phylogeny is congruent with morphological taxonomy, discordances were caused by the chloroplast phylogeny. Introgressive hybridization was the most likely explanation for these discordances. We also detected that subsections pollinated by several wasp genera had smaller figs and were pollinated by smaller wasps than subsections pollinated by a single wasp genus.

Conclusions

As hypothesized, we discovered evidences of past hybridization in *Ficus* section *Galoglychia*. Further, introgression was only detected in subsections presenting incongruence between plant and pollinator phylogenies and taxonomy. This supports the hypothesis that host shift is the cause for plant-pollinator incongruence. Moreover, small fig size could facilitate host shifts. Eventually, this study demonstrates that non-coding chloroplast markers are valuable to resolve deep nodes in *Ficus* phylogeny.

Key words: phylogeny, fig size, codivergence, cospeciation, introgression, non coding cpDNA, ITS, ETS

Background

Frequent natural hybridization events in vascular plants were documented by early biologists (e.g. [1, 2]). However, the prevalence of hybridization events is unevenly distributed across the plant phylogeny and seems to be concentrated within a small fraction of families and genera [3]. Grant [4] pointed out that the frequency of natural hybridizations varies with factors such as life history, breeding system, environmental disturbance, genetic predisposition, and eventually pollination syndrome. In this context, plant species pollinated by specific pollinator species are predicted to present almost no hybridization events. Indeed, in specific mutualisms as plant-pollinator associations, the associated species usually present co-adaptations involved in attraction, recognition and physical compatibility with the other species. Such adaptations can be viewed as pre-zygotic barriers limiting interspecific hybridization [5-7].

Host shift is one of the main ecological processes that can break patterns of strict cospeciation between tightly bound interacting species [8-11], as observed in some cases affecting postulated strict co-evolutionary pattern in specific pollination mutualisms [12]. When associated with a transitional use of two hosts by the pollinator, a host shift may allow plant hybridization and hence leave a signature of genetic introgression in the plant genome. Following such hybridization, recurrent backcrosses of offspring with one parent may also allow the introgression of a limited set of foreign alleles into a plant species [13]. Hence, plant lineages in which pollinators have jumped from one host to another are more likely to exhibit signs of introgression than other lineages.

Detecting past events of host shifts requires tracking incongruence between phylogenetic trees of interacting taxa [8, 14, 15]. However, pollinator species duplication in the absence of host speciation, followed by cospeciation and subsequent asymmetric extinction, a phenomenon we will call duplication/extinction hereafter, is another mechanism that can cause incongruence between host phylogeny and partner phylogeny [16]. Although different in essence, these phenomena lead to similar patterns of phylogenetic incongruence, and discriminating between them is often largely speculative [17]. Nevertheless, and contrary to host shift, duplication/extinction does not involve plant hybridization. Consequently, detecting introgressed genes in a plant lineage for which there is incongruence between plant and pollinator phylogenies would support the host shift hypothesis.

Fig trees (*Ficus*, Moraceae) and their pollinating wasps (Agaonidae, Chalcidoidea) constitute one of the well-known cases of species-specific pollination mutualism. In this system, the rule is that a wasp species generally pollinates a single host species [18, 19]. Further *Ficus* and Agaonid wasp morphological taxonomy and molecular phylogenies are steadily improving, providing a unique set of information to infer past history of the association on a broad set of species. Based on phylogenetic data, several authors have suggested parallel diversification of host and associated pollinating wasps [5, 20-23]. Moreover, fig tree sections or subsections appear to be generally monophyletic and pollinated by a single wasp genus or by a few genera. Nevertheless, exceptions to the rule of a single agaonid wasp species associated with a single *Ficus* species are quite frequent and rather well documented based on morphology [24, 25] as well as on molecular methods [12, 26-28]. Indeed, Rasplus *et al.* [29] estimated for Africa, on the basis of morphology, that 17% of fig species were pollinated by more than one wasp species and that 15% of fig pollinator wasps used two or more hosts. However, molecular data on Australian pollinators suggest that when two or more wasps pollinate the same host they seem to often be sister species [30].

Within this general context, species of section *Galoglychia* show an unusual pattern of association with their pollinating wasps within genus *Ficus*. The 77 described species of section *Galoglychia* are distributed into six subsections [18, 31], several of which are unambiguously defined. Seven morphologically unambiguous genera of fig wasps are known to pollinate section *Galoglychia* (Fig. 1). Some genera are both subsection specific and the sole pollinators of the subsection. Other genera pollinate fig trees from different subsections and some subsections are pollinated by several wasp genera. Further some individual *Ficus* species may be pollinated by wasps belonging to different genera (e.g. *Ficus natalensis* pollinated by *Alfonsiella* and *Elisabethiella*) and one wasp species may pollinate several *Ficus* species (e.g. *Elisabethiella stuckenbergii* pollinates *F. burkei*, *F. natalensis*, *F. petersii* and *F. lingua depauperata* [32]).

Recent molecular phylogenies supported, with some caveats, the monophyly of subsections within section *Galoglychia* [33] and that of wasp genera [26]. Hence the interactions between figs of section *Galoglychia* and their pollinating wasps present us with a series of instances of lack of strict specificity, and lack of congruence between fine wasp and *Ficus* taxonomy and phylogeny [18, 34]. Comparing the phylogenies of trees and insects suggests additional lack

of congruence, with respect to the order of branching of *Ficus* subsections and wasp genera (Fig. 1). In the plant phylogeny [33], based on ITS and ETS nuclear markers, subsections *Cyathistipulae* and *Crassicostae* are grouped with subsections *Galoglychia* and *Caulocarpae* within a clade, while the remaining subsections are clustered together into a second clade. In the wasp pollinator phylogeny [26], based on 28S and ITS markers, subsection *Galoglychia* pollinators and *Caulocarpae* pollinators constitute the outgroups of a clade including the pollinators of the remaining subsections. Fig tree phylogeny is thus incongruent with pollinating wasp phylogeny with respect to the branching of subsections *Cyathistipulae* and *Crassicostae* versus *Chlamydodora* and their pollinators. An additional finding of the molecular phylogeny of section *Galoglychia* is a suggestion that neotropical species of section *Americana* could be enclosed within section *Galoglychia*, although this result was only weakly supported by the various node validation methods [33]. If section *Galoglychia* is really paraphyletic to section *Americana*, pollinators of this last section could provide an example, pending phylogenetic and taxonomic re-evaluation, of discrepancy between the phylogenetic histories of *Ficus* and their pollinating wasps as the pollinators of section *Americana* and of section *Galoglychia* have been classified into different subfamilies (Agaoninae and Blastophaginae) [19].

Incongruence between pollinator and host phylogenies has been traditionally attributed to host shifts [12, 35]. However, ancient pollinator species duplication events could also have lead to different wasp pollinator genera pollinating the same host figs. Subsequent cospeciation and asymmetric extinction could be responsible of the apparently haphazard host association exhibited by *Alfonsiella* and *Elisabethiella* pollinators [26, 36].

The goal of this contribution is to study introgression events in the phylogeny of section *Galoglychia*. If host shifts and not duplications/extinctions are effectively responsible for plant-pollinator taxonomic and phylogenetic incongruence, then we expect to detect introgression events, especially in the plant lineages implied in this incongruence. We compared nuclear and chloroplast phylogenies to detect cases of genetic introgression between *Ficus* species, because the chloroplast genome of vascular plants possess several features that facilitate its introgression into a new plant species after hybridization compared to the nuclear genome [13, 37, 38], as illustrated by the numerous cases of chloroplast introgression documented in the literature [37, 39, 40]. Our working hypothesis is that the current complex pattern of host association of genera *Elisabethiella*, *Nigeriella* and

Alfonsiella result from host shifts. The prediction is that we may detect evidence for cytoplasm transfer between species of subsection *Chlamydodora*, *Platyphyllae* and *Crassicostae*. We will also search for other potential cases of genetic introgression and especially for subsection *Cyathistipulae* for which available data suggests mismatch between the phylogenetic position of the subsection [33] and the phylogenetic position of its pollinators, namely genus *Agaon* [26]. To do so, we first developed specific chloroplast markers for *Ficus* and sequenced them in 58 individuals representing 38 species of section *Galoglychia* and ten species of section *Americana*. We also sequenced ITS and ETS markers on these same individuals. Second, we searched for cases of discordance between chloroplast and nuclear phylogenies and tested their significance. Third, we showed that genetic introgression was the most likely mechanism to explain the cases of discordance. Besides, because section *Galoglychia* is the sole *Ficus* section subdivided into six subsections all of which are well defined (pending revision of the taxonomic position of some species assigned to subsection *Platyphyllae*, see discussion) and the sole section pollinated by six monophyletic agaonid wasp genera, this section offers a unique opportunity to investigate host shifts and their determinants. Within the section, known and suspected examples of pollinators using several hosts seem to involve host species presenting rather small figs, such as *F. burkei*, *F. natalensis*, *F. petersii* and *F. lingua depauperata* [32]. Therefore we investigated whether the fig subsections involved effectively presented small figs and whether their pollinator genera presented small body size.

Results

Data set partitioning

Chloroplast marker development allowed producing fourteen pairs of primer specifically developed for *Ficus* (additional file S2). Five of them – *atpB-rbcL*, *FcB*, *FcJ*, *FcL* and *trnL-trnF* – were selected for this analysis based on the phylogenetic information provided by a subsample of the final data set used in this study. The final combined chloroplast matrix contained 58 accessions and 3,604 base pairs (additional file S3; TreeBASE accession number: SN4278). One hundred and twenty-two sites were variable and 52 (1.5%) were potentially parsimony informative. The model of sequence evolution varied across chloroplast partitions (Table 1). We used arithmetic mean $-\ln L$ of Bayesian posterior topologies to measure the ability of data partitioning to explain the entire data set. The analysis separating

non-coding and coding regions (strategy S_3) of the chloroplast data set returned a decisively better description of the data than the other two analyses according to Bayes factors (Table 2).

The combined ITS and ETS matrix contained 1,320 base pairs of which 348 were variable and 158 (12%) were potentially parsimony informative. The model of sequence evolution differed between the ITS and ETS partitions (Table 1). The strategy individualizing each of these partitions (S_5) was decisively better than the strategy combining the entire nuclear data set (S_4) (Table 2). In the following we will only present the results obtained using the best strategy for each data set.

Chloroplast phylogeny of section Galoglychia

The combined chloroplast matrix produced twelve most parsimonious trees of length 141. Consistency and Retention Indexes (CI = 0.88; RI = 0.93) suggested the presence of a single island of trees [41]. The strict consensus of all most parsimonious trees (SCMP) is shown in Figure 2A with bootstrap percentages (MP_BS). The two different runs during the BI analysis with partitioning strategy S_3 reached model parameter convergence and generated a similar tree. This tree is shown in Figure 3 with the posterior probabilities (PP) indicated for each node. ML analysis produced one most likely tree (-lnL = 5960.04; not shown), slightly less well resolved than the Bayesian tree. Bootstrap percentages (ML_BS) are indicated over the Bayesian tree in Figure 3.

The SCMP tree, the ML and the Bayesian trees were all similar in their overall topology. The monophyly of section *Americana* was strongly supported (91;89;1 for %MP_BS, %ML_BS and PP respectively; node supports are given in a similar order in the following) and formed a sister clade to section *Galoglychia* which was monophyletic (88;92;1). Subsections *Galoglychia* and *Caulocarpae* constituted monophyletic groups (80;90;1 and 86;93;1, respectively) and were basal within section *Galoglychia*. A third well supported clade was composed by species belonging to subsections *Cyathistipulae*, *Crassicostae*, *Chlamydodora* and *Platyphyllae* (94;96;1; CCCP clade hereafter). However and contrary to the previous subsections, within the CCCP clade subsections did not form monophyletic groups. For example, the two accessions of *F. trichopoda* branched within distinct clades, as did the two accessions of *F. craterostoma*, and *F. densistipulata* (subsection *Cyathistipulae*) branched within a clade only composed of taxa of subsection *Chlamydodora*.

One hundred and four distinct gaps were coded for the chloroplast data set. Five insertions and eight deletions were synapomorphic (Fig. 3; additional file S4). A 10 bp insertion (chloroplast indel n°4) supported the monophyly of section *Galoglychia* since it was found neither in the outgroup nor in section *Americana*. A 7 bp deletion (n°6) supported the CCCP clade.

Nuclear phylogeny of section Galoglychia

The combined nuclear matrix produced 15,000 most parsimony informative trees of length 521 (RI = 0.84; CI = 0.75). The SCMP of all 15,000 trees is shown in Figure 2B. Figure 3 illustrates the unique Bayesian tree produced by the two different runs during the analysis and using the S_5 partitioning strategy. Only one most likely tree (-lnL = 5150.78; not shown) was produced by the ML analysis. As with the chloroplast data, it differed from the Bayesian tree only in being slightly less resolved.

Here again, the SCMP, the ML and the Bayesian trees were all similar in their overall topology. A basal trichotomy divided the ingroup into three clades. The first one was the strongly supported section *Americana* (91;99;1). The second grouped subsection *Chlamydodora* and part of subsection *Platyphyllae* (Clade A). Although it was not strongly supported by bootstraps (55;55), this clade was found in the 15,000 most parsimonious trees, and had a Bayesian posterior probability of one. The third clade grouped accessions from subsections *Caulocarpace*, *Galoglychia*, *Cyathistipulae*, *Crassicostae* and part of *Platyphyllae* (Clade B). It was not strongly supported by the bootstrap distribution (71;72) but it was present in every most parsimonious trees and is maximally supported by the BI analysis.

Sixty five indels were coded for the nuclear data set and 10 of them were synapomorphic (Fig. 3; additional file S4). The ingroup is supported by one deletion and one insertion. Inside the ingroup, all accessions of section *Americana* presented a 3 bp deletion in their ETS sequence while all accessions of section *Galoglychia* shared a one base pair insertion in the ITS sequence.

Investigation of the cyto-nuclear discordance

While different phylogenetic reconstruction methods gave rise to similar topologies for a given data set, chloroplast and nuclear phylogenies appeared discordant for several internal and external nodes. Templeton tests applied to MP trees, and SH tests to ML trees always

returned congruent results (Table 3). Both SCMP and ML trees explained significantly better the data set used to reconstruct these trees than the rival data set, demonstrating the overall discordance between nuclear and chloroplast topologies. We then tested the significance of local discordances involving *F. trichopoda*, *F. craterostoma* and *F. densistipulata* accessions. The original chloroplast topology in which both accessions of *F. trichopoda* are paraphyletic explained significantly better the chloroplast data than a modified topology in which these accessions were branched to form a monophyletic relationship. The same was true with the *F. craterostoma* chloroplast accessions. However, the original chloroplast topology was not significantly better than the topology in which *F. densistipulata* was branched with other *Cyathistipulae* taxa to make this subsection monophyletic. Alternatively, modifying the nuclear topology to branch one of the two *F. trichopoda* and one of the two *F. craterostoma* accessions with *Chlamydorae* and with *Cyathistipulae* and *Crassicostae* accessions respectively, as in the chloroplast topology, produced significantly worse topologies than the unmodified nuclear topologies for the nuclear dataset. The same was true when *F. densistipulata* was branched with the *Chlamydorae* species as in the chloroplast topology. For these last tests, it was possible to artificially reconstruct several alternative topologies. Table 3 gives results only for possibilities giving the highest *P*-value.

Besides the terminal cases of discordance tested above, a cyto-nuclear discordance was generated by the branching of subsections *Cyathistipulae* and *Crassicostae*. To avoid interference due to cases of terminal discordance, we tested the significance of this deep discordance by comparing topologies differing only by the branching of these two subsections (Table 3). Therefore, branching was left unresolved (i.e. with a polytomy) inside subsections *Galoglychia* and *Caulocarpae* for the chloroplast topology, and inside the CCCP clade, Clade A and Clade B (see Figure 3). Both Templeton and SH tests showed that the position of subsections *Cyathistipulae* and *Crassicostae* obtained in trees reconstructed using one data set (nuclear or cytoplasmic) produced a significantly better topology than the rival topology for that set of data.

The Bayesian approach used to statistically evaluate discordances gave similar results (not shown). For each Bayesian analysis, we did not find any branching supporting a rival phylogenetic hypothesis in any tree of the 95% credible set sampled once stationarity was reached. The sole exception was constituted by *F. densistipulata* in the chloroplast data set

analysis, which branched with other *Cyathistipulae* species in 17% of the trees of the 95% credible set.

To check that apparent cyto-nuclear incongruence was not an artefact due to saturation by mutations, we plotted the estimated number of transitions and transversions against the TN93 [42] genetic distances (additional file S5). For all three analyses, transitions outnumber transversions indicating that substitutions are not saturated in the three types of data (chloroplast, ETS and ITS). Finally, considering ITS plus ETS as a single nuclear unit, a scenario based on gene duplication without genetic introgression and explaining the deep node discordance between nuclear and cytoplasmic phylogenies requires one duplication event at the base of section *Galoglychia* followed by three extinction events, each at the base of a group of subsections (Fig. 5).

Fig and pollinator size

We investigated if fig and pollinator size were morphological traits associated at the subsection/genus level with plant-pollinator phylogenetic incongruence. Indeed, because of the as yet limited number of documented cases of incongruence, we limited our investigation to crude comparisons. Dry fig diameter of species belonging to section *Galoglychia* was distributed into two size categories (Fig. 4A). Species belonging to subsections pollinated by a single pollinator genus (subsection *Cyathistipulae*, *Caulocarpae* and *Galoglychia*) presented significantly larger figs than species belonging to sections pollinated by at least two pollinator genera ($t=7.63$; $p<10^{-3}$; not corrected for phylogenetic correlations). Similarly, body length of pollinating wasps was distributed into two size categories (Fig. 4B): wasp genera that pollinated fig trees from a single subsection (i.e. *Agaon*, *Paragaon*, *Courtella* and *Allotriozon*) were significantly larger than genera pollinating several plant subsections ($t=7.43$; $p<10^{-3}$; not corrected for phylogenetic correlations). Graphically, genus *Paragaon* presented small-sized wasps and therefore escapes this pattern. However, this genus includes only two species and for only one of them is the host known. Wasp size in this genus is thus little informative.

Discussion

The monophyly of section Galoglychia

Rønsted *et al.* [33] did not demonstrate the monophyly of the African section *Galoglychia*. Based on ITS+ETS data, they suggested that section *Galoglychia* might be paraphyletic to the American section *Americana*, although this result was ambiguously supported. With our ITS+ETS data set, all three types of analysis (MP, ML and BI) also failed to resolve the placement of section *Americana* relative to section *Galoglychia*, although a 3 bp deletion in the ETS sequence characterises section *Americana* and a 1 bp insertion in the ITS sequence characterises section *Galoglychia*. However, the chloroplast phylogeny provides strong support for the monophyly of section *Galoglychia* with each of the three methods. Moreover, a 10 bp insertion and a 1 bp deletion in fragment *FcL* characterises species of section *Americana* and a 10 bp insertion in the same gene characterises species of section *Galoglychia*. Hence the phylogeny based on chloroplast data enables to resolve the trifurcation of the phylogeny based on ITS+ETS data and the result is further supported by insertion-deletion data which were not taken into account in the phylogenetic reconstructions. Eventually, our chloroplast data set provided less information on terminal nodes, but offered better resolution for deeper nodes. Because chloroplast genes overall evolve more slowly than nuclear genes do [43], they are less prone to homoplasy and are therefore usually recommended for deep phylogenies [44-46].

Morphologically, sections *Galoglychia* and *Americana* share the presence of two bracts subtending the figs as opposed to generally three in genus *Ficus*. Section *Galoglychia* is characterised by all the bracts of the ostiole (the passage leading into the fig) turned inward, the orifice of the ostiole forming a bilabiate slit. In section *Americana*, the bracts of the ostiole are imbricate so that generally three bracts are visible as closing the ostiole. The monophyly of section *Galoglychia*, as recognized by botanists since the major revision of African figs ([47]; recognized as subgenus *Bibracteatae* at that time), is hence most probably correct.

The infrasectional classification of section Galoglychia

The nuclear phylogeny proposes the division of section *Galoglychia* into two clades that do not include the same subsections as the two clades evidenced by the chloroplast analyses. Although the two nuclear clades are weakly supported (as previously observed by Rønsted *et al.* [33]), several features suggest that this phylogeny is reliable. First, the alternative topology suggested by the chloroplast data set was demonstrated to be significantly less likely than the nuclear topology. Second, although 35 % of our taxa were different from those used by Rønsted *et al.*, we found a similar topology with similar values of bootstrap supports. It is

thus unlikely that the sampling composition (which covers half of the species of section *Galoglychia*) biased the phylogenetic reconstruction. Third, the position of most species remains consistent with classical taxonomy [18]. A pattern seems to appear in the taxonomic modifications revealed by molecular phylogenies in *Ficus*, suggesting that they do not result from identification/manipulation errors or from limiting phylogenetic information. Indeed, within subsection *Platyphyllae*, *F. platyphylla* and *F. jansii* both present clusters of small figs at the axils of the leaves or just below, a trait they share with a number of species of subsection *Crassicostae*, with which they are grouped in the phylogenetic tree. Conversely, *F. glumosa* and *F. stuhlmannii* only present two figs at the axils of the leaves and are grouped with species from subsection *Chlamydodora*, which all have figs in pairs at the axils of the leaves. Preliminary data on some other species of subsection *Platyphyllae* seem to support this splitting of the species into two different units, mainly according to this trait. Fourth, a potential source of error using ITS and ETS could be the amplification of paralogous copies of ITS and ETS leading to an incorrect phylogenetic reconstruction [48-50]. However, here again, the monophyly of each subsection is hardly compatible with the amplification of paralogous copies of ITS and ETS that generally evolve independently [51]. Four events of extinction following the duplication would have to be hypothesised to obtain an observed nuclear topology corresponding to the chloroplast topology (Fig. 5). Although this number of duplication-extinctions can reasonably occur given the considered time lapse [39], these four events would have had to occur in the ancestor of each of the four extant subsections and not within the subsections, a pattern which is unlikely. Last, the separation of section *Galoglychia* into two nuclear clades is supported by ecological and biogeographical arguments [33]. All these lines of evidence lead us to conclude that subsections *Galoglychia*, *Cyathistipulae*, *Crassicostae*, *Caulocarpace* and part of *Platyphyllae* on one hand, and subsection *Chlamydodora* and part of *Platyphyllae* on the other hand, form two monophyletic clusters.

A discordant chloroplast topology

Our investigation of the overall discordance between the chloroplast and the nuclear topology revealed two types of discordances occurring at two taxonomic levels. The first type concerns phylogenetic relationships between subsections: subsections *Cyathistipulae* and *Crassicostae* branch with subsections *Chlamydodora* and *Platyphyllae* in the chloroplast topology, and with subsections *Galoglychia* and *Caulocarpace* in the nuclear topology. As highlighted above, the nuclear topology is consistent with morphological and biogeographical data. We can thus confidently assume that it is the chloroplast and not the nuclear topology that is

discordant with the species tree. The second type of discordance concerns the monophyly of subsections: *Chlamydodora* and *Platyphyllae* are paraphyletic in the chloroplast phylogeny (at least for *F. craterostoma* and *F. trichopoda*), and monophyletic in the nuclear phylogeny. The monophyly of subsections found in the nuclear tree is consistent with the morphological traits that led botanists to recognize these subsections for a century [47], meaning that here again, the nuclear but not the chloroplast phylogeny is most likely to reflect the species tree.

Examples of cyto-nuclear discordance are rather common in plants [52-57] and a limited number of causes leading to such discordances has been identified. Discordance between cytoplasmic and nuclear phylogenetic trees may either be artefactual or evidence distinct evolutionary histories followed by cytoplasmic and nuclear genomes [39]. Long Branch Attraction (LBA hereafter; [58]) is a source of artefact that may generate cyto-nuclear discordance [59, 60]. Several arguments allow us to exclude LBA artefacts for our data set. First, the branches leading to discordance are not the longest branches of the tree, while some longer branches do not have unexpected positions in the tree. Further, the molecular data sets are not saturated which means that our markers did not evolve rapidly enough for phylogenetic signals to be lost. Finally, the discordance was the same whatever the phylogenetic reconstruction method employed. Although methodological concordance does not rule out LBA [61], parsimony has a stronger bias towards grouping long branches together than ML and BI which take into account unequal rates or branch lengths [62, 63]. It is therefore unlikely that the observed cases of discordance resulted from LBA.

The cyto-nuclear discordance has therefore to be explained in terms of difference between species and chloroplast history. Two mechanisms may be involved: incomplete chloroplast lineage sorting and introgression. Differentiating between hypotheses of introgression and lineage sorting is difficult because both processes can generate very similar phylogenetic patterns [64]. Ancestral polymorphism is expected to be completely sorted out with a high probability approximately $4N_e$ generations (N_e = effective population size) after the separation [65]. The branch supporting the CCCP clade (BL = 0.1774) is longer than the branch supporting clades in which reciprocal monophyly has been reached, like the *Caulocarpace* clade (BL = 0.1699), the *Galoglychia* clade (BL = 0.1269) and section *Americana* (BL = 0.1756). Hence, this branch is expected to be long enough to have allowed chloroplast genes to reach reciprocal monophyly as well and hence to complete allele sorting. However, the sorting depends on N_e and not on branch length directly. Therefore, any variation in N_e over

time and across species would alter conclusions based on relative branch length comparison. Methods based on coalescent simulations to calculate the probability of observing lineage sorting are also based on N_e remaining constant over time and among species [66] and both hypotheses are not realistic when comparing savannah and forest lineages over longer periods of time. Moreover, they require estimating ancestral population sizes that can only be obtained by intensive intraspecific sampling and this is not compatible with the phylogenetic approach used here. Eventually, these methods require generation time to be known and do not allow this parameter to vary across the phylogeny. *Ficus* lineages included in our phylogeny are highly different in term of height and life forms, with some lineages represented mainly by canopy emergent trees (e.g. subsection *Caulocarpae*) and other mainly by small shrubs and trees in open habitats (e.g., subsection *Chlamydodora*). Because the generation time is known to vary with life form and height in trees [67], substantial variation in the generation time across the *Galoglychia* phylogeny would generate strong uncertainty for results based on coalescent methods. In conclusion, because of too many uncertainties in the inputs required for more subtle models, we assess that the method based on relative branch lengths, although not optimal, is appropriate to discuss lineage sorting with the current data set. Although we are not able to fully exclude it, the incomplete lineage sorting hypothesis thus does not appear a likely explanation for the observed cyto-nuclear discordance.

Host shift mediated introgression as the cause of phylogenetic discordance

Introgressive hybridization, i.e. the introduction of an allele inside the genome of a foreign taxon after hybridization [13], could explain the observed pattern. Introgressive hybridization in *Ficus* necessitates that a pollinating wasp fertilizes two hosts. Hybridization may result from sporadic visitation of an alternative host by a wasp, and may in that case not require successful development of wasp offspring. However, it may also result from transient or even stabilised utilisation of two hosts by a pollinating wasp species, and could result in subsequent speciation of the wasps on the new host. Thus, pollinator taxonomy and phylogeny may be informative to analyse whether introgressive hybridization could account for discrepancies between markers in the *Ficus* phylogeny.

Subsections *Cyathistipulae* and *Crassicostae* present discordant positions between the chloroplast and nuclear phylogenies. Subsection *Cyathistipulae* is strictly associated with genus *Agaon*. The phylogenetic analysis concludes to nuclear monophyly and does not reject

cytoplasmic monophyly. On the other hand subsection *Crassicostae* is pollinated by what could be its own genus of pollinators, *Paragaon* (two species known, and host association only known for one species) but also more frequently by *Elisabethiella* and in one case *Nigeriella*, i.e. genera also associated with subsection *Chlamydodora*e and *Platyphyllae*. Hence this part of the cyto-nuclear mismatch could be associated with repeated pollinator host shifts. Erasmus *et al.* [26] published a phylogeny of wasps pollinating *Ficus* species of section *Galoglychia* that may shed light on the origin of genera *Agaon* (the exclusive pollinators of subsection *Cyathistipulae*) and *Paragaon* (pollinating some species of *Crassicostae*). Based on a combined ITS/28S data set, they found the following topology: (((*Agaon* & *Paragaon*), (*Alfonsiella* & *Nigeriella*), *Elisabethiella*), *Courtella*, *Allotriozoon*). Hence, the cytoplasmic CCCP clade is exclusively pollinated by the *Agaon*, *Paragaon*, *Alfonsiella*, *Nigeriella* and *Elisabethiella* clade. One may therefore suggest that a first host shift of a pollinator from an ancestor of the *Chlamydodora*e to an ancestor of the *Cyathistipulae* and *Crassicostae* may have given rise to the *Agaon-Paragaon* lineage (detailed in Figure 6). This would be the first and major event bringing about the cyto-nuclear discordance. Then we need to hypothesize a second wasp transfer between these two large clades, from *Chlamydodora*e to *Crassicostae*. In this scenario, the pollinators bring with them the cytoplasm of their original host. A simple way to achieve such an a priori unexpected cytoplasm transfer is the stabilised occurrence of a species of fig pollinating wasps using two host species. Such a situation allows backcrosses of hybrids with both parental species. Indeed field observations report several cases of two *Ficus* species being pollinated by the same wasp species [26]. For instance *Elisabethiella stuckenbergii* is known to pollinate *F. burkei*, *F. natalensis natalensis*, *F. lingua depauperata* and *F. petersii* [32], a feature which could potentially allow cytoplasm transfer in any direction independently of which of these *Ficus* species, if any, was the original host of the wasp.

The scenario presented above hinges on the relative phylogenetic position of the different genera pollinating section *Galoglychia*. A recently published pollinator phylogeny [68] proposed that pollinators of subsections *Chlamydodora*e, *Platyphyllae*, *Cyathistipulae* and *Crassicostae* are monophyletic, as in the phylogeny of Erasmus *et al.* [26], and thus does not contradict the scenario proposed above. This new phylogeny did not include *Allotriozoon* and *Paragaon* and did not fully resolve the position of the different genera.

The second type of discordance, concerning the monophyly of subsections, suggests that introgressive hybridization has occurred inside the CCCP clade, with at least two evidenced instances. Both cases concerned species from subsections that are pollinated by several wasp genera. All wasp phylogenies demonstrated the monophyly of wasp genera [26, 68]. Taken together, these observations support the hypothesis that hosts shift could account for several wasp genera pollinating a single *Ficus* subsection. Erasmus *et al.* [26] and Jousselin *et al.* [36] recently suggested that original pollinator species duplication followed by more recent extinction of one or other of the lineages in host figs, rather than host shifts, could explain this unusual association pattern. Beyond evidence for cytoplasmic gene transfers in our phylogenies, there are two reasons that do not favour this hypothesis. First we would need to apply this line of reasoning to subsection *Crassicostae* which is pollinated by three wasp genera (*Paragaon*, *Elisabethiella*, *Nigeriella*). As *Elisabethiella* and *Nigeriella* also pollinate subsection *Chlamydodora*, these genera should then have differentiated into well defined genera very early in section *Galoglychia* history and only much later would one genus have become extinct in some species or even subspecies and survived in other species or even subspecies. The second reason is that Erasmus *et al.* [26] provided direct evidence of ongoing local pollinator exchange between *Ficus* species. Indeed *Alfonsiella binghami* samples from Nelspruit (South Africa), formed a single genetic group whether collected on *F. petersii* or on *F. sthulmannii* comparatively to samples of *A. binghami* also collected on *F. sthulmannii* but in Lekgalameetse (South Africa) and in Amani (Tanzania). Finally, recovering a phylogenetic signal suggesting some amount of parallel cladogenesis is not sufficient to infer that parallel cladogenesis is relevant within these taxonomic groups. Indeed, if host shifts preferentially occur between closely related hosts [69], then the limited amount of evidence suggesting host-pollinator co-cladogenesis in section *Chlamydodora* [36] could easily result from host shifts. Hence the evidence in favour of the occurrence of host shifts within the pollinators of part of section *Galoglychia* is overwhelming.

Fig size: a factor facilitating host shifts?

Although we statistically demonstrated only three instances of cyto-nuclear discordance, it is remarkable that these signs of introgressive hybridization were only found in subsections that are implied in the incongruence between plant and insect phylogeny and taxonomy. If these results strongly supports the hypothesis that host shift is the cause of this incongruence, it also stresses the fact that host shifts have occurred preferentially in some particular lineages of section *Galoglychia*. Hence, host shift could have been facilitated by particular phenotypic

traits in wasps or host plants rather than resulting from purely stochastic processes. When we compare the six subsections of section *Galoglychia*, a striking pattern emerges. All subsections that share pollinator genera and present cytoplasmic introgression also present, on average, small figs and are pollinated by small wasps. Conversely all subsections that have their own chloroplast lineage (or for which cytoplasm monophyly cannot be ruled out) and are pollinated by their own genus of agaonid wasps produce large figs and are pollinated by large wasps. Among pollinators of section *Galoglychia*, large wasp species seem therefore to be more host-specific than small wasp species, a pattern of association that could correlate with a simpler and shorter ostiole passages into small figs of section *Galoglychia*. Indeed, in Gabon, Michaloud [70] showed that *Courtella* sp. pollinating *Ficus ottoniifolia* (subsection *Caulocarpae*) loses its wings and antennae on entering the tight ostiole of the fig, while *Alfonsiella fimbriata* pollinating *F. natalensis leprieuri* (section *Chlamydodora*) only loses its wings and retains its antennae, going through an ostiole that Michaloud describes as a simple slit.

Conclusions

As predicted by the hypothesis that host shifts are responsible for the lack of congruence between insect and plant phylogeny and taxonomy in section *Galoglychia*, we found evidence for several cases of chloroplast genetic introgression events. The alternative to host shifts would be duplication of insect taxa on a host followed by later extinction. Separating these two hypotheses is extremely difficult [69, 71]. However, in our case, because two genetically independent markers, nuclear DNA carried by insects and chloroplasts, tell the same story, the hypothesis of host shifts becomes much more likely. Knowing which lineages are involved in host shifts within a *Ficus* section allowed us to infer that a facilitating factor for host shift could be small fig size. Detecting identical or analogous patterns to those described here in other sections of the genus will constitute the real test of their generality. Finally, while chloroplast DNA presents limited variability in *Ficus* [72], our initial screening of markers allowed increasing the number of polymorphic sites. We were eventually able to resolve deep nodes in the *Ficus* phylogeny for which ITS and ETS nuclear markers were not sufficiently informative. Non-coding chloroplast markers could thus be very helpful to resolve basal branching inside the *Ficus* phylogeny; a necessary step for an in depth analysis of the history of the coevolution between figs and their pollinating wasps.

Material and Methods

Taxon sampling

Forty four specimens belonging to 38 species of *Ficus* out of 77 from the section *Galoglychia* were analyzed in addition to ten species from section *Americana*. *Ficus carica*, *F. ingens*, *F. variegata* and *F. orthoneura* were chosen as outgroups because they belong to distinct sections originating from Asia. This sampling encompasses all six subsections of section *Galoglychia* sensu Berg [31] and confirmed by Rønsted *et al.* [33]. For six species, two specimens from different locations were collected and used for validation. Most of the material consisted of dried leaves collected during field work mainly by FK, from herbaria and from living collections. Forty-one ITS and 33 ETS accessions were retrieved from GenBank (following previously published papers [5, 22, 33]). We chose to acquire our own set of nuclear sequences for species for which available sequences came from individuals of unknown geographic origin and belonging to species with large pan-african distribution ranges, or for species recognized to include several subspecies or morphotypes. This procedure allowed us to be confident that we were comparing chloroplast and nuclear genes of the same taxa. A list of taxa with origin and GenBank accession numbers is provided in additional file S3.

Chloroplast DNA laboratory work

Although chloroplast markers were used in early *Ficus* phylogenetic studies [20], their low level of polymorphism and the associated difficulty to obtain resolved phylogenies down to genus level excluded them from subsequent studies [72]. We therefore developed our own chloroplast primer pairs. We focused on non-coding to select informative sequences. The detailed protocols used during molecular laboratory works are provided in additional file S6. We designed primer pairs in *Ficus* using the Oligo4 software. The Staden Package was used to compile contiguous sequences of each accession and all polymorphic sites were checked against the original electrophoregrams. They were re-examined and adjusted manually using BioEdit [73]. To identify the position of *Ficus* sequences, they were blasted with the completely sequenced cp

genome, phylogenetically closest to *Ficus*, namely that of *Morus indica* [74]. Pairwise *P*-distance and average value with standard error estimated by 500 bootstraps were computed using the MEGA software version 3.1 [75].

Using universal primers developed by Grivet *et al.* [76] in the large single copy (LSC) region of chloroplast genome, we tested 25 amplifying DNA fragments smaller than 3 Kb on a sub-sample of 9 species (additional file S1) representing genus *Ficus* diversity with 6 sections and 3 species within section *Galoglychia*. We also tested two previously described sequences, *atpB-rbcL* [77] and *trnL* (UAA)3'exon-trnF(GAA) [78]. Based on the resulting sequences and their polymorphism, we selected 10 primer pairs and defined new primer pairs to amplify potentially informative sequences shorter than 1 kb (additional file S2). For the 3 fragments longer than 1 kb, we designed 2 to 3 primer pairs focusing on the most informative sequences with 6 to 15 substitutions. These 14 amplified sequences were distributed throughout the LSC chloroplast genome. We examined the amount of phylogenetic information conveyed by each marker on 33 *Ficus* species including 19 from sections *Galoglychia* and *Americana*. Finally, we eventually selected five markers on the basis of the phylogenetic information they conveyed: *atpB-rbcL*, *FcB*, *FcL*, *FcJ* and *trnL-trnF*.

ITS and ETS laboratory work

Based on 100 ITS sequences previously published [5, 22, 33], we designed new primer pairs specifically defined in *Ficus* using the Oligo4 software. The primers ITSF1 (ACAAGGTTTCCGTAGGTGA) and ITSF4 (GTATAGTTATTCGCCTCCT) were defined as forward and reverse primers in 18S and 26S ribosomal RNA genes, respectively. An additional reverse primer ITSF5 (CGGAGGTTACGCTGGGGTC) was defined in the zone at the junction between ITS2 and 26S conferring a high specificity in *Ficus* species. Indeed, ITS amplification using the ITSF1-ITSF5 primer pair was selective and

successful for some samples producing a single, clearly defined, DNA fragment. For the remaining samples, we used the ITS1-ITS4 primer pair that amplified several bands since they were defined on conserved sequences. For samples which produced several amplified DNA fragments, we selected the one presenting the expected size based on previous published ITS sequences [22, 72] by isolating it from the agarose gel following a long electrophoresis migration. We then performed a second PCR on this isolated DNA fragment for sequencing. ETS amplification was performed using the ETS-Hel-1/18S-ETS primer pair [79].

Alignment and phylogenetic analyses

Sequence alignment was performed using the program Clustal X [80] with manual adjustment and excluding ambiguous alignment positions (0.2% and 3.5% of the original chloroplast and nuclear data sets, respectively). All analyses with the chloroplast genes were performed with the five markers combined into a single matrix. This concatenation is not expected to modify the inference of the chloroplast history since this organelle is not known to recombine in wild Angiosperms [81]. For nuclear analyses, we followed previous studies that demonstrated that ITS and ETS sequences could be combined into a single matrix in *Ficus* [5, 22]. In maximum parsimony (MP) analyses, all characters were equally weighted and treated as unordered. Heuristic searches using PAUP* 4b10 [82] were started with 1000 stepwise random addition sequence replicates, holding 10 trees at each step, followed by tree bisection-reconnection branch (TBR) swapping, saving maximally 100 most parsimonious trees. All shortest trees retained in memory were used as starting trees for a second round of searching using TBR branch swapping until all trees were found or a pre-set maximum of 15,000 trees were found. We eventually computed the strict consensus of the most parsimonious trees (SCMP). Relative levels of homoplasy in both nuclear and chloroplast data sets were assessed from all characters using the consistency index (CI) and the retention index (RI) as implemented in PAUP. Bootstrap support was assessed using 1000 replicates each consisting of 10 random addition sequence replicates with TBR swapping and no limits on the number of trees saved.

Bayesian inferences (BI) were performed with several partitioning strategies for each data matrix. Three strategies were tested with the chloroplast matrix: all chloroplast data combined

(P_1 in Table 1), one partition each for the five markers (P_2 - P_6) and one partition each for the coding and non-coding DNA (P_7 - P_8). Two strategies were tested with the nuclear matrix: all chloroplast data combined (P_9) and one partition each for ITS and ETS (P_{10} - P_{11}). The appropriate model of sequence evolution for each partition was determined using the likelihood ratio test (LRT) implemented with MrModeltest [83].

We tested every partitioning strategy with BI with the program MrBayes 3.1.2 [84]. Model parameters were always taken as unlinked and the rate multipliers set as variable across partitions. Initial runs were conducted starting with random neighbour joining trees to check the number of simultaneous Markov Chain Monte Carlo (MCMC) chains necessary not to get caught on local optima. A Dirichlet distribution was assumed for the rate matrix and base frequency and every tree topology was assumed to be equally probable. The MCMC process was set so that two simultaneous independent analyses starting from different random trees with four chains (three heated) ran simultaneously over 4,000,000 generations, and every 200th tree saved into a file. Variation in likelihood scores was examined graphically to determine apparent stationarity for each independent run resulting in a burnin of 10 % trees. MCMC convergence was also explored by examining the Potential Scale Reduction Factor (PSRF) convergence diagnostics for all parameters in the model. Finally, we determined posterior probabilities of the phylogenies and associated branches based on the stationary trees pooled from the two runs for each analysis.

The results for each partitioning strategy were compared to the strategy with the best arithmetic mean of the likelihoods ($-\ln L$; sampled from the posterior) using Bayes factors (see [85] for details). Bayes factors were approximated by the ratio of the harmonic means of the likelihoods of the two strategies being tested [86]. Arithmetic and harmonic means were calculated using the *sump* command of MrBayes. In this study, we used the traditional cutoff criterion of $2 \ln$ Bayes factor of ≥ 10 as very strong evidence against the compared strategy [87].

Maximum likelihood (ML) analyses were performed with RaxML [88] which offers the possibility to partition data. For each data set, we therefore ran analyses using the best partitioning strategy and model of sequence evolution calculated with MrModeltest. We replicated 200 heuristic searches using a randomized maximum parsimony tree as starting tree. The confidence of branching was assessed using 1000 non-parametric bootstrap

resamplings generated as heuristic searches. The information from the 1000 bootstrapped trees was drawn on the best-scoring ML tree from the 200 runs.

Gap characters representing mainly deletions were studied separately from the nucleotide matrices. Gaps were coded for each data set using SeqState [89] with the Simple Indel Coding approach (SIC) [90]. This method scores all gaps, regardless of length, as separate presence/absence characters. Sequences with gaps that extend beyond both the 5'- and 3'-termini of the gap being coded, as well as sequences with gaps that extend beyond one terminus and to the other terminus, are scored as missing data for that character. We used preferentially the SIC approach because it has been demonstrated to outperform the classical method consisting in coding gaps as 5th states for each and every nucleotide position regardless of length [91]. We eventually mapped gaps onto the topologies resulting from the analyses of the respective data sets. Because of limited space along branches, we mapped only synapomorphic gaps.

Analyses of cyto-nuclear discordances

Chloroplast and nuclear phylogenies appeared discordant for several internal and external nodes. The nested pattern of the discordances prevented us from testing their individual significance by using the classical method of removing all discordant taxa and then adding them individually in comparative tests (see [39] for an example). Instead, we used Templeton tests [92] with trees modified from the SCMPs, and Shimodaira-Hasegawa tests (SH; [93]) with trees modified from the ML trees. SH tests were implemented in PAUP, using the REEL approximation with 10,000 bootstrap replicates. Because PAUP cannot account for partitions, models of sequence evolution specified in SH tests were those calculated with all data combined for each data set. Basically, for two phylogenetic hypotheses to be discordant, a topology A must explain significantly better a data set A than a topology B does it, and a topology B must explain significantly better a data set B than a topology A does it. Both conditions need to be full-filled to be able to assess the significance of a given discordance. For example, the global discordance between nuclear and chloroplast topology was tested by comparing the length of the SCMP of the chloroplast trees with the length of the nuclear SCMP, based on the chloroplast data set. The test was then repeated based on the nuclear data set. Similarly, we tested several local discordances individually by comparing a SCMP topology with the same SCMP topology in which only one discordant accession was moved manually in the position as found in the rival topology; each time duplicating the test with

both data sets. Last, we tested discordances with a Bayesian approach [85]. This was achieved by first building 95% credible sets of unique trees (sampled at stationarity) with the *sumt* command in MrBayes for each the chloroplast and nuclear analysis. Then, we used the SumTrees [94] software to check for the presence of alternative phylogenetic hypotheses in the 95% credible set. If they were absent, these hypotheses could be rejected statistically.

Because saturation in substitutions can lead to incorrect phylogenetic inferences [95], we plotted the estimated number of transitions and transversions against the genetic distance using DAMBE V 4.5.32 [96]. In an unsaturated data set, transitions and transversions are both expected to increase linearly with the genetic distance, with a steeper slope for transitions than for transversions. Analyses were conducted with chloroplast, ITS and ETS sequence matrixes separately.

We investigated the hypothesis of amplifying paralogous copies of nuclear genes by assessing the minimum number of gene duplications and extinctions that are required to generate the observed discordant pattern with the help of the GeneTree 1.3.0 software [97, 98]. This procedure was only done for the nuclear genes because gene duplication has not been documented for plastid loci [99]. Because ITS and ETS belong to the same unit of transcription, we did not separate them for this analysis.

Fig and pollinator size

Data on fig and pollinator size were retrieved from Berg & Wiebes [18]. As the data were collected by the same observers, the values were comparable. Dry fig diameter was averaged for each subsection of section *Galoglychia*. We used diameter of dry figs because it was the sole fig-size measure available for all species. We also used body length for each pollinator wasp species. Body length included head, thorax and gaster but excluded the ovipositor. Simple mean comparison tests were performed to compare fig/wasp size between subsections/genera involved in strict association with their partners and subsections/genera involved in multiple associations.

Authors's contribution

JPR, FK and BK jointly designed and coordinated the study. FK carried out the sampling and BK, CG and SS the laboratory analyses. JPR analysed the sequence data sets and wrote the

first draft of the manuscript. FK and BK finalized the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank E. Renoult and M. Charpentier for constructive comments on earlier version of the manuscript. We also thank N. Rønsted who kindly provided nine samples from her studies. JPR was supported by a grant from the French Ministry of Research. The work was supported by the ANR grant NICEFig.

References

1. Losty JP: **Evolution by means of hybridization**. Dordrecht: Nijhoff; 1916.
2. Raven PH: **Systematics and plant population biology**. *Syst. Biol.* 1976, **1**:284-316.
3. Ellstrand NC, Whitkus R, Rieseberg LH: **Distribution of spontaneous plant hybrids**. *Proc. Natl Acad. Sci. USA* 1996, **93**:5090-5093.
4. Grant V: **Plant speciation**, 2nd edn. New-York: Columbia University Press; 1981.
5. Jousselin E, Rasplus JY, Kjellberg F: **Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus***. *Evolution* 2003, **57**:1255-1269.
6. Levin RA, Raguso RA, McDade LA: **Fragrance chemistry and pollinator affinities in Nyctaginaceae**. *Phytochemistry* 2001, **58**:429-440.
7. Pellmyr O: **Yuccas, yucca moths, and coevolution: A review**. *Ann. Missouri Bot. Gard.* 2003, **90**:35-55.
8. Charleston MA, Robertson DL: **Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny**. *Syst. Biol.* 2002, **51**:528-535.
9. Jackson AP: **A reconciliation analysis of host switching in plant-fungal symbioses**. *Evolution* 2004, **58**:1909-1923.
10. Menken SBJ: **Pattern and process in the evolution of insect-plant associations: *Yponomeuta* as an example**. *Entomol. Exp. Appl.* 1996, **80**:297-305.
11. Schoonhoven LM, Jermy T, van Loon JJA: **Insect-plant biology. From physiology to evolution**. London: Chapman and Hall; 1998.

12. Machado CA, Robbins N, Gilbert MTP, Herre EA: **Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism.** *Proc. Natl Acad. Sci. USA* 2005, **102**:6558-6565.
13. Mallet J: **Hybridization as an invasion of the genome.** *Trends Ecol. Evol.* 2005, **20**:229-237.
14. Huysse T, Volckaert FAM: **Comparing host and parasite phylogenies: *Gyrodactylus* flatworms jumping from goby to goby.** *Syst. Biol.* 2005, **54**:710-718.
15. Ricklefs RE, Fallon SM, Bermingham E: **Evolutionary relationships, cospeciation, and host switching in avian malaria parasites.** *Syst. Biol.* 2004, **53**:111-119.
16. Reed DL, Light JE, Allen JM, Kirchman JJ: **Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice.** *BMC Biol.* 2007, **5**.
17. Jackson JA: **Analysis of parasite host-switching: limitations on the use of phylogenies.** *Parasitology* 1999, **119**:S111-S123.
18. Berg CC, Wiebes JT: **African fig trees and fig wasps.** Amsterdam: North Holland; 1992.
19. Wiebes JT: **Agaonidae (Hymenoptera-Chalcidoidea) and *Ficus* (Moraceae) - Fig wasps and their figs .14. (Conclusion - Old-World).** *Proc. Koninklijke Nederl. Akad. Wetens. (C)*1994, **97**:491-495.
20. Herre EA, Machado CA, Bermingham E, Nason JD, Windsor DM, McCafferty SS, VanHouten W, Bachmann K: **Molecular phylogenies of figs and their pollinator wasps.** *J. Biogeog.* 1996, **23**:521-530.
21. Machado CA, Jousselein E, Kjellberg F, Compton SG, Herre EA: **Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps.** *Proc. R. Soc. Lond. Series-B Biol. Sci.* 2001, **268**:685-694.
22. Rønsted N, Weiblen GD, Cook JM, Salamin N, Machado CA, Savolainen V: **60 million years of co-divergence in the fig-wasp symbiosis.** *Proc. R. Soc. Lond. Series-B Biol. Sci.* 2005, **272**:2593-2599.
23. Weiblen GD: **Phylogenetic relationships of fig wasps pollinating functionally dioecious *Ficus* based on mitochondrial DNA sequences and morphology.** *Syst. Biol.* 2001, **50**:243-267.
24. Michaloud G, Michaloudpelletier S, Wiebes JT, Berg CC: **The co-occurrence of two pollinating species of fig wasp and one species of fig.** *Proc. Koninklijke Nederl. Akad. Wetens. (C)* 1985, **88**:93-119.

25. Ramírez WB: **Host specificity of fig wasps (Agaonidae).** *Evolution* 1970, **24**:680-691.
26. Erasmus JC, van Noort S, Jousselein E, Greeff JM: **Molecular phylogeny of fig wasp pollinators (Agaonidae, Hymenoptera) of *Ficus* section *Galoglychia*.** *Zool. Scripta* 2007, **36**:61-78.
27. Lopez-Vaamonde C, Rasplus JY, Weiblen GD, Cook JM: **Molecular phylogenies of fig wasps: Partial cocladogenesis of pollinators and parasites.** *Mol. Phylogenet. Evol.* 2001, **21**:55-71.
28. Molbo D, Machado CA, Sevenster JG, Keller L, Herre EA: **Cryptic species of fig-pollinating wasps: Implications for the evolution of the fig-wasp mutualism, sex allocation, and precision of adaptation.** *Proc. Natl Acad. Sci. USA* 2003, **100**:5867-5872.
29. Rasplus JY: **The one-to-one species specificity of the *Ficus*-Agaoninae mutualism: How casual?** *Biodiv. Afric. Pl.* 1996:639-649.
30. Lopez-Vaamonde C, Dixon DJ, Cook JM, Rasplus JY: **Revision of the Australian species of *Pleistodontes* (Hymenoptera : Agaonidae) fig-pollinating wasps and their host-plant associations.** *Zool. J. Linn. Soc.* 2002, **136**(4):637-683.
31. Berg CC: **Subdivision of *Ficus* subg *Urostigma* sect *Galoglychia* (Moraceae).** *Proc. Koninklijke Nederl. Akad. Wetens. (C)* 1986, **89**(2):121-127.
32. Underhill JG: **Phylogeny and codivergence in the fig-fig wasp mutualism: Sycoecine and Agaonid fig wasps (Chalcidoidea, Hymenoptera) associated with *Ficus* section *Galoglychia* (Moraceae).** Cape Town: Faculty of Science; 2008.
33. Rønsted N, Salvo G, Savolainen V: **Biogeographical and phylogenetic origins of African fig species (*Ficus* section *Galoglychia*).** *Mol. Phylogenet. Evol.* 2007, **43**(1):190-201.
34. Compton SG, Vannoort S: **Southern African fig wasps (Hymenoptera, Chalcidoidea) - Resource utilization and host relationships.** *Proc. Koninklijke Nederl. Akad. Wetens. (C)* 1992, **95**(4):423-435.
35. Cook JM, Rasplus JY: **Mutualists with attitude: Coevolving fig wasps and figs.** *Trends Ecol. Evol.* 2003, **18**(5):241-248.
36. Jousselein E, van Noort S, Berry V, Rasplus JY, Rønsted N, Erasmus JC, Greeff JM: **One fig to bind them all: Host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps.** *Evolution* 2008, **62**(7):1777-1797.

37. Liston A, Kadereit JW: **Chloroplast DNA evidence for introgression and long-distance dispersal in the desert annual *Senecio flavus* (Asteraceae).** *Pl. Syst. Evol.* 1995, **197**(1-4):33-41.
38. Rieseberg LH, Soltis DE: **Phylogenetic consequences of cytoplasmic gene flow in plants.** *Evol. Trends Pl.* 1991, **5**(1):65-84.
39. van der Niet T, Linder HP: **Dealing with incongruence in the quest for the species tree: A case study from the orchid genus *Satyrium*.** *Mol. Phylogenet. Evol.* 2008, **47**(1):154-174.
40. Van Droogenbroeck B, Kyndt T, Romeijn-Peeters E, Van Thuyne W, Goetghebeur P, Romero-Motochi JP, Gheysen G: **Evidence of natural hybridization and introgression between *Vasconcellea* species (Caricaceae) from southern Ecuador revealed by chloroplast, mitochondrial and nuclear DNA markers.** *Ann. Bot.* 2006, **97**(5):793-805.
41. Maddison DR: **The discovery and importance of multiple islands of most-parsimonious trees.** *Syst. Zool.* 1991, **40**(3):315-328.
42. Tamura K, Nei M: **Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees.** *Mol. Biol. Evol.* 1993, **10**(3):512-526.
43. Wolfe KH, Li WH, Sharp PM: **Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.** *Proc. Natl Acad. Sci. USA* 1987, **84**(24):9054-9058.
44. Clegg MT, Gaut BS, Learn GH, Morton BR: **Rates and patterns of chloroplast DNA evolution.** *Proc. Natl Acad. Sci. USA* 1994, **91**(15):6795-6801.
45. Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R *et al*: **Angiosperm phylogeny based on *matK* sequence information.** *Am. J. Bot.* 2003, **90**(12):1758-1776.
46. Schuettpelez E, Korall P, Pryer KM: **Plastid *atpA* data provide improved support for deep relationships among ferns.** *Taxon* 2006, **55**(4):897-906.
47. Mildbraed J, Burret M: **Die Afrikanischen Arten der Gattung *Ficus*.** *Linn. Bot. Jarhb.* 1911, **46**:163-269.
48. Buckler ES, Ippolito A, Holtsford TP: **The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications.** *Genetics* 1997, **145**(3):821-832.
49. Hartmann S, Nason JD, Bhattacharya D: **Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*.** *J. Mol. Evol.* 2001, **53**(2):124-134.

50. Ko KS, Jung HS: **Three nonorthologous ITS1 types are present in a polypore fungus *Trichaptum abietinum***. *Mol. Phylogenet. Evol.* 2002, **23**(2):112-122.
51. Alvarez I, Wendel JF: **Ribosomal ITS sequences and plant phylogenetic inference**. *Mol. Phylogenet. Evol.* 2003, **29**(3):417-434.
52. Dorado O, Rieseberg LH, Arias DM: **Chloroplast DNA introgression in southern California Sunflowers**. *Evolution* 1992, **46**(2):566-572.
53. Kellogg EA, Appels R, MasonGamer RJ: **When genes tell different stories: The diploid genera of Triticeae (Gramineae)**. *Syst. Bot.* 1996, **21**(3):321-347.
54. Natcheva R, Cronberg N: **Recombination and introgression of nuclear and chloroplast genomes between the peat mosses, *Sphagnum capillifolium* and *Sphagnum quinquefarium***. *Mol. Ecol.* 2007, **16**(4):811-818.
55. Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N, Kato M: **Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae)**. *Mol. Biol. Evol.* 2005, **22**(2):285-296.
56. Soltis DE, Kuzoff RK: **Discordance between nuclear and chloroplast phylogenies in the *Heuchera* Group (Saxifragaceae)**. *Evolution* 1995, **49**(4):727-742.
57. Sutton BCS, Flanagan DJ, Gawley JR, Newton CH, Lester DT, Elkassaby YA: **Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones**. *Theor. Appl. Genet.* 1991, **82**(2):242-248.
58. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading**. *Syst. Zool.* 1978, **27**(4):401-410.
59. Kennedy M, Holland BR, Gray RD, Spencer HG: **Untangling long branches: Identifying conflicting phylogenetic signals using spectral analysis, neighbor-net, and consensus networks**. *Syst. Biol.* 2005, **54**(4):620-633.
60. Sanderson MJ, Wojciechowski MF, Hu JM, Khan TS, Brady SG: **Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants**. *Mol. Biol. Evol.* 2000, **17**(5):782-797.
61. Bergsten J: **A review of long-branch attraction**. *Cladistics* 2005, **21**(2):163-193.
62. Pol D, Siddall ME: **Biases in maximum likelihood and parsimony: A simulation approach to a 10-taxon case**. *Cladistics* 2001, **17**(3):266-281.

63. Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO, Rogers JS: **Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods.** *Syst. Biol.* 2001, **50**(4):525-539.
64. Holder MT, Anderson JA, Holloway AK: **Difficulties in detecting hybridization.** *Syst. Biol.* 2001, **50**(6):978-982.
65. Pamilo P, Nei M: **Relationships between gene trees and species trees.** *Mol. Biol. Evol.* 1988, **5**(5):568-583.
66. Buckley TR, Cordeiro M, Marshall DC, Simon C: **Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada Dugdale*).** *Syst. Biol.* 2006, **55**(3):411-425.
67. King DA: **Allometry and life history of tropical trees.** *J. Trop. Ecol.* 1996, **12**:25-44.
68. Lopez-Vaamonde C, Wikström N, Kjer KM, Weiblen GD, Rasplus JY, Machado CA, Cook JM: **Molecular dating and biogeography of fig-pollinating wasps.** *Mol. Phylogenet. Evol.* 2009, doi:10.1016/j.ympev.2009.05.028.
69. De Vienne DM, Giraud T, Shykoff JA: **When can host shifts produce congruent host and parasite phylogenies? A simulation approach.** *J. Evol. Biol.* 2007, **20**(4):1428-1438.
70. Michaloud G: **Figuiers tropicaux et pollinisation** (Motion picture film). In. Edited by Service du Film de Recherche Scientifique P, France; 1982.
71. Light JE, Hafner MS: **Codivergence in heteromyid rodents (Rodentia : Heteromyidae) and their sucking lice of the genus *Fahrenholzia* (Phthiraptera : Anoplura).** *Syst. Biol.* 2008, **57**(3):449-465.
72. Rønsted N, Yektaei-Karin E, Turk K, Clarkson JM, Chase MW: **Species-level phylogenetics of large genera: prospects of studying co-evolution and polyploidy.** In: *Reconstructing the tree of life: taxonomy and systematics of species rich taxa.* Edited by Hodkinson T, Parnell J, Waldren S. Boca Raton, Florida: Systematic Association Series. CRC Press; 2007.
73. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl. Ac. Symp. Ser.* 1999, **41**:95-98.
74. Ravi V, Khurana JP, Tyagi AK, Khurana P: **The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis.** *Tree Genet. Genom.* 2007, **3**(1):49-59.

75. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment.** *Brief. Bioinf.* 2004, **5**(2):150-163.
76. Grivet D, Heinze B, Vendramin GG, Petit RJ: **Genome walking with consensus primers: application to the large single copy region of chloroplast DNA.** *Mol. Ecol. Notes* 2001, **1**(4):345-349.
77. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH: **Use of DNA barcodes to identify flowering plants.** *Proc. Natl Acad. Sci. USA* 2005, **102**(23):8369-8374.
78. Taberlet P, Gielly L, Pautou G, Bouvet J: **Universal primers for amplification of 3 noncoding regions of chloroplast DNA.** *Pl. Mol. Biol.* 1991, **17**(5):1105-1109.
79. Baldwin BG, Markos S: **Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae).** *Mol. Phylogenet. Evol.* 1998, **10**(3):449-463.
80. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucl. Ac. Res.* 1997, **25**(24):4876-4882.
81. Trabelsi S, Gargouri-Bouزيد R, Vedel F, Nato A, Lakhoua L, Drira N: **Somatic hybrids between potato *Solanum tuberosum* and wild species *Solanum vernei* exhibit a recombination in the plastome.** *Plant Cell Tiss. Org. Cult.* 2005, **83**(1):1-11.
82. Swofford DL: **PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0b4.** In.: Sunderland, MA: Sinauer Associates; 2000.
83. Nylander JAA: **MrModeltest v2. Program distributed by the author.** In. Uppsala University: Evolutionary Biology Centre; 2004.
84. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**(8):754-755.
85. Brandley MC, Schmitz A, Reeder TW: **Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards.** *Syst. Biol.* 2005, **54**(3):373-390.
86. Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL: **Bayesian phylogenetic analysis of combined data.** *Syst. Biol.* 2004, **53**(1):47-67.
87. Kass RE, Raftery AE: **Bayes Factors.** *J. Am. Stat. Assoc.* 1995, **90**(430):773-795.

88. Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**(21):2688-2690.
89. Müller K: **SeqState - primer design and sequence statistics for phylogenetic DNA data sets.** *Appl. Bioinf.* 2005, **4**:65-69.
90. Simmons MP, Ochoterena H: **Gaps as characters in sequence-based phylogenetic analyses.** *Syst. Biol.* 2000, **49**(2):369-381.
91. Simmons MP, Muller K, Norton AP: **The relative performance of indel-coding methods in simulations.** *Mol. Phylogenet. Evol.* 2007, **44**(2):724-740.
92. Templeton AR: **Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes.** *Evolution* 1983, **37**(2):221-244.
93. Shimodaira H, Hasegawa M: **Multiple comparisons of log-likelihoods with applications to phylogenetic inference.** *Mol. Biol. Evol.* 1999, **16**(8):1114-1116.
94. **SumTrees: Summarization of split support on phylogenetic trees. Version 1.0.2. Part of the DendroPy phylogenetic computation library version 2.1.3**
[<http://sourceforge.net/projects/dendropy>]
95. Swofford DL, Olsen G, Waddell PJ, Hillis DM: **Phylogenetic inference in molecular systematics.** In. Sunderland: Sinauer Associates Inc.; 1996: 407-514.
96. Xia X, Xie Z: **DAMBE: Software package for data analysis in molecular biology and evolution** *J. Hered.* 2001, **92**(4):371-373.
97. Page RDM: **GeneTree: comparing gene and species phylogenies using reconciled trees.** *Bioinformatics* 1998, **14**(9):819-820.
98. Page RDM, Charleston MA: **From gene to organismal phylogeny: Reconciled trees and the gene tree species tree problem.** *Mol. Phylogenet. Evol.* 1997, **7**(2):231-240.
99. Soltis DE, Soltis PS, Doyle JJ: **Molecular systematics of plants II.** Boston: Kluwer Academic Publishers; 1998.

Figure 1. Associations between subsections of section *Galoglychia* in *Ficus* (left) and their genera of Agaonidae pollinating wasps (right). Numbers of taxa described for each wasp genus and each fig subsection are given within brackets. Linking lines inform on pollination associations between wasp genera and host subsections, with, for each link, the number of fig species known to be pollinated by a given wasp genus provided within circles. Colours highlight major clades to stress the discrepancy between plant (from [33]) and wasp (from [26]) phylogenies.

Figure 2. Strict consensus trees of all most parsimonious trees. Analyses performed with (A) the chloroplast data set (length = 141 steps) and (B) the nuclear ITS+ETS data set (length = 521 steps). Values above branches indicate bootstrap supports.

Figure 3. ITS+ETS (left) and chloroplast (right) phylogenetic trees obtained with the Bayesian Inference (BI) analyses. Values above and under branches indicate posterior probability of BI and bootstrap support performed with the Maximum Likelihood analysis, respectively. Synapomorphic indels of each data set are mapped and numbered onto the corresponding tree (details about indels are given in additional file S4). Genera of wasps known to pollinate a subsection are indicated between brackets.

Figure 4. Mean dry fig diameter and mean pollinator body length for each subsection of section *Galoglychia* and each pollinating wasp genus. Figures above whiskers give the number (n) of species included in the mean calculation. Whiskers give 95% credible interval (provided when $n > 3$, otherwise single values are plotted).

Figure 5. The cyto-nuclear discordance explained by ITS and ETS duplication-extinction. The scenario requires a basal duplication (D) followed by three extinction events: one extinction of allele 1 (E1) and two extinctions of allele 2 (E2).

Figure 6. Comparative phylogeny and the host shift scenario. If we accept current phylogenies, then a pollinating wasp of an ancestor of *Chlamydodorae* and *Platyphyllae* must have shifted to an ancestor of *Cyathistipulae* and *Crassicostae*, an event required to explain the discrepancy between ITS+ETS (left) and pollinator (right) phylogenies. However, under the hypothesis of a wasp suddenly changing host (grey arrows), there is no cytoplasm transfer and the *Ficus* lineage resulting from the hybridization necessary bears a chloroplast related to

the *Caulocarpae* and *Galoglychia* lineages (middle bottom), a feature which is not observed in the cytoplasm phylogeny. On the contrary, if a wasp uses two hosts for a prolonged period of time (black arrows), repeated backcrosses may allow the introduction of a chloroplast originating in the *Chlamydodora* plus *Platyphyllae* lineage inside the ancestor of *Cyathistipulae* and *Crassicostae* (middle top), as observed in the data set. The set of phylogenetic trees therefore support the role of an intermediate stage in which a pollinating wasp has consistently used two host species.

Table 1. Data partitions, genome to which they are associated, total number of characters of each partition used in phylogenetic analyses and their estimated model of sequence evolution.

Partition	Genome	Number of characters	Selected Model
P_1 : All chloroplast data combined	Chloroplast	3604	GTR+I+ Γ
P_2 : <i>atpB-rbcL</i>	Chloroplast	820	HKY+ Γ
P_3 : <i>FcB</i>	Chloroplast	779	GTR
P_4 : <i>FcJ</i>	Chloroplast	760	HKY
P_5 : <i>FcL</i>	Chloroplast	777	GTR+ Γ
P_6 : <i>trnL-trnF</i>	Chloroplast	468	HKY
P_7 : Coding chloroplast DNA	Chloroplast	964	HKY
P_8 : Non-coding chloroplast DNA	Chloroplast	2640	GTR+ Γ
P_9 : All nuclear data combined	Nuclear	1320	GTR+ Γ
P_{10} : <i>ITS</i>	Nuclear	810	GTR+ Γ
P_{11} : <i>ETS</i>	Nuclear	510	HKY+ Γ

Table 2. Arithmetic mean $-\ln L$ for each partition strategy (S_i) and $2\ln$ Bayes factors results of comparisons between the strategies with the highest likelihood and alternative strategies.

Partition strategy	Mean $-\ln L$	$2\ln$ Bayes factor
$S_1: P_1$	6158.25	S_1 vs S_3 : 214.3
$S_2: P_2+P_3+P_4+P_5+P_6$	6030.59	S_2 vs S_3 : 11.38
$S_3: P_7+P_8$	6028.45	
$S_4: P_9$	5584.63	S_4 vs S_5 : 21.06
$S_5: P_{10}+P_{11}$	5578.51	

Table 3. Results of Templeton tests with most parsimonious (MP) trees and SH tests with maximum likelihood (ML) trees. *P*-values of a given line correspond to the test comparing the length (MP trees) or the likelihood (ML trees) of the topology of the corresponding line with the length/likelihood of the above topology written in bold characters. In the topology description, “,” indicates that branching between subsections are left unresolved, and “+” indicates that the branching among and between subsections are left unresolved. Likelihoods were computed with PAUP.

Topology	Templeton test		SH test	
	Length	<i>P</i> -Value	-LnL	<i>P</i> -Value
<i>Chloroplast Data</i>				
Chloroplast	143		5956.50	
Nuclear	195	<0.001	6241.39	<0.001
chloroplast - <i>F. trichopoda</i> monophyletic	147	0.045	5981.30	0.021
chloroplast - <i>F. craterostoma</i> monophyletic	152	0.003	5999.10	0.001
chloroplast - <i>F. densistipulata</i> within <i>Cyathistipulae</i>	144	0.564	5961.60	0.280
(Americana,(Galoglychia,Caulocarpae,(Cyathistipulae +Crassicostae+Chlamydodora+Platyphyllae))	226		6061.80	
(Americana,(Galoglychia+Caulocarpae+Cyathistipulae +Crassicostae),(Chlamydodora+Platyphyllae))	276	0.016	6340.26	0.005
<i>Nuclear Data</i>				
Nuclear	539		5170.68	
Chloroplast	741	<0.001	6091.60	<0.001
nuclear - <i>F. trichopoda</i> 2 within <i>Chlamydodora</i>	560	<0.001	5259.70	0.000
nuclear - <i>F. craterostoma</i> 2 within <i>Cyathistipulae</i> and <i>Crassicostae</i>	560	<0.001	5211.23	0.001
nuclear - <i>F. densistipulata</i> within <i>Chlamydodora</i>	559	<0.001	5256.40	<0.001
(Americana,(Galoglychia+Caulocarpae+Cyathistipulae +Crassicostae),(Chlamydodora+Platyphyllae))	818		5407.92	
(Americana,(Galoglychia,Caulocarpae,(Cyathistipulae +Crassicostae+Chlamydodora+Platyphyllae))	962	0.001	5890.20	0.002

Additional files

Additional file S1. List of species used for the development of non coding chloroplast DNA markers for *Ficus*. Format DOC.

Additional file S2. Description of the new chloroplast primer pairs specifically defined in *Ficus*. Format DOC.

Additional file S3. List of samples included in the final analyses with origin and GenBank accession numbers. Format XLS.

Additional file S4. Insertions and deletions in chloroplast and nuclear markers. Format DOC.

Additional file S5. Substitution pattern of the chloroplast markers, ITS and ETS genes. Format PDF.

Additional file S6. Protocols for molecular laboratory works. Format DOC.

**Systematics of the *Podarcis hispanicus*-complex (Sauria,
Lacertidae) II: the valid name of the north-eastern Spanish form**

JULIEN P. RENOULT¹, PHILIPPE GENIEZ², PAUL BACQUET¹, CLAUDE P.
GUILLAUME², PIERRE-ANDRE CROCHET^{1,3}

¹ CNRS-UMR 5175, Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende,
34293 Montpellier cedex 5, France

² EPHE-UMR 5175, Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende,
34293 Montpellier cedex 5, France

³ E-mail: pierre-andre.crochet@cefe.cnrs.fr

Running head: Nomenclature of north-eastern Iberian *Podarcis*

Corresponding author: Pierre-André Crochet

Abstract

Recent genetic works have suggested that the Iberian wall lizard *Podarcis hispanicus* (Steindachner, 1870) sensu lato is in fact a species complex. Several forms have already been elevated to species rank, including the north-eastern Spanish form actually named *Podarcis atratus* (Bosca, 1916). However, we show here that lizards from Valencia are conspecific with the north-eastern populations in spite of them carrying different mitochondrial lineages and that consequently the nomen *Lacerta muralis* var. *liolepis* Boulenger, 1905 has precedence over *atratus* to designate the north-eastern species of the *hispanicus* complex.

[cyto-nuclear discordance, introgression, nomenclature, Wall Lizard, Spain, Reptilia]

Introduction

Recent genetic and morphological data indicate that *Podarcis hispanicus* (Steindachner, 1870) as traditionally understood (see for example Pleguezuelos *et al.* 2003) is actually composed of several genetically distinct lineages, most of which deserve a specific rank (Oliverio *et al.* 2000; Sá-Sousa 2000; Harris and Sá-Sousa 2001, 2002; Harris *et al.* 2002a, 2002b; Sá-Sousa *et al.* 2002, Busack *et al.* 2005; Pinho *et al.* 2006; Pinho *et al.* 2007, Pinho *et al.* 2008, Renoult *et al.* submitted). Few formal systematic changes have been proposed to date, however.

The insular populations from the Columbretes islands have been the first to be formally raised to species rank under the binomen *Podarcis atratus* (Boscá, 1916) (as *Podarcis atrata*, but see Böhme and Köhler 2005 for the gender of *Podarcis*) on the basis of their large amount of genetic divergence (mitochondrial DNA sequences) compared to specimens from the mainland (Castilla *et al.* 1998a, b). Later, the North African and South Iberian taxon *vaucheri* (Boulenger, 1905) was raised to species rank by Oliverio *et al.* (2000) and Busack *et al.* (2005). In the same publication, Busack *et al.* (2005) proposed to treat the north-eastern Spanish form (= “*P. hispanica* morphotype 3” in Pinho *et al.* 2006) as a distinct species under the nomen *Podarcis atratus*, arguing that the Columbretes populations are conspecific to the mainland specimens of the north-eastern Spanish form.

We do not question that the Columbretes populations are conspecific with the north-eastern Spanish form, a conclusion that was already suggested by Geniez (2001) on morphological grounds, but we wish to stress that the oldest available nomen for the north-eastern Spanish population is *Lacerta muralis* var. *liolepis* Boulenger, 1905, as already pointed out by Guillaume (1987) and Geniez (2001). This name is not in concurrence with *atratus* to designate a taxon limited to the Columbretes islands, but it has precedence over *atratus* if the north-eastern Iberian populations and the Columbretes taxon are united in the same species, as proposed by Busack *et al.* (2005) and as we think is correct.

The north-eastern Spanish form has been genetically characterised based on specimens from Barcelona (Harris and Sá-Sousa 2002; Pinho *et al.* 2003), Burgos and Medinaceli (Castilla y León) (Pinho *et al.* 2006), Girona, Tarragona and the central southern Pyrenees (Harris and Sá-Sousa 2002), and Andorra (Harris *et al.* 2002a), and corresponds to the mitochondrial lineage named “*Podarcis hispanicus* type 3” (Pinho *et al.* 2006). The syntypes of *Lacerta muralis* var. *liolepis* Boulenger, 1905 originate mainly from Valencia and surroundings (see below), an area where *Podarcis* lizards are characterised by a different mitochondrial lineage, named “*Podarcis hispanica sensu stricto*” in Pinho *et al.* (2006).

Although these two mitochondrial lineages diverged several million years ago (Pinho *et al.* 2006), whether they correspond to two distinct taxonomic units is still unclear. Indeed, based on a study examining allozyme variation, Pinho *et al.* (2007) showed that specimens from both lineages were grouped into a single nuclear cluster although this result was later contradicted by the finding of a null gene flow between the two lineages, based on the sequences of two nuclear introns (Pinho *et al.* 2008). Recently, Renoult *et al.* (submitted) demonstrated that the various populations bearing the mitochondrial lineage “*P. hispanicus sensu stricto*” belong to several distinct taxonomic units and that the populations located in the province of Valencia belong to the same group as “*P. hispanicus* type 3” based on morphological and microsatellites data.

The aim of this article is to expand on the analyses of Renoult *et al.* (submitted) using recently available nuclear introns sequences, focusing on the relationships of the populations of the Valencia province with typical “*P. hispanicus* type 3” populations. We were not able to conduct a genetic analyse of the syntypes of *liolepis* but we compared them morphologically with recent specimens from the type locality. We also compared Valencia syntypes with specimens from the other Iberian members of the *P. hispanicus* complex. Based on these findings, we formally allocate the nomen *liolepis* to the north-eastern Spanish form of the *P. hispanicus* complex.

Material and methods

Morphology: material examined, specimens identification and abbreviations

Morphological comparisons are based on specimens coming from a large database comprising more than 2700 specimens of the *Podarcis hispanicus* complex originating from 270 Spanish, 54 Portuguese and 59 French localities. Most specimens belong to the former collection of the “Biogéographie et Ecologie des Vertébrés” laboratory (BEV) now housed at the Centre d’Ecologie Fonctionnelle et Evolutive (Montpellier, France). Others were borrowed from the Natural History Museum, London, UK (BMNH); Museu Bocage, Lisbon, Portugal (MBL); private collection of Dr Charles P. Blanc (Montpellier, France: CPB); Estación Biológica de Doñana, Seville, Spain (EBD); Naturhistorisches Museum, Bâle, Switzerland (NHMB); Muséum national d’Histoire naturelle, Paris, France (MNHN); Naturhistorisches Museum Wien, Austria (NMW); private photographic collection of Philippe Geniez, Vendargues, France (PGe). In the following text “the Code” designates the current edition of the International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature, 1999).

In this study, we first excluded from this database specimens with incomplete data, and selected only adult males, as morphological separation of the taxa in the *Podarcis hispanicus* complex is usually more efficient on adult males (pers. obs.). In addition, we excluded specimens that come from vague localities or localities that we were unable to locate. We then selected among these animals a series of reference specimens for each lineage of the *Podarcis hispanicus* complex currently recognised in the Iberian Peninsula (with the exception of the well defined species *P. bocagei* and *P. carbonelli*): *Podarcis hispanicus* sensu stricto (as redefined in Geniez *et al.* 2007; = *hispanicus* hereafter), *Podarcis vaucheri* (Iberian specimens only; = *vaucheri* hereafter), and *Podarcis hispanicus* type 1 (=TYP1), type 2 (= TYP2) and type 3 (= north-eastern Spanish form = *atratus*, sensu Busack *et al.* 2005: = TYP3 hereafter).

For *hispanicus* specimens, we retained all those identified as such in Geniez *et al.* (2007) that fulfilled the above criteria. For the other lineages, we selected specimens that fulfilled two criteria: they had to come from localities within the known range of each mitochondrial lineage either based on published genetic data or on our additional genetic data gathered during this study (see below) and to morphologically conform to the characters defined by Geniez (2001) for the corresponding morphotypes. Using these two criteria, assignation of many specimens to one of the above lineages was straightforward. In addition, a few misclassified individuals would not alter our conclusions. After this selection, a total of 652 reference specimens were analysed: 97 *hispanicus*, 160 *vaucheri*, 126 TYP1, 69 TYP2 and 200 TYP3. TYP3 specimens were further divided into two subtypes corresponding to the two morphological types identified by Guillaume and Geniez (1986), Guillaume (1987) and Geniez (2001): “liolepis” and “robusta”. The “liolepis” morphotype corresponds to specimens from the eastern part of the range of TYP3 (the Ebro basin and adjacent areas) and the “robusta” morphotype from the western part of the range (the Duero basin, around Zamora, Leon, Burgos).

In addition, the database includes nine of the syntypes of *liolepis* (= syntypes hereafter) and seven recent specimens from the town of Valencia but not part of the type series (= “Valencia” specimens hereafter). These specimens need to be compared with the reference specimens to determine their relationships.

Morphology: analysed characters

Three categories of characters were measured: (i) quantitative morphometric variables (*SVL* = snout-vent length; *HeNe* = head length + neck length from the tip of the snout to the

insertion of the forelimb; *PiLL* = pileus length; *HeH* = head height), (ii) quantitative pholidosis variables (*Dors* = number of longitudinal rows of dorsal scales at mid-body; *Venl* = number of transversal rows of ventral plates; *FPor* = number of femoral pores [mean of left and right side]; *Lame* = number of enlarged scales [infradigital lamellae] under the 4th toe [left side]; *ID10* = mean size of the masseteric shields of the two sides as computed from Guillaume (1988) and (iii) semi-quantitative variables describing colour pattern not modified in the preserved specimens (*Vert* = prominence of the dark vertebral line [0 = absent, 1 = on part of the back only, 2 = complete, broken or unbroken]; *Bif* = bifurcation of anterior part of the vertebral line [0 = no, 1 = yes]; *DoLa* = prominence of the pale dorsolateral stripes [0 = absent, 1 = weakly prominent, 2 = strongly prominent]; *Frag* = fragmentation of the pale dorsolateral stripes [0 = no, 1 = yes]; *SDLa* = prominence of the dark supradorsolateral stripes [0 = absent, 1 = less than 4 scales wide, 2 = 4 scales wide or more]; *Pari* = pale dorsolateral stripes extend on the parietal plates [0 = no, 1 = weakly, 2 = strongly]; *PilP* = prominence of dark pigmentation of the pileus [1 = no dark spots, 2 = thin dark spotting, 3 = well marked dark dots].

Morphology: data analyses

We used discriminant function analysis based on the 16 variables described above to (i) separate morphologically the specimens of the various lineages and (ii) investigate if the *liolepis* syntypes could be morphologically assigned to one of these lineages. These specimens were therefore treated as supplementary (non-active) individual: their position along the discriminant axis was calculated after the analysis but they did not participate in the calculation of the axis. Their position was therefore not influenced by any *a priori* assignation. We also used Principal Component Analysis (PCA) based on the same 16 variables to visually represent the morphological variability inside the TYP3, and between TYP3, syntypes and Valencia specimens.

Genetic data: sampled specimens

All genetic samples were linked to a voucher specimen in the BEV collection. For the study of mitochondrial DNA polymorphism, we selected 44 reference specimens of *hispanicus*, TYP2 and TYP3 based on their morphology and range as explained for the morphological data. Since we could successfully amplify DNA from only three specimens from Valencia and surroundings, we included also in the analysis 22 specimens originating within a radius of 200 km around Valencia and having the TYP3 morphology. They constitute

a further category of specimens, called “southern TYP3” hereafter. We supposed that they belong to the same taxon as the specimens from Valencia, but this assumption was checked by morphology and nuclear DNA analyses (see results). Eventually, we included 14 sequences from Pinho *et al.* (2006) to represent each of the major clades known in the *Podarcis* genus in Iberia (Pinho *et al.* 2006)

For nuclear DNA analyses, we genotyped 104 individuals belonging to *hispanicus*, TYP2, reference TYP3 and southern TYP3. Details about the specimens sequenced for the mitochondrial gene, the GenBank accession numbers retrieved from Pinho *et al.* (2006) and the localities of the genotyped individuals are provided in Renoult *et al.* (submitted). Last, we compared sequence polymorphism of 24 reference TYP3 with 26 southern TYP3 for the β fib intron, and 26 reference TYP3 with 26 southern TYP3 for Pgd intron.

Genetic data: extraction, amplification and typing

Total cellular DNA was extracted using DNeasy tissue kit (Qiagen, Santa Clarita, CA) following the supplier’s procedure. For the mitochondrial study, we amplified and sequenced a 460 bp segment including parts of the central and right domain of the control region (CR) previously described in Crochet *et al.* (2004). PCR amplification and sequencing conditions are provided in Renoult *et al.* (submitted).

We amplified β fib fragments of 575 bp length with the primers BF8 (F: 5’- CAC CAC CGT CTT CTT TGG AAC ACT G - 3’) and β fib (R: 5’- CAG GGA GAG CTA CTT TTG ATT AGA C - 3’) from Pinho *et al.* (2008). PCR was carried out in 20 μ L volumes containing 2X reaction buffer (100mM Tris-HCl + 500mM KCl), 1.6 mM MgCl₂, 1.6 mM dNTP, 0.4 μ M each primer, 0.2 U Taq DNA polymerase (Goldstar, Eurogentec) and 3 μ L of DNA solution. PCR conditions consisted of a pre-denaturing step of 3 min at 92°C followed by 35 cycles of a denaturing step at 92°C for 30 s, annealing at 54°C for 30 s and extension at 72°C for 40 s with a final extension at 72°C for 4 min.

For Pgd fragments, amplifications were performed with the following conditions: 20 μ l of total reaction volume containing 1.5 μ l DNA solution (variable concentration), 2.5x buffer (Tris-HCL 100 mM + KCl 500mM), 3.2 mM MgCl₂, 2 mM dNTP at 20 μ M, 0.5 μ M of each primer and 0.2 unit Taq DNA polymerase Goldstar, Eurogentec. Primers used to amplify the Pgd intron were PgdP7F, 5’ GACATGCAGCTGATCTGTGAGGCC and PgdP8R, 5’ GAGTCCAGCTCAGTCTT-ATTCCAC. The following conditions were used for PCR: pre-denaturing step of 3 min at 92°C followed by 35 cycles of a denaturing step at 92°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 90 s with a final extension at 72°C for 5

min. All sequencing reactions for nuclear products were performed by AGOWA (www.agowa.de).

For the microsatellites data, we genotyped eight (Pb10, Pb11, Pb20, Pb37, Pb47, Pb50, Pb55 and Pb73) out of the nine polymorphic microsatellite loci developed by Pinho *et al.* (2004). For multiplexing convenience, we excluded Pb66 from our analysis. Details on the conditions of the amplification are provided in Renoult *et al.* (submitted).

Genetic data: data analysis

Sequences were compiled and aligned manually using Mega v.3.1 (Kumar *et al.* 2004). Phase resolution (i.e., the determination of sequences for both alleles) for multiples heterozygotes was performed with PHASE v.2.1.1 (Stephens *et al.* 2001). Three independent phylogenetic reconstruction analyses were performed with the three data sets (CR, β fib and Pgd). We did not concatenate sequences from the two nuclear genes into a single matrix because we were more interested in generating two independent results informing on the evolution of the nuclear genome rather than a single one. Also, because we were not interested in phylogenetic relationships among lineages, we limited our phylogenetic analyses to a crude method (neighbour-joining (NJ) trees with 100 bootstrap replications).

Details of microsatellites analyses are provided in Renoult *et al.* (submitted). Briefly, we analysed microsatellites with the Geneland software Guillot *et al.* (2005), a clustering method based on a Bayesian model that assigns individuals probabilistically to clusters without *a priori* knowledge of population units and limits. The software allows obtaining two results: the number of clusters having the highest posterior probability to occur in the studied area and the posterior probability for each individual to belong to the defined clusters. Parameters used for running the software are provided in Renoult *et al.* (submitted).

Results

The type specimens of Lacerta muralis var. liolepis Boulenger, 1905

The central argument developed here is that the majority of the name-bearing types of *liolepis* belong to the north-eastern Spanish form. It is therefore crucial to accurately establish which specimens constitute the name-bearing types of *liolepis*.

Boulenger (1905) did not list precisely the specimens used for his description but he based it partly on the description made by Bedriaga (1879) of several specimens originating from Valencia. These specimens are thus part of the type series (see below). He also assigned

to this form “numerous specimens from Valencia in the British Museum” collected by Lilford, Boscá and Bolivar and specimens from Seville received from Prof. Calderon. C. J. McCarthy (Natural History Museum, London) examined for us the catalogues of the Natural History Museum and together we were able to establish that the following specimens currently in this collection are part of the original type series: BMNH 1946.9.1.33-35 (formerly 76.2.28.6-8), Valencia, Spain / Lord Lilford; BMNH 86.12.29.1-10 A-H, Valencia, Spain / Prof. E. Boscá; BMNH 1946.9.1.36-45 (formerly 1905.2.27.5-14), Valencia, Spain / Prof. E. Boscá (Madrid museum); BMNH 1920.1.20.321, Valencia, Spain / Prof. E. Boscá (Lataste collection); BMNH 1905.2.27.15-22, Foyos [= Foios], near Valencia / Prof. E. Boscá (Madrid Museum); BMNH 94.6.1.15-17, Sevilla / Prof. Calderon.

Bedriaga (1879) does not give a list of the specimens he examined. He explicitly refers to several specimens he saw along the road from Valencia to Dehesa de la Albufera, several of which were collected by himself but died and got rotten before he could measure them, except for one specimen whose measurements are given in his paper. Despite numerous enquiries in various collections, we were unable to locate this specimen or any other *Podarcis* specimen collected by Bedriaga near Valencia. The Bedriaga syntypes are thus probably lost.

Because Valencia and Sevilla are inhabited by different taxa of the *hispanicus* complex (see Guillaume 1987; Geniez 2001; Harris *et al.* 2002a; Harris and Sá-Sousa 2002, Pinho *et al.* 2006, Pinho *et al.* 2007, Pinho *et al.* 2008), the name *liolepis* could potentially apply to either of these taxa. Nevertheless, the types of *liolepis* include at least 39 specimens from Valencia and its surroundings but only three specimens from Sevilla. Furthermore, Boulenger (1905) starts the description of *liolepis* by referring to specimens seen at Valencia by Bedriaga, and in subsequent publications (Boulenger 1921, pp. 402-403, in the appendix) he only lists specimens from Valencia as types. It is thus obvious that Boulenger intended to use the name *liolepis* for the taxon occurring in and around Valencia. This has been followed by Salvador (1986), Guillaume (1987) and Geniez (2001) although none of these authors realised that the type locality actually encompassed Valencia, Foios and Sevilla.

As the Sevilla syntypes are probably *Podarcis vaucheri* (the species currently inhabiting Sevilla and its surroundings) it is necessary to designate a lectotype from the Valencia syntypes to permanently fix the use of the nomen *liolepis*. Among the original syntypes we select here the specimen BMNH 86.12.29.3 (fig. 5A) from Valencia as lectotype of *Lacerta muralis* var. *liolepis* Boulenger, 1905 (see systematic section below).

Which taxon of the Podarcis hispanicus complex occurs in Valencia?

Morphology: assignation of the syntypes and Valencia specimens

Figure 1 shows the results of the discriminant analyses with modern specimens from Valencia and *liolepis* syntypes from Valencia treated as non-active individuals. In these figures, all lineages differ substantially in morphology, but there is a variable amount of overlap between all morphotypes that were analysed together. Usually, this overlap is due to a few outliers, with most individuals falling outside the bivariate distribution of the vast majority of the specimens of the other morphotypes. The lineage that is the most difficult to characterise morphologically seems to be *vaucheri*, whose outliers fall far in the distribution of the TYP1 and TYP3 specimens (Fig. 1A). Valencia specimens and syntypes fall all outside the distribution of TYP1 specimens (Fig. 1A) and all but one (not a syntype) fall outside the distribution of *hispanicus* specimens (Fig. 1B). Most of Valencia individuals and syntypes also fall outside the distribution of TYP2 (Fig. 1B). Valencia and syntypes are less morphologically discriminated from *vaucheri* specimens, nevertheless some of these specimens fall outside the distribution of *vaucheri*, and the others are outside the region of the axis 1 x axis 2 plan with a high density of *vaucheri* specimens: they fall instead in areas where only a few outliers of *vaucheri* can be found (Fig. 1A). On the contrary, virtually all (bare one in Fig. 1B) specimens from Valencia or from the type series have a morphology which places them fully within the distribution of TYP3 specimens.

The Valencia and syntypes sample thus falls morphologically clearly outside the variation of *hispanicus*, TYP1 and TYP2 lineages. Its separation from *vaucheri* (based on the characters we analysed) is less clear cut but most specimens are obviously outside the morphological variation of this form. On the contrary, it falls well within the typical variation of the TYP3 lineage.

Mitochondrial DNA relationships of the “Valencia” and “southern TYP3” specimens

Figure 2A shows the phylogenetic relationships of the control region sequences of our specimens together with GenBank accessions deposited by Pinho *et al.* (2006) to allow assignation of our haplotypes to previously identified lineages. It is noteworthy that the tree results from the analysis of a single marker and that the relationships among the main clades are neither supported nor informative. Nevertheless, this analysis identifies several terminal clades with strong support values corresponding to the lineages previously identified within the *Podarcis hispanicus* complex (see for example Pinho *et al.* 2006).

As expected, all our reference TYP3 specimens (including “*liolepis*” and “*robusta*” morphotypes) group with Pinho *et al.*’s (2006) CR sequences of the “*P. hispanicus* type 3”

lineage. Similarly, our TYP2 specimens fall into the expected clade. Our *hispanicus* specimens, however, fall into two very divergent mtDNA clades: the “*P. hispanicus* Galera” clade and the “*P. hispanicus* Valencia” clade. Detailed analysis of this phenomenon is beyond the scopes of the present paper and is presented in details elsewhere (Renoult *et al.* submitted). Briefly, this mtDNA differentiation is not accompanied by differentiation in nuclear DNA: specimens carrying either of these mtDNA types exhibit the same alleles for the nuclear intron and the same microsatellite genotypes. The presence of these two lineages in *hispanicus* (in the restricted sense of Geniez *et al.* 2007) corresponds to a mtDNA polymorphism, probably resulting from an ancient introgression. The original *hispanicus* haplotype appears to be the “*P. hispanicus* Galera” lineage: this lineage is only found in *hispanicus*, where it exhibits a relatively deep intra-taxon divergence indicating long evolutionary history in this taxon, while the “*P. hispanicus* Valencia” lineage is found in other morphotypes as well and shows no polymorphism at all (a single haplotype) in *hispanicus*, indicating recent introduction in *hispanicus*.

All “Valencia” and “southern TYP3” specimens fall (with part of the *hispanicus* specimens, as detailed above) into the “*P. hispanicus* Valencia” mtDNA clade (called “southeastern clade” or “SE clade” hereafter). They are thus clearly different from the TYP3 reference specimens in mtDNA, despite their TYP3 morphological characters. Other markers are thus needed to decide whether this difference in mtDNA corresponds to mitochondrial polymorphism within a single lineage or to different lineages. We will investigate below whether further morphological data and nuclear markers confirm this differentiation.

Nuclear differentiation between *hispanicus*, TYP2, reference and southern TYP3 specimens

The Geneland Software attributed the highest posterior probability to a model with 3 nuclear clusters (see Renoult *et al.*, submitted, for detailed results). Figure 3 summarizes the assignation of each specimen to one of the three clusters defined. In the most frequent clustering pattern generated by the algorithm, *hispanicus* specimens constituted a distinct nuclear cluster. This was also the case for the TYP2 specimens. The third cluster included all individuals from “Valencia”, “southern TYP3” and reference TYP3. The geographic delimitation of TYP2, *hispanicus* and TYP3 (including Valencia and southern populations) as it was described in Geniez (2001) according to the morphology thus corresponds to nuclear units defined on the basis of microsatellite loci.

Morphological differentiation between reference TYP3 and “southern TYP3” + “Valencia” specimens

The morphology of the “southern TYP3” and “Valencia” specimens (including the syntypes) overlap broadly with the morphology of the reference TYP3 specimens of “liolepis” morphotype (Fig. 4). Furthermore, there is as much morphological differentiation between TYP3 reference specimens of “robusta” and “liolepis” morphotypes (both carrying TYP3 mtDNA) as between “liolepis” reference specimens and “southern TYP3” plus Valencia specimens (carrying the SE mtDNA).

Nuclear DNA differentiation between reference and southern TYP3

B and 2C shows the relationships between alleles of β fib and Pdg, respectively, of reference TYP3 and southern TYP3 specimens. For both nuclear markers, reference TYP3 and “southern TYP3” specimens were mixed in all branches of the trees, meaning that populations carrying TYP3 mtDNA were not differentiated from those carrying the “south-eastern” mtDNA clade in any of these two markers.

Discussion

Discriminant analyses results suggest that the syntypes of *liolepis* and the modern specimens from Valencia (the type locality of *liolepis*) belong to the same morphotype as specimens of the TYP3 lineage of the *Podarcis hispanicus* complex (the morphotype 3). Our mitochondrial analysis showed, however, that the southern populations of this morphotype (our “southern TYP3” and “Valencia” specimens) all carry a mitochondrial lineage (the SE clade or “*P. hispanicus* Valencia” lineage) unrelated to the TYP3 mtDNA lineage. As expected, the TYP3 mtDNA lineage is the only one found in northern populations of the morphotype 3 (our reference TYP3 specimens).

This geographic bipartition of the mitochondrial polymorphism found in the morphotype 3 was, however, not mirrored by microsatellite loci variation. The geographic distribution of the three nuclear clusters of *Podarcis hispanicus* generated by Geneland for the south-eastern Iberian Peninsula matches with the one based on morphotypes, and therefore do not identify the southern TYP3 specimens, carrying the SE mtDNA clade, as a nuclear unit distinct from the reference TYP3 specimens.

The distribution of the two nuclear introns alleles in reference TYP3 and “southern TYP3” specimens further support that they belong to a single nuclear gene pool: we could not find any separation between specimens carrying either of the two mtDNA clades.

Altogether, these results suggest that mitochondrial DNA alone is not a reliable marker of evolutionary units in the *Podarcis hispanicus* complex of the south east Iberian Peninsula. Renoult (2006) and Renoult *et al.* (submitted) clearly demonstrated that specimens of *Podarcis hispanicus* (sensu Geniez *et al.* 2007) that do not differ in morphology, microsatellites or β fib intron polymorphism can have two very different mtDNA lineages. This apparently also occurs in the morphotype 3, as the southern and northern populations of this morphotype carry very divergent mtDNA lineages but do not differ significantly in morphology, microsatellites or β fib and Pgd intron polymorphism. Interestingly, the same mtDNA lineage (the south-eastern clade) is responsible for this mtDNA polymorphism in both the *hispanicus* and TYP3 morphotypes.

It is beyond the scope of this systematic paper to discuss the evolutionary scenarios responsible for this situation. Briefly, Renoult (2006) and Renoult *et al.* (submitted) suggest that it results from past introgression following hybridization events between an unknown taxon (currently extinct or extremely restricted in distribution) and both *hispanicus* and the “type 3” taxon.

Conclusion: the correct name of the north-eastern form of the *Podarcis hispanicus* complex

Busack *et al.* (2005) argued that the north-eastern Spanish form (= “*P. hispanica* morphotype 3” in Pinho *et al.* 2006) is conspecific with the Columbretes populations and proposed to use the binomen *Podarcis atratus* to designate this species. Several lines of evidence reported in the present work support the conclusion that the populations of the *Podarcis hispanicus* complex from the town of Valencia and the surrounding areas, including most original syntypes and the lectotype of *liolepis*, belong to the same species as the typical populations of “*P. hispanica* type 3” located in the north-east of Spain. The valid nomen of the north-eastern form of the *Podarcis hispanicus* complex (= *Podarcis hispanica* type 3) is thus *liolepis*, which has precedence over *atratus* when the two nomina are in concurrence to designate a species-rank taxon. Last, since multiple recent evidence of reproductive isolation between the morphotype 3 and both morphotype 2 and *hispanicus* have been reported recently (Pinho *et al.* 2007; Renoult 2006; Renoult *et al.* submitted), we concur with Busack *et al.*

(2005) and Geniez (2001) that the morphotype 3 should be considered as a full species: *Podarcis liolepis*.

Systematic section

Podarcis liolepis (Boulenger, 1905)

Lacerta muralis var. *liolepis*, Boulenger, 1905, Transactions of the zoological Society of London 17: 363. Name-bearing type: Natural History Museum (London) BMNH 86.12.29.3, an adult male from Valencia, coll. or leg by Prof. E. Boscá, lectotype by present designation (see Fig. 5). The list of the known original syntypes is given above. Type locality: Valencia, Spain, by present lectotype designation. Originally “Valencia”, “Foyos” (= Foios), near Valencia, “Sevilla”, and along the road from Valencia to Dehesa de la Albufera, Spain.

Junior synonym

Lacerta muralis var. *hesperica*, Schreiber, 1912, Herpetologia europaea. Eine systematische Bearbeitung der Amphibien und Reptilien welche bisher in Europa aufgefunden sind: 943. Name-bearing types: Five syntypes, still extant: Naturhistorisches Museum Wien NMW 16020 and NMW 16021: 1-4, syntypes (Tiedemann *et al.* 1994). Type locality: Zaragoza, Spain.

Remarks: 1) This nomina was treated as an infrasubspecific nomen by Mertens and Wermuth (1960) and Salvador (1986), and therefore as not available (“nomen illegitimum”). This is clearly a mistake: Schreiber (1912) described *hesperica* as a variety of *Lacerta muralis* (see heading of page 943), treated it as a variety of *muralis* (see page 410). This is confirmed by the original labels of the *hesperica* syntypes in Wien, which writes as: “NMW 16020 *Lacerta bocagei liolepis* Blgr (Lectotypus of *L. muralis* var. *hesperica* Schreib.) Saragossa; Southoff leg, VIII.1912 Coll. Schreiber / NMW 16021:1-4 *Lacerta muralis* var. *hesperica* Zaragoza, Spaniel 1915.II.22E Coll. Schreiber Nr. 84,89 leg. Southoff”.

2) Despite the mention on the labels that one of the specimen is a lectotype, neither Tiedemann *et al.* (1994) nor ourselves were able to find a published lectotype designation for *hesperica*. The specimens of the original type series thus all remain syntypes.

Proposed status: subjective junior synonym of *Podarcis liolepis*. This nomen would be available for the northern type 3 populations (carrying the type 3 mtDNA lineage) if they were recognised as distinct from the southern type 3 ones (carrying the south-eastern mtDNA lineage = the “*Podarcis hispanica sensu stricto*” lineage of previous workers and “Valencia” lineage of Renoult et al. submitted).

A series of other nomina apply to various populations of *Podarcis liolepis*. They are either subjective junior synonyms of *liolepis* (if *Podarcis liolepis* is treated as a monotypic species) or valid nomina for subspecies of *Podarcis liolepis*. As a review of the morphological variation and subspecific systematics of *Podarcis liolepis* is beyond the scope of this paper, we do not address their status here but will do it in a forthcoming work dedicated to the infraspecific systematics of *Podarcis liolepis*.

Acknowledgements: We thank J. Cabot (EBD, Seville), C. McCarthy (BMNH, London) and F. Tiedemann (NMW, Vienna) for their help in examining specimens housed in these collections. A. Dubois, A.-M. Ohler, J. Schmidtler and P. Wagner helped us in interpreting the original German text of Schreiber (1912). We thank J.-P. Baron, A. Bea, W. Böhme, S. Boissinot, G. Bruneau, O. Chaline, M. Cheylan, P. David, P. Escudié, F. Geniez, M. Geniez, A. Gosa, G. Guillaume, V. Joubert, J. A. Mateo, U. Mathis, G. Moraguès, T. Menut, J.-M. Pleguezuelos, A. Ravayrol, D. Renoult, M. Siol, F. Veyrunes, J. Viglione and J. Wolff, for helping to realize this study including help with references, field work and pictures or specimens’ collections.

References

- Bedriaga, J. von (1879) Herpetologische Studien (Fortsetzung). *Archiv für Naturgeschichte, Berlin*, 45(1), 243–339.
- Böhme, W. & Köhler, J. (2005) (“2004”) Do endings of adjective flexible species names affect stability? A final note on the gender of *Podarcis* Wagler, 1830 (Reptilia, Lacertidae). *Bonner zoologische Beiträge*, 53(3/4), 293–295.
- Boulenger, G.A. (1905) A contribution to our knowledge of the varieties of the Wall-Lizard (*Lacerta muralis*) in Western Europe and North Africa. *Transactions of the Zoological Society of London*, 17, 351–437.
- Boulenger, G.A. (1921) *Monograph of the Lacertidae, vol. II*. Trustees of the British Museum of Natural History, London.
- Busack, S.D., Lawson, R. & Arjo, W.M. (2005) Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphibia–Reptilia*, 26, 239–256.
- Castilla, A.M., Fernández-Pedrosa, V., Backeljau, T., González, A., Latorre, A. & Moya, A. (1998a) Conservation genetics of insular *Podarcis* lizards using partial cytochrome *b* sequences. *Molecular Ecology*, 7, 1407–1411.
- Castilla, A.M., Fernández-Pedrosa, V., Harris, D.J., González, A., Latorre, A. & Moya, A. (1998b) Mitochondrial DNA divergence suggests that *Podarcis hispanica atrata* (Suamata: Lacertidae) from the Columbretes Islands merits specific distinction. *Copeia*, 1998, 1037–1040.
- Crochet, P.A., Chaline, O., Surget-Groba, Y., Debain, C., & Cheylan, M. (2004) Speciation in mountains: phylogeography and phylogeny of the rock lizards genus *Iberolacerta* (Reptilia : Lacertidae). *Molecular Phylogenetics and Evolution*, 30, 860–866.
- Geniez, P. (2001) *Variation géographique des lézards du genre Podarcis (Reptilia, Sauria, Lacertidae) dans la péninsule Ibérique, l’Afrique du Nord et le sud de la France*. Unpublished E.P.H.E. diploma thesis. Montpellier: Université Montpellier II and Ecole Pratique des Hautes Etudes, 342 pp.
- Geniez, P., Cluchier, A., Sá-Sousa, P., Guillaume, C.P. & Crochet, P.-A. (2007) Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetological Journal*, 17, 69–80.
- Guillaume, C.P. (1987) *Les petits Lacertidés du Bassin méditerranéen occidental (genera Podarcis et Archaeolacerta essentiellement)*. *Sur quelques problèmes d’ordre*

- systematique et biogeographique*. Unpublished doctoral thesis. Montpellier: Université des Sciences et Techniques du Languedoc, 474 pp.
- Guillaume, C.P. (1988) Réflexions sur l'utilisation du critère "plaque massétérique" en systématique des Lacertidés et proposition d'un indice nouveau: l'indice massétérin [M]. *Bulletin de la Société Herpétologique de France*, 46, 10–21.
- Guillaume, C.P. & Geniez, P. (1986) Contribución a la biogeografía y la sistemática de las lagartijas del género *Podarcis* en península Ibérica y Africa del Norte Abstracts of the communications of 1º Congreso Nacional d'Herpetologia de l'A.H.E. (Benicasim, 1-3.11.1986), 60.
- Guillot, G., Mortier, F. & Estoup, A. (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, 5, 712–715.
- Harris, D.J. & Sá-Sousa, P. (2001) Species distinction and relationships of the Western Iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetological Journal*, 11, 129–136.
- Harris, D.J. & Sá-Sousa, P. (2002) Molecular phylogenetics of Iberian wall lizards (*Podarcis*): is *Podarcis hispanica* a species complex? *Molecular Phylogenetics and Evolution*, 23, 75–81.
- Harris, D.J., Batista, V., Carretero, M.A. & Pinho, C. (2002a) Mitochondrial DNA sequence data confirms the presence of *Podarcis carbonelli*, Pérez-Mellado, 1981 in southern Spain. *Herpetozoa*, 15, 188–190.
- Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C. & Ferrand, N. (2002b) Complex biogeographical distribution of generic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, 29, 1257–1262.
- International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature*. Fourth Edition. The International Trust for Zoological Nomenclature, London, xxix + 306 pp.
- Kumar, S., Tamura, K. & Nei, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 5, 150–163.
- Mertens, R. & Wermuth, H. (1960) *Die Amphibien und Reptilien Europas*. Verlag Waldemar Kramer, Frankfurt am Main, 264 pp.
- Oliverio, M., Bologna, M.A. & Mariottini, P. (2000) Molecular biogeography of the Mediterranean lizards *Podarcis* Wagler, 1830 and *Teira* Gray, 1838 (Reptilia, Lacertidae). *Journal of Biogeography*, 27, 1403–1420.

- Pinho, C., Ferrand, N. & Harris, D.J. (2006) Reexamination of the Iberian and North African *Podarcis* (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, 38, 266–273.
- Pinho, C., Harris, D.J. & Ferrand, N. (2003) Genetic polymorphism of 11 allozyme loci in populations of wall lizards (*Podarcis sp.*) from the Iberian Peninsula and North Africa. *Biochemical Genetics*, 41, 343–359.
- Pinho, C., Harris, D.J., & Ferrand, N. (2007) Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (*Podarcis*, Lacertidae). *Biological Journal of the Linnean Society: A*, 91, 121–133.
- Pinho, C., Harris, D.J. & Ferrand, N. (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis spp.*) are an assemblage of incipient species. *Bmc Evolutionary. Biology*, 8.
- Pinho, C., Sequeira, F., Godinho, R., Harris, D.J. & Ferrand, N. (2004) Isolation and characterization of nine microsatellite loci in *Podarcis bocagei* (Squamata : Lacertidae). *Molecular Ecology Notes*, 4, 286–288.
- Pleguezuelos, J.M., Márquez, R. & Lizana, M. (Eds) (2002) *Atlas y libro rojo de los anfibios y reptiles de España, 2nd edition*. Madrid: Dirección General de Conservación de la Naturaleza – Asociación Herpetologica Española.
- Renoult J.P. (2006) *Histoire évolutive du Lézard ibérique Podarcis hispanicus (Squamata: Lacertidae): discordances entre morphologie, ADN nucléaire et ADN mitochondrial*. In *MsC Thesis*. Université Montpellier II Sciences et Techniques du Languedoc.
- Renoult, J.P., Geniez, P., Bacquet, P., Benoit, L & Crochet, P.-A. (submitted) The Necessity of Multi-Marker Analyses for Systematic Studies: the Case of the Iberian Wall Lizard *Podarcis hispanicus* (Squamata, Lacertidae) Species Complex.
- Salvador, A. (1986) *Podarcis hispanica* (Steindachner, 1870) – Iberische Mauereidechse. In *Handbuch der Reptilien und Amphibien Europas, Band 2 (2), Echsen 3*, Aula, Wiesbaden, pp. 71-82.
- Sá-Sousa, P. (2000) A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetological Journal*, 10, 1–11.
- Sá-Sousa, P., Vicente, L. & Crespo, E.G. (2002) Morphological variability of *Podarcis hispanica* (Sauria: Lacertidae) in Portugal. *Amphibia-Reptilia*, 23, 55–69.

- Schreiber, E. (1912) *Herpetologia europaea. Eine systematische Bearbeitung der Amphibien und Reptilien welche bisher in Europa aufgefunden sind*. Verlag von Gustav Fischer, Jena.
- Stephens, M., Smith, N.J. & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68, 978–989.
- Tiedemann, F., Häupl, M. & Grillitsch, H. (1994) Katalog der Typen der herpetologischen Sammlung nach dem Stand vom Jänner 1994, Teil II: *Reptilia*. *Kataloge der wissenschaftlichen Sammlungen des Naturhistorischen Museums in Wien*, Wien. 10 (Vertebrata 4): Selbstverlag Naturhistorisches Museum, Wien, 110 pp.

Figure captions

Figure 1. First and second components of the Discriminant Analyses performed with 16 morphological variables on specimens of the *Podarcis hispanicus* complex. Specimens from Valencia and *liolepis* syntypes from Valencia were treated as non-active individuals. We did not include more than three lineages in a single analysis to maximise the plot scattering. Two distinct analyses were thus performed here.

Figure 2. NJ phylogenetic tree of mitochondrial control region haplotypes (A) and β fib (B) and Pgd (C) alleles, with bootstrap support indicated for values above 50%. A: Symbols give the name of the mitochondrial lineages according to Pinho *et al.* (2006): filled rhombus = “*P. hispanicus sensu stricto*” lineage (“*P. hispanicus* Valencia” is this work), open triangles = “*P. hispanicus* type 2” lineage, open squares = “*P. hispanicus* type 3” lineage and open circles = “*P. hispanicus* Galera type”. Underlined specimen numbers refer to individuals from Valencia. Specimens of *Podarcis hispanicus* with type 3 mtDNA lineage are also indicated. For specimen in (B) and (C), mitochondrial lineage has been inferred from the geographical distribution when mtDNA has not been sequenced.

Figure 3. Map of the south-east Iberian Peninsula summarizing the results obtained with Geneland (Renoult *et al.* submitted). The locations sampled for the nuclear clustering are mapped with symbols corresponding to mitochondrial clade membership (sometimes inferred from the geographical distribution of the mitochondrial lineages): filled rhombus = “*P. hispanicus* Valencia” lineage, open triangles = “*P. hispanicus* type 2” lineage, open squares = “*P. hispanicus* type 3” lineage and open circles = “*P. hispanicus* Galera type”. Light gray, medium gray and dark gray mark include populations that were attributed to “type 2”, “type 3” and *hispanicus* clusters, respectively, with a posterior probability > 0.9. One “*P. hispanicus* type 2” was attributed to “type 2” nuclear cluster with posterior probability of 85 % and is thus located outside the light gray shading. Dashed lines delimit the morphotype 2 (west), morphotype 3 (north) and *hispanicus* (south) distributions.

Figure 4. First and second components of the Principal Component Analysis performed on TYP3, syntypes of *liolepis* and Valencia specimens only using 16 morphological variables.

Figure 5. Latero-dorsal views of *Podarcis liolepis* (Boulenger, 1905). A. Specimen BMNH 86.12.29.3, lectotype of *Lacerta muralis* var. *liolepis* Boulenger, 1905, Valencia, Spain. Photo

C. McCarthy. B. Specimen BEV 9853 with “*Podarcis hispanicus* Valencia” mitochondrial lineage, male, collected in Rafelbuñol, north of Valencia, Spain. Photo PGe. C. Specimen BEV 9815 with “*Podarcis hispanicus* type 3” mitochondrial lineage, male, collected 3 Km NW of Miami-Platja along the national road 340, Spain. Photo PAC. Note the great morphological resemblance of both *Podarcis liolepis* (BEV 9853 and BEV 9815) showing however a deep mitochondrial divergence.

Appendix 1. GenBank accession numbers of β fib and Pgd sequences. Individual labels correspond to BEV collection numbers excepted for individuals retrieved from Pinho *et al.* (2006).

individual	locality	GenBank Acc.Number	
		β fib	Pgd
1803	Villanueva de Huerva	-	FJ586260
1808	Codos (between Belchite and Cariñena)	FJ586299	-
1865	Albarracín (West of Teruel)	FJ586300	FJ586247
1867	Albarracín (West of Teruel)	FJ586301	FJ586248
1868	Albarracín (West of Teruel)	-	FJ586261
1870	Albarracín (West of Teruel)	FJ586302	-
1872	Albarracín (West of Teruel)	FJ586303	FJ586249
1873	Albarracín (West of Teruel)	FJ586304	FJ586250
1875	Albarracín (West of Teruel)	FJ586305	FJ586262
4551	2 km North of Farena (41,273°N / 1,065°E)	FJ586306	-
9815	GPS: 41,0311°N / 0,9666°E	FJ586307	FJ586263
9818	GPS: 41,0311°N / 0,9666°E	FJ586308	FJ586264
9819	GPS: 41,0311°N / 0,9666°E	FJ586309	-
9820	GPS: 41,0311°N / 0,9666°E	FJ586310	FJ586251
9821	GPS: 41,0311°N / 0,9666°E	-	FJ586252
9822	GPS: 41,0311°N / 0,9666°E	FJ586311	FJ586253
9823	San Carles de la Rapita	FJ586312	FJ586265
9825	San Carles de la Rapita	FJ586313	FJ586254
9826	9 Km North of San Carles de la Rapita	-	FJ586266
9827	10 Km North of San Carles de la Rapita	FJ586314	FJ586267
9829	11 Km North of San Carles de la Rapita	-	FJ586268
9830	12 Km North of San Carles de la Rapita	FJ586315	-
9831	13 Km North of San Carles de la Rapita	-	FJ586269
9832	14 Km North of San Carles de la Rapita	FJ586316	-
9833	L'Aldea	-	FJ586270
9835	L'Aldea	FJ586317	FJ586271
9836	Road of Peñiscola [40,3911°N / 0,3759°E]	FJ586318	FJ586272
9837	Road of Peñiscola [40,3911°N / 0,3759°E]	-	-
9838	Torreblanca	FJ586319	-
9839	Torreblanca	FJ586320	-
9840	Torreblanca	FJ586321	-
9841	Oropesa del Mar	FJ586322	-
9842	Oropesa del Mar	-	FJ586273
9843	Oropesa del Mar	FJ586323	FJ586274
9844	Northern entance in Benicasim	FJ586324	-
9845	Northern entance in Benicasim	-	FJ586275
9846	Northern entance in Benicasim	-	FJ586276
9847	Castellón de la Plana	-	FJ586277

9848	Castellón de la Plana	-	FJ586278
9851	Sagunto	FJ586325	-
9852	Rafelbuñol (North of Valencia)	FJ586326	FJ586256
9853	Rafelbuñol (North of Valencia)	FJ586327	-
9854	Between Maises and the rio Turia	FJ586328	FJ586257
9855	Massanassa (South of Valencia)	FJ586329	FJ586258
9856	Perellonet (South of Valencia)	FJ586330	FJ586259
9858	Mareny Blau	FJ586331	-
9861	Mareny de San Llorenç de Cullera (South of Valencia)	FJ586332	FJ586279
9870	Southern entrance in Ibi coming from Tibi	-	FJ586280
9871	Southern entrance in Ibi coming from Tibi	FJ586334	-
9872	Southern entrance in Ibi coming from Tibi	-	FJ586281
9873	Southern entrance in Ibi coming from Tibi	FJ586335	FJ586282
9874	Ibi, city-center	-	FJ586283
9875	0.5 Km after Ibi going to Bañeres	FJ586336	FJ586284
9879	7 km after Sella going to Alcolecha	-	FJ586285
9880	7 km after Sella going to Alcolecha	-	FJ586286
9882	1 km before Port de Tudons coming from Orxeta	FJ586337	FJ586287
9883	1 km before Port de Tudons coming from Orxeta	FJ586338	FJ586288
9884	0.5 Km after Port de Tudons dir. Alcolecha	FJ586339	-
9885	1 km before Alcolecha coming from Orxeta	FJ586340	-
9886	1 km before Alcolecha coming from Orxeta	FJ586341	-
9887	1 km before Alcolecha coming from Orxeta	-	FJ586289
9888	1 km before Alcolecha coming from Orxeta	FJ586342	-
9889	Benifato, northern part of the village	FJ586343	FJ586290
9890	Benifato, northern part of the village	FJ586344	FJ586291
9891	Castello de Guadalest	FJ586345	-
9893	Castello de Guadalest	FJ586346	-
9894	Castello de Guadalest	-	FJ586292
9895	4 km before Alcoy coming from Benasau	FJ586347	FJ586293
9896	2 km before Alcoy coming from Benasau	-	FJ586255
9898	2 km before Alcoy coming from Benasau	FJ586348	FJ586294
9899	2 km before Alcoy coming from Benasau	-	FJ586295
9900	2 km before Alcoy coming from Benasau	-	FJ586296
9901	2 km before Alcoy coming from Benasau	-	FJ586297
9903	Baneyres, road toward Villena	FJ586349	-
T1200	Sagunto	-	FJ586298
muralis1	<i>allele B80 in Pinho et al. (2008)</i>	EU269550	-
muralis2	<i>allele B79 in Pinho et al. (2008)</i>	EU269549	-
muralis-1	<i>allele P65 in Pinho et al. (2008)</i>	-	EU269659
muralis-2	<i>allele P64 in Pinho et al. (2008)</i>	-	EU269658
muralis-3	<i>allele P63 in Pinho et al. (2008)</i>	-	EU269657

Figure 1.

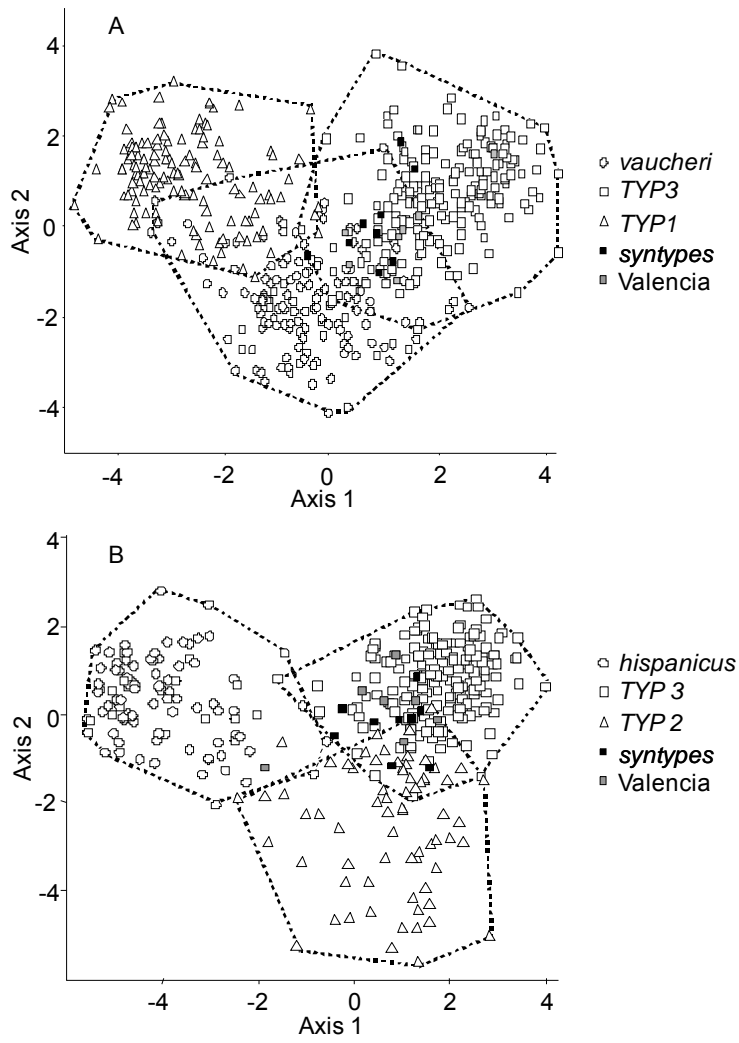


Figure 2.

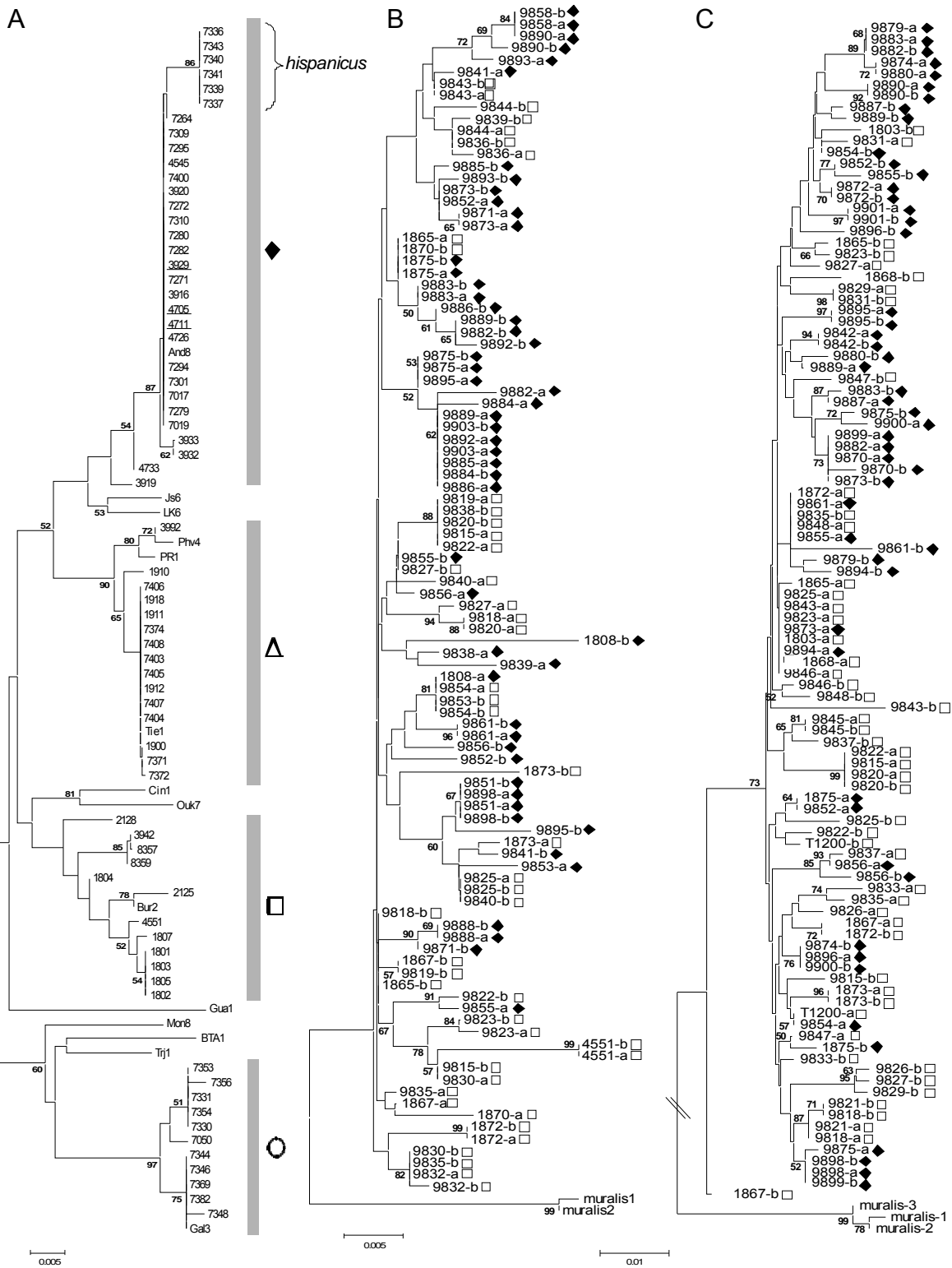


Figure 3.

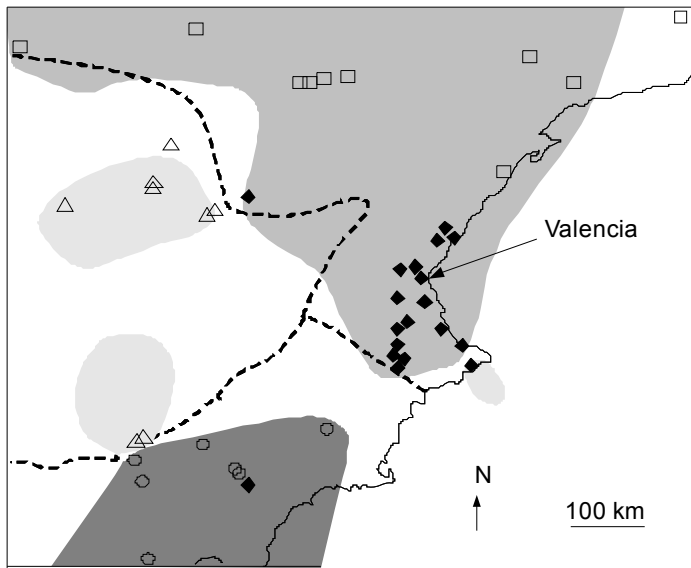


Figure 4.

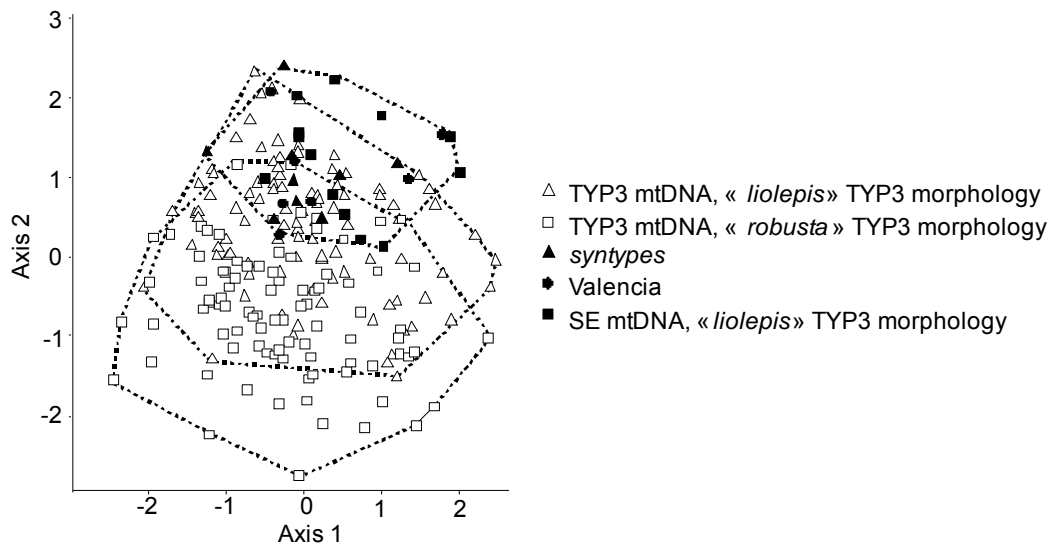
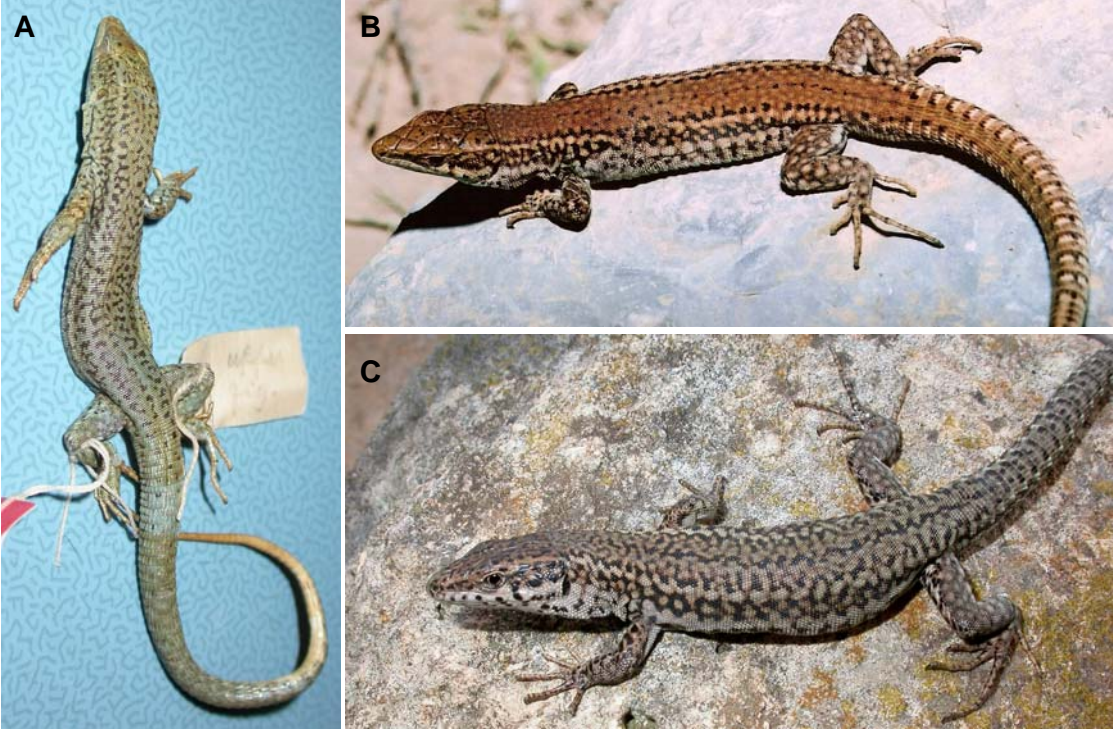


Figure 5.



ANNEXES

AUTRES MANUSCRITS REDIGES DURANT LA THESE

Article S1

When assumptions on visual system evolution matter: nestling colouration and parental visual performance in birds **p.214**

Julien P. Renoult, Alexandre Courtiol, Finn Kjellberg

Journal of Evolutionary Biology, in press

Article S2

*Cyto-nuclear discordance in the phylogeny of *Ficus* section Galoglychia and host shifts in plant-pollinator associations* **p.229**

Julien P. Renoult, Finn Kjellberg, Cinderella Grout, Sylvain Santoni, Bouchaïb Khadari

Amphibia-Reptilia, in press.

Article S3

Typhlops rajeryi, a new species of Malagasy blind snake (Serpentes: Typhlopidae) **p.244**

Julien P. Renoult & Achille P. Raselimanana

Submitted to Zootaxa

When assumptions on visual system evolution matter: nestling colouration and parental visual performance in birds

Julien P. RENOULT¹, Alexandre COURTIOL^{2,3}, Finn KJELLBERG¹

¹*UMR 5175 Centre d'Ecologie Fonctionnelle et Evolutive, 34293 Montpellier cedex 5, France*

²*Université Montpellier II, Place Eugène Bataillon, C.C. 065, 34095 Montpellier cedex 5, France*

³*UMR 5554 Institut des Sciences de l'Evolution, Place Eugène Bataillon, C.C. 065, 34095 Montpellier cedex 5, France*

Running title: Assumptions on avian visual system evolution

Corresponding authors:

JPR: E-mail: julien.renault@cefe.cnrs.fr, phone: +33 4 67 61 32 62, fax : +33 4 67 41 21 38

FK: E-mail: finn.kjellberg@cefe.cnrs.fr, phone: +33 4 67 61 32 29, fax : +33 4 67 41 21 38

Abstract

Comparative studies in visual ecology of birds often rely on several assumptions on the evolution of avian vision. In this study, we show that when these assumptions are not upheld, conclusions may be strongly affected. To illustrate this purpose, we reanalysed the data of Avilés & Soler (2009) who demonstrated that nestling gape colouration in altricial birds is associated with visual system. We show that a slight change in analysis methodology leads to opposite conclusions. Such conflicting result raises the problem of applying powerful methods developed for continuous variables to a small sample and a small number of independent events of qualitative visual system shift in comparative analyses. Further, we show that the current trend to assume strong phylogenetic inertia of avian visual systems is contradicted by data and that the sequencing of the SWS1 opsin gene should be considered as an alternative approach.

Key words: avian visual system, UVS, VS, SWS1 opsin, phylogenetic inertia, comparative analyses

The evolution of avian colouration and colour vision has drawn attention ever since the earliest biologist (reviewed in Bennett & Théry, 2007). This field of evolutionary ecology has become even trendier over the last decade due to the development of portable spectrometers that allow factual quantitative measurements of colours. Another major progress was the development of models of avian vision allowing to analyse how colours are effectively perceived by birds (Vorobyev *et al.*, 1998; Goldsmith & Butler, 2003; Endler & Mielke, 2005). The necessity of such an approach has become obvious ever since birds have been shown to perceive colours differently from humans (Huth & Burkhardt, 1972; Håstad & Ödeen, 2008). Avian vision models are based on physiological data such as the area of the pupil, the transmittance of ocular media and oil droplets, the spectral sensitivity of photopigments or the relative proportion of the different photopigments on the retina. Indeed these measurements allow to determine how a colour spectrum arriving onto a bird eye is processed. Obviously, such physiological data are available only for few bird species (Hart, 2001; Hart & Vorobyev, 2005). While most parameters do not seem to exhibit marked variation among species, the range of spectral absorbance of one of the four avian photopigments strongly varies according to a bimodal distribution such that bird eyes are usually classified as being either Ultra-Violet Sensitive (UVS) or Violet Sensitive (VS).

Two methods are mainly used to get around the lack of information about a specific visual system. First, the effects of varying the visual system can be assessed by performing analyses with both visual systems and by checking the consistency of results (Uy & Endler, 2004; Doucet *et al.*, 2007; Gomez & Théry, 2007; Schaefer *et al.*, 2007; Schaefer *et al.*, 2008; Langmore *et al.*, 2009). If the visual system has limited effects on vision performance, knowing accurately the visual system will be largely irrelevant to the particular study. This method is, however, not always applicable, especially when the specific purpose of a study is to compare any characteristics between UVS and VS species. In that case, a second method may be more useful. It consists in assuming that a species has the same visual system as the most closely related species for which this information is available. This method is currently the most widely used one and is usually justified by supposedly strong phylogenetic inertia of avian visual systems (*e.g.* Benites *et al.*, 2007; Eaton, 2007; Avilés *et al.*, 2008; Bridge *et al.*, 2008; Cassey *et al.*, 2008a; Cassey *et al.*, 2008b; Wilson *et al.*, 2008; Avilés & Soler, 2009).

The incidence of uncertainties of visual system assignation on results in studies of visual ecology has been poorly explored. In this contribution, we use a recently published paper by Avilés and Soler (2009) to discuss these effects. Avilés and Soler compared the detectability of nestling gape and skin colouring in UV-tuned bird eyes and VS-tuned bird

eyes. The main part of their study compared gape and body skin colouration of nestlings belonging to UVS and VS species. After controlling for phylogenetic effects, they showed that gape colouration but not body skin colouration differs significantly between species according to their visual systems. By construction, this result strongly depends on the initial visual system classification. We revisit the visual system classification of the species studied by Avilés and Soler and we show that our classification procedure leads to different conclusions. We then analyse the widely accepted assumption of strong phylogenetic inertia of avian visual system and show that this assumption is not reliable. We eventually discuss the validity of a few other methods and assumptions commonly used in comparative studies of visual ecology.

To assign a vision system to 15 of their 22 studied species, Avilés and Soler used a parsimonious approach based on the hypothesis of strong phylogenetic inertia in the evolution of bird photopigments, because “no splits in the type of vision have so far been reported within a bird family”. The authors rely on the results of Ödeen and Håstad (2003) who inferred the evolution of UVS pigments across the bird phylogeny from the analysis of key mutation sites in the SWS1 opsin gene. Ödeen and Håstad showed that all the passerine species they studied possessed UVS pigments, with the exception of two Tyrannidae species and two Corvidae species. Following this, Avilés and Soler classified their 14 passerine species as UVS, with the exception of four Corvidae species, and also considered their two shrike species (*Lanius spp.*) as UVS (Fig. 1).

We obtained a different visual system classification following the same line of reasoning *i.e.* based on the assumption of strong phylogenetic inertia and using current knowledge of bird phylogeny. Shrikes form a family close to the Corvidae, a phylogenetic affinity supported by comprehensive phylogenetic (Barker *et al.*, 2004) and phylogenomic (Hackett *et al.*, 2008) studies. The Laniidae and Corvidae families are included in a clade called Corvoidea, which forms a sister lineage to the Passerida clade. Both the Corvoidea and Passerida clades belong to the Oscines lineage. Tyrannidae belong to the Suboscines clade, a basal lineage of passerine birds. To date, all occurrences of UVS pigments reported in passerine birds concern the Passerida clade and, to our knowledge, all investigated species of Passerida present UVS pigments. Further, no occurrence of UVS pigments has been detected among passerine species that do not belong to the Passerida clade. Hence the unique Cys90Ser substitution in the SWS1 opsin gene allowing the shift to UVS pigments from VS pigments (Yokoyama *et al.*, 2000) must have appeared at the base of the Passerida lineage. As

a consequence, current data suggest a single origin of UVS pigments in passerine birds. Classifying both species of shrikes as UVS requires a second, independent, origin of UVS pigments, an event which is not supported by any published data. The most parsimonious visual system classification of the passerine bird species studied by Avilés and Soler would thus classify all species as VS except species belonging to the Passerida lineage (Fig. 1).

In addition, the authors included two species of owls in their analyses that they modelled as presenting a violet-tuned eye, based on the work of Bowmaker and Martin (1978). Bowmaker and Martin used microspectrophotometer measurements to reveal the presence of only three classes of visual pigments in the tawny owl *Strix aluco*, instead of the four classes usually encountered in birds. Because the pigment presenting its maximal absorbance in the shortest wavelength peaks around 463 nm, it is now generally considered that this species lacks the SWS1 opsin protein (Vorobyev *et al.*, 1998). This proposition has been supported by a failure to amplify a fragment of the SWS1 gene in this species (Ödeen & Håstad, 2003). No physiological or molecular data have been published on any other nocturnal bird species; however, lack of the SWS1 opsin protein is suggested by behavioural experiments on another species of owl (Koivula *et al.*, 1997). The absence of SWS1 pigment in the tawny owl is probably linked to its nocturnal habit (Bennett & Cuthill, 1994; Cuthill *et al.*, 2000). If we assume strong phylogenetic inertia of avian visual pigment evolution, then it would be most likely that both owl species analysed by Avilés and Soler share this characteristic. Hence, pending new data, no owl should be included within a dataset comparing the performance of tetrachromatic species (*i.e* having VS or UVS-tuned eyes) because the owls should be classified as trichromatic.

To assess the consequences of using the most parsimonious visual system classification, we reinvestigated the relationship between parental visual system and gape colouration first using the same method as Avilés and Soler, and then with a slightly different method. First we ran two General Linear Models (GLM) in R 2.8 (R Development Core Team), with variables defining gape colouration (using PC1 and PC2 colour scores for flanges and mouth, provided in Table S1 in Avilés & Soler, 2009) as dependent variables and parental visual system and nesting site as independent variables. The first GLM used Avilés and Soler's parental visual system classification, and the second GLM used the most parsimonious classification (Fig. 1). Using the most parsimonious classification did not lead to qualitatively different results, with a significant effect of parental visual system on PC1 scores for flanges and PC2 scores for mouths (Table 1; note that in the original table given in Avilés & Soler, 2009, several result lines are inverted). Then, to take into account the

statistical dependence induced by phylogenetic inertia, we performed GLMs using Felsenstein's phylogenetic independent contrasts method (Felsenstein, 1985), as did Avilés and Soler. With the most parsimonious classification, the significant effect of parental visual system on both PC1 scores for flanges and PC2 scores for mouths vanishes (Table 1). Further, this non-significant result persists even when we only removed both owl species from the analyses, and classified both *Lanius* species as UVS birds.

Contrasts measure differences between character values in pairs of sister species and/or nodes, meaning that contrasts involving longer periods of time (estimated by branch length) are expected to present higher absolute values and are therefore given greater weight by the GLM (Garland *et al.*, 1992). This effect is particularly likely to affect analyses relying on dissymmetric topologies, such as the topology used in Avilés and Soler. Contrast data should therefore be standardized before performing any regression analysis, by dividing the raw data by the square root of the sum of corrected branch lengths (Garland *et al.*, 1992). In a final analysis, we therefore ran the GLM using standardized contrast values and the most parsimonious classification. Again, the results do not support the relationship between parental visual system and nestling gape colouration: only the nesting site variable is significant in this final model (Table 1).

The analyses presented here show that visual system classification can strongly influence the results of the study. Because removing owls from the data set is enough to make the significance vanish, the limited number of species investigated is likely to have influenced the results. Although we acknowledge that gathering data on many bird species may be laborious, careful attention should be paid to the number of species included in future comparative analyses of avian visual systems.

More interestingly, Avilés and Soler's (2009) study raises the question of whether assuming strong phylogenetic inertia is appropriate when inferring the vision system of bird species. This question has become crucial given the increasing interest in research in visual ecology. To our knowledge, the shift from an ancestral VS to a derived UVS vision has occurred independently at least six times within the limited set of bird species investigated so far: in the blue-crowned trogon *Trogon curucui* (Carvalho *et al.*, 2007), in the common rhea *Rhea americana*, in hummingbirds (Chen & Goldsmith, 1986; Herrera *et al.*, 2008) and once each in passerines, parrots and Laridae (Ödeen & Håstad, 2003). Further, both UVS (in *Larus spp.* and *Rissa tridactyla*) and VS (in *Sterna spp.*) vision has been reported in the Laridae (Håstad *et al.*, 2005). Thus, UVS vision seems to be a phylogenetically labile character

presenting limited inertia, and additional instances of UVS evolution are likely to be detected as the number of investigated species increases. Hence strong phylogenetic inertia should not be assumed for the evolution of avian visual systems, limiting the usefulness of parsimony-based inference in this context. So, how should we classify species as possessing a UVS or VS visual system before conducting comparative analyses?

An increased ability to absorb UV wavelengths entails higher risks of damaging retinal tissues (Vannorren & Schellekens, 1990). Therefore, the cost associated with the maintenance of UVS visual systems coupled with the occurrence of six independent shifts from a VS to a UVS state and the lack of evidenced cases of reversals, all support an adaptive function of the UVS visual system. According to the sensory-drive theory (Ryan *et al.*, 1990; Endler, 1992), abiotic factors or inter-specific interactions would be the primary selective agents in the evolution of UVS pigments. Ecological factors could therefore serve as an alternative proxy to phylogeny for inferring visual systems when the goal is to investigate the use of visual sense for intra-specific purposes, as in Avilés and Soler (2009). Unfortunately, there is as yet no evidence of a clear relationship between visual ecology and the evolution of UVS pigments (Hart & Vorobyev, 2005). Two studies have compared the performance of UVS and VS eyes (Vorobyev *et al.*, 1998; Schaefer *et al.*, 2007). Both predicted higher performance of UVS eyes in the discrimination of particular colours and one (Schaefer *et al.*, 2007) suggests an advantage of UVS vision for fruit foraging. However, both studies used simulated visual systems and were as such not true comparative studies based on real UVS and VS species. Besides, a study investigating the correlative evolution of the avian visual system with several ecological traits across the bird phylogeny is still lacking. Therefore, given current knowledge, ecological factors cannot be used to classify visual systems in birds.

Until recently, information on the spectral sensitivities of cone types, data necessary to determine the visual system, could only be acquired by microspectrophotometric measurements (Hart *et al.*, 1999). This technique has a limited application in the field of evolutionary ecology because it requires the sacrifice of substantial numbers of birds, which is both frequently impossible and ethically questionable. An alternative technique has arisen recently with the discovery that key mutational changes in the sequence of the SWS1 opsin gene are responsible for the shift from the VS to the UVS state (Wilkie *et al.*, 2000). Ödeen and Håstad (2003) and Ödeen *et al.* (2009) provide a set of primers that allow fast and easy amplification of the sequence fragment presenting the key mutation sites. Thus it is now easy and relatively cheap to obtain direct, non-invasive evidence of the visual system of any bird

species. It is thus surprising that sequencing of the SWS1 opsin gene has so far never been used as a preliminary step in any study in avian visual ecology.

A final issue raised by Avilés and Soler's study concerns the violation of assumptions about the model of visual system evolution when applying comparative phylogenetic methods. The authors assumed a Brownian motion (BM) model of visual system evolution in both the comparative methods they used. Several studies have demonstrated that caution is required when applying phylogenetic correction methods with a BM model when the evolution of the character investigated seriously violates the model of purely random drift evolution with no selection (e.g. Díaz-Uriarte & Garland, 1996; Price, 1997; Harvey & Rambaut, 2000). Thus, the BM model is probably not the most appropriate if avian visual systems are assumed to be either adaptatively constrained, or, alternatively, phylogenetically constrained (Martins *et al.*, 2002). Further, comparative methods do not remove the effects of phylogeny from data correlation, but simply reduce the Type I error (Rohlf, 2006). The rate of Type I error is linked to the quality of the fit between the predicted phylogenetic distribution of character states according to the model considered and their actual distribution. When coding a discrete character like the vision system as a continuous character, the fit will be estimated better if several character shifts are analysed (Díaz-Uriarte & Garland, 1996). Consequently, the one or two events of visual system shift (depending on whether you consider our own or Avilés and Soler's classification, respectively) encompassed in the data set are too limited to allow the application of such phylogenetic comparative methods. Indeed, when few events of trait shift are included, results of correlation tests reflect the correlation with phylogeny more than the correlation with other traits, even if phylogenetic effects are accounted for.

The comparative study of Avilés and Soler (2009) is, to our knowledge, the first that specifically aimed at investigating performance differences between UVS and VS species. This study, as many other comparative studies in visual ecology, is based on several assumptions that we have here shown to be unreliable. As violation of these assumptions may have drastic consequences, the interesting hypothesis that selection has resulted in nestlings presenting increased detectability by feeding parents through tuning of their gape colouration to their parents visual performance, has yet to be tested.

Finally, because shift in colour vision is a discrete event, we encourage researchers working on visual ecology to sample species in a way that maximizes the number of independent events of visual system shift. We also recommend the sequencing of the SWS1

opsin gene to characterize the visual system of investigated species. Although constraining, these precautions are necessary to reach reliable conclusions on the evolution of avian vision and avian colouration.

References

- Avilés, J. M., Perez-Contreras, T., Navarro, C. & Soler, J. J. 2008. Dark nests and conspicuousness in color patterns of nestlings of altricial birds. *Am. Nat.* **171**: 327-338.
- Avilés, J. M. & Soler, J. J. 2009. Nestling colouration is adjusted to parent visual performance in altricial birds. *J. Evol. Biol.* **22**: 376-386.
- Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. 2004. Phylogeny and diversification of the largest avian radiation. *Proc. Natl Acad. Sci. USA.* **101**: 11040-11045.
- Benites, P., Tubaro, P. L., Lijtmaer, D. A., Lougheed, S. C. & Eaton, M. D. 2007. Color differences among closely related species of red-breasted meadowlarks (*Sturnella*). *Condor.* **109**: 605-616.
- Bennett, A. T. D. & Cuthill, I. C. 1994. Ultraviolet vision in birds - What is its function. *Vision Res.* **34**: 1471-1478.
- Bennett, A. T. D. & Thery, M. 2007. Avian color vision and coloration: Multidisciplinary evolutionary biology. *Am. Nat.* **169**: S1-S6.
- Bowmaker, J. K. & Martin, G. R. 1978. Visual pigments and color vision in a nocturnal bird, *Strix aluco* (Tawny owl). *Vision Res.* **18**: 1125-1130.
- Bridge, E. S., Hylton, J., Eaton, M. D., Gamble, L. & Schoech, S. J. 2008. Cryptic plumage signaling in *Aphelocoma* Scrub-Jays. *J. Ornithol.* **149**: 123-130.
- Carvalho, L. S., Cowing, J. A., Wilkie, S. E., Bowmaker, J. K. & Hunt, D. M. 2007. The molecular evolution of avian ultraviolet- and violet-sensitive visual pigments. *Mol. Biol. Evol.* **24**: 1843-1852.
- Cassey, P., Ewen, J. G., Blackburn, T. M., Hauber, M. E., Vorobyev, M. & Marshall, N. J. 2008a. Eggshell colour does not predict measures of maternal investment in eggs of *Turdus* thrushes. *Naturwissenschaften.* **95**: 713-721.
- Cassey, P., Honza, M., Grim, T. & Hauber, M. E. 2008b. The modelling of avian visual perception predicts behavioural rejection responses to foreign egg colours. *Biol. Lett.* **4**: 515-517.
- Chen, D. M. & Goldsmith, T. H. 1986. 4 Spectral classes of cone in the retinas of birds. *J. Comp. Physiol. A* **159**: 473-479.
- Cuthill, I. C., Partridge, J. C., Bennett, A. T. D., Church, S. C., Hart, N. S. & Hunt, S. 2000. Ultraviolet vision in birds. *Adv. Stud. Behav.* **29**: 159-214.

- Díaz-Uriarte, R. & Garland Jr, T. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* **45**:27–47.
- Doucet, S. M., Mennill, D. J. & Hill, G. E. 2007. The evolution of signal design in manakin plumage ornaments. *Am. Nat.* **169**: S62-S80.
- Eaton, M. D. 2007. Avian visual perspective on plumage coloration confirms rarity of sexually monochromatic North American passerines. *Auk*. **124**: 155-161
- Endler, J. A. 1992. Signals, signal conditions, and the direction of evolution. *Am. Nat.* **139**: S125-S153.
- Endler, J. A. & Mielke, P. W. 2005. Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc.* **86**: 405-431.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* **125**: 1-15
- Garland, T., Harvey, P. H. & Ives, A. R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* **41**: 18-32.
- Goldsmith, T. H. & Butler, B. K. 2003. The roles of receptor noise and cone oil droplets in the photopic spectral sensitivity of the budgerigar, *Melopsittacus undulatus*. *J. Comp. Physiol. A* **189**: 135-142.
- Gomez, D. & Thery, M. 2007. Simultaneous crypsis and conspicuousness in color patterns: Comparative analysis of a neotropical rainforest bird community. *Am. Nat.* **169**: S42-S61.
- Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K. L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C. & Yuri, T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**: 1763-1768.
- Hart, N. S. 2001. Variations in cone photoreceptor abundance and the visual ecology of birds. *J. Comp. Physiol. A* **187**: 685-697.
- Hart, N. S., Partridge, J. C. & Cuthill, I. C. 1999. Visual pigments, cone oil droplets, ocular media and predicted spectral sensitivity in the domestic Turkey (*Meleagris gallopavo*). *Vision Res.* **39**: 3321-3328.
- Hart, N. S. & Vorobyev, M. 2005. Modelling oil droplet absorption spectra and spectral sensitivities of bird cone photoreceptors. *J. Comp. Physiol. A* **191**: 381-392.
- Harvey, P. H. & Rambaut, A. 2000. Comparative analyses for adaptive radiations. *Philos. Trans. R. Soc. B.* **355**:1599–1605.

- Håstad, O., Ernstdotter, E. & Ödeen, A. 2005. Ultraviolet vision and foraging in dip and plunge diving birds. *Biol. Lett.* **1**: 306-309.
- Håstad, O. & Ödeen, A. 2008. Different ranking of avian colors predicted by modeling of retinal function in humans and birds. *Am. Nat.* **171**: 831-838.
- Herrera, G., Zagal, J. C., Diaz, M., Fernandez, M. J., Vielma, A., Cure, M., Martinez, J., Bozinovic, F. & Palacios, A. G. 2008. Spectral sensitivities of photoreceptors and their role in colour discrimination in the green-backed firecrown hummingbird (*Sephanoides sephaniodes*). *J. Comp. Physiol. A* **194**: 785-794.
- Huth, H. H. & Burkhard, D. 1972. Spectral visual range of *Colibri serrirostris*. *Naturwissenschaften.* **59**: 650-650.
- Koivula, M., Korpimäki, E. & Viitala, J. 1997. Do Tengmalm's owls see vole scent marks visible in ultraviolet light? *Anim. Behav.* **54**: 873-877.
- Langmore, N. E., Stevens, M., Maurer, G. & Kilner, R. M. 2009. Are dark cuckoo eggs cryptic in host nests? *Anim. Beh.* **78**: 461-468.
- Martins, E. P., Diniz-Filho, J. A. F. & Housworth, E. A. 2002. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution.* **56**: 1-13.
- Ödeen, A., Hart, N. S. & Håstad, O. 2009. Assessing the use of genomic DNA as a predictor of the maximum absorbance wavelength of avian SWS1 opsin visual pigments. *J. Comp. Physiol. A* **195**: 167-173.
- Ödeen, A. & Håstad, O. 2003. Complex distribution of avian color vision systems revealed by sequencing the SWS1 opsin from total DNA. *Mol. Biol. Evol.* **20**: 855-861.
- Price, T. 1997. Correlated evolution and independent contrasts. *Philos. Trans. R. Soc. B.* **352**: 519-529.
- R Development Core Team. 2008. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.
- Rohlf, F. J. 2006. A comment on phylogenetic correction. *Evolution.* **60**: 1509-1515.
- Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. 1990. Sexual selection for sensory exploitation in the frog *Physalaemus pustulosus*. *Nature.* **343**: 66-67.
- Schaefer, H. M., McGraw, K. & Catoni, C. 2008. Birds use fruit colour as honest signal of dietary antioxidant rewards. *Funct. Ecol.* **22**: 303-310.
- Schaefer, H. M., Schaefer, V. & Vorobyev, M. 2007. Are fruit colors adapted to consumer vision and birds equally efficient in detecting colorful signals? *Am. Nat.* **169**: S159-S169.

- Uy, J. A. C. & Endler, J. A. 2004. Modification of the visual background increases the conspicuousness of golden-collared manakin displays. *Beh. Ecol.* **15**: 1003-1010.
- Vannorren, D. & Schellekens, P. 1990. Blue-light hazard in rat. *Vision Res.* **30**: 1517-1520.
- Vorobyev, M., Osorio, D., Bennett, A. T. D., Marshall, N. J. & Cuthill, I. C. 1998. Tetrachromacy, oil droplets and bird plumage colours. *J. Comp. Physiol. A* **183**: 621-633.
- Wilkie, S. E., Robinson, P. R., Cronin, T. W., Poopalasundaram S., Bowmaker J. K. & Hunt D. M. 2000. Spectral tuning of avian violet- and ultraviolet-sensitive visual pigments. *Biochemistry.* **39**: 7895-7901.
- Wilson, R. E., Eaton, M. & McCracken, K. G. 2008. Color divergence among Cinnamon Teal (*Anas cyanoptera*) subspecies from North America and South America. *Ornitol. Neotrop.* **19**: 307-314.
- Yokoyama, S., Radlwimmer, F. B. & Blow, N. S. 2000. Ultraviolet pigments in birds evolved from violet pigments by a single amino acid change. *Proc. Natl Acad. Sci. USA.* **97**: 7366-7371.

Figure caption

Figure 1. Phylogenetic relationships among species used to compute phylogenetic independent contrasts. The visual system classification used by Avilés & Soler is indicated along with the most parsimonious classification used in the reanalysis. In the most parsimonious classification, nocturnal birds have lost the SWS1 opsin protein (event n°1) and UVS pigments in passerine birds have a single origin (event n°2). The Suboscines lineage has been added to the tree with a dotted branch to highlight the lack of UVS pigments in passerine birds outside the Passerida lineage (#: J.P. Renoult, unpublished data based on molecular markers). * indicates species belonging to a family for which molecular or microspectrophotometric data indicating the visual system is available for at least one species (Ödeen & Håstad, 2003).

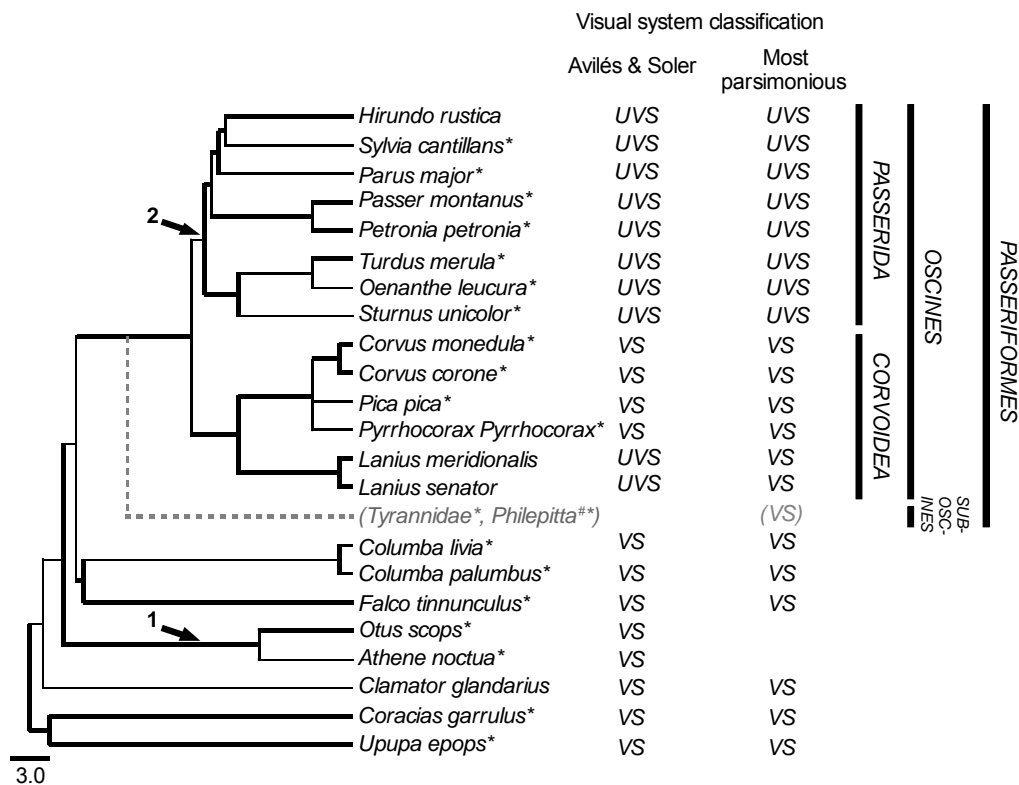


Table 1. Results of general linear models, including variables defining gape colouration (PC1 and PC2 for flanges and mouth) as dependent variables, and nesting site and parental visual system as independent variables. In the most parsimonious classification, *Lanius spp.* were treated as VS species and owls were removed. Contrasts refer to phylogenetic independent contrasts (see text for details).

variables	F	P-value	PC1 (flange)		PC2 (flange)		PC1 (mouth)		PC2 (mouth)		
			t	P-value	t	P-value	t	P-value	t	P-value	
Raw data with Avilés & Soler classification											
<i>Nesting site</i>	9.1313	0.0005	3.885	0.0009 [#]	0.358	0.7240	3.961	0.0008	1.554	0.1367†	
<i>Parental visual system</i>	8.0494	0.0009	4.054	0.0006 [#]	0.53	0.6020	1.899	0.0728	4.278	0.0004†	
Raw data with the most parsimonious classification											
<i>Nesting site</i>	8.2218	0.0013	3.551	0.0025	0.167	0.8690	3.929	0.0011	0.221	0.8276	
<i>Parental visual system</i>	5.9512	0.0052	3.936	0.0011	0.087	0.9320	1.619	0.1238	3.48	0.0029	
Contrasts with Avilés & Soler classification											
<i>Nesting site</i>	8.7178‡	0.0006	3.776	0.0013	0.715	0.4830	4.014	0.0007	1.649	0.1160	
<i>Parental visual system</i>	2.6772‡	0.0698	1.824	0.0839	0.362	0.7220	0.943	0.3576	2.54	0.0200	
Contrasts with <i>Lanius</i> as UVS and nocturnal birds removed											
<i>Nesting site</i>	8.1628	0.0013	3.945	0.0010	0.591	0.5620	3.913	0.0011	1.468	0.1603	
<i>Parental visual system</i>	2.0984	0.1353	1.533	0.1437	0.425	0.6760	0.803	0.4331	2.419	0.0270	
Contrasts with the most parsimonious classification											
<i>Nesting site</i>	8.0081	0.0014	3.653	0.0020	0.56	0.5830	4.181	0.0006	0.992	0.3350	
<i>Parental visual system</i>	1.7175	0.2020	1.788	0.0917	0.005	0.9960	1.196	0.2482	1.655	0.1160	
Standardized contrasts with the most parsimonious classification											
<i>Nesting site</i>	12.1609	0.0002	3.382	0.0035	-1.377	0.1860	3.036	0.0075	2.228	0.0397	
<i>Parental visual system</i>	1.1513	0.3733	0.796	0.4368	0.249	0.8060	-0.394	0.6983	1.45	0.1653	

[#], † and ‡ indicate lines that are inverted in the table provided in Avilés & Soler (2009).

**An isolated population of *Podarcis vaucheri* (Sauria: Lacertidae)
in south-eastern Spain: genetic data suggest human-mediated
range expansion**

Julien P. Renoult¹, Philippe Geniez², Menad Beddek¹, Pierre-André Crochet^{1,3}

¹ CNRS-UMR 5175, Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende,
34293 Montpellier cedex 5, France

² EPHE-UMR 5175, Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende,
34293 Montpellier cedex 5, France

³ E-mail: pierre-andre.crochet@cefe.cnrs.fr

Corresponding author: Pierre-André Crochet

Short Title: An isolated population of *Podarcis vaucheri*

Abstract

In this study, we investigate the identity and origin of a population of Iberian Wall Lizards (*Podarcis hispanicus* complex) that we recently discovered inside the distribution range of the nominotypical form *P. hispanicus*. In the field, these animals were clearly identified as different from this species but their identity remained problematic. Using morphological and mitochondrial DNA analyses, we here identify this population as *P. vaucheri*, which constitutes a significant range extension for this species. Molecular results suggest a North African origin to this population. However, according to morphological results, these individuals are closer to the Spanish morphotype than to the North African morphotypes of *P. vaucheri*. Taken together, these results suggest a human-mediated introduction as the origin of this population, with local adaptation or phenotypic plasticity responsible of phenotypic convergence.

Key words: Iberian Wall Lizard, Human introduction, mitochondrial Control Region, ND4, morphology.

Thanks to its geographical position as a glacial refuge during the climatic oscillations of the Pleistocene (Hewitt, 1996; Taberlet et al., 1998) and its complex geography, the western part of the Mediterranean Basin is one of the world biodiversity hotspots (Myers et al., 2000). The *Podarcis hispanicus* complex nicely illustrates this diversity, with recent systematic investigations revealing the occurrence of at least nine lineages within the Iberian Peninsula only (Harris et al., 2002; Busack et al., 2005; Pinho et al., 2006; Geniez et al., 2007).

Some of these lineages have long been treated as distinct species (e.g. *Podarcis bocagei* Seoane, 1884) while others have only recently been recognised: *P. carbonelli* Pérez-Mellado, 1981, by Sá-Sousa and Harris (2002); *P. vaucheri* Boulenger, 1905 by Busack et al. (2005); *P. hispanicus* (Steindachner, 1870) by Geniez et al. (2007) and *P. liolepis* (Boulenger, 1905), elevated to species status under the junior synonym *P. atratus* (Boscá, 1916) by Busack et al. (2005). Three other lineages have not been formally elevated to species status yet but several lines of arguments show that there is reproductive isolation between them and with the other species in this complex, even if it is sometimes incomplete, indicating that they constitute valid biological species as well (Harris and Sá-Sousa, 2001; Pinho et al., 2008;). These are the widespread “Type 1A”, “Type 1B” and “Type 2” lineages (Pinho et al., 2006). The ninth lineage, called “*P. hispanica sensu stricto*” in Pinho et al. (2006), has been renamed “Valencia lineage” by Renoult et al. (*in press*) to avoid confusion with the species called *P. hispanicus*. It is currently known by its mitochondrial genome only and is found, among others, in specimens of *P. hispanicus* and *P. liolepis* (Renoult et al., *in press*). The taxon originally carrying this lineage remains to be discovered.

Most of these species have wholly allopatric or parapatric ranges except for *P. bocagei* and *P. carbonelli* which occur in sympatry with another species over most of their range. Based on Geniez et al. (2007) and on additional own unpublished data, *P. vaucheri* and *P. hispanicus* come into contact in southern Spain south and east of the Sierra Nevada where they replace each other abruptly with no reported overlap. However, during a field trip in April 2002, we found a population of *Podarcis* in Alguazas (province of Murcia, Spain), well inside the range of *P. hispanicus* and some 230 km away from the eastern limit of the known distribution of *P. vaucheri*. The morphology of these individuals immediately differentiated them from *P. hispanicus* but was somewhat reminiscent of *P. vaucheri* (Fig. 1). In this contribution, we use morphological and mitochondrial DNA tools to elucidate the identity of this isolated population and discuss its origin.

Three adult males (M) and three adult females (F) were collected in Alguazas, along the Rio de Segura [GPS: 38.056°N/1.237°W] on April 19th 2002, and conserved in 95% alcohol in the “Biogéographie et Ecologie des Vertébrés” (BEV) collection now at the Centre d’Ecologie Fonctionnelle et Evolutive (Montpellier, France) (BEV.7317-BEV.7323). Total genomic DNA was extracted for the six specimens. A 460 bp fragment of the mtDNA control region (CR; primers described in Crochet et al. 2004) and a 795 bp fragment of mtDNA comprising the 3’ end of the NADH dehydrogenase subunit 4 gene and adjacent tRNAs (ND4; primers described in Arévalo et al. 1994) were amplified and sequenced according to the protocols provided in Pinho et al. (2006). We built a first sequence alignment with the two fragments concatenated and adding sequences retrieved from Pinho et al. (2006) that represent each of the major mitochondrial lineages documented in the *Podarcis* genus in Iberia and North Africa. We built a second alignment with ND4 sequences only, including our own data and sequences retrieved from Pinho et al. (2007) that represent each of the nine geographic haplogroups documented in *P. vaucheri*. Sequences not published previously have been deposited in GenBank (Accessions Numbers FJ764794 and FJ764795). We performed Maximum Likelihood (ML) analyses using the heuristic search algorithm PHYML 2.4.4 (Guindon and Gascuel, 2003) and 100 bootstrap replications (BS) to assign haplotypes of Alguazas to one of several known mitochondrial lineages.

Introgression of foreign mtDNA into a species’ nuclear background is a recurrent phenomenon in animals (Mallet, 2005) that has recently been demonstrated in the *Podarcis hispanicus* complex as well (Renoult et al, *in press*) and evidently prevents specific identification based on mtDNA only. To avoid falsely identifying the Alguazas specimens on the basis of their mtDNA only, we used Discriminant Analyses (DA) to morphologically classify the Alguazas individuals. This procedure allowed a better discrimination between lineages than the use of a Principal Component Analysis (results not shown). If morphological assignment matches the mtDNA genotype, mtDNA introgression would be unlikely.

Discriminant Analysis requires that specimens are classified into groups prior to the analysis. To avoid potential biases originating from this procedure, the Alguazas specimens were entered as supplementary individuals: they were not attributed to any group prior to the analyses, but were attributed to one of the groups using the results of the DA based on the active specimens. Active specimens (the specimens used by the DA to compute discriminant functions among groups) were reference specimens of each of the lineages of the *P. hispanicus* species complex occurring in the southern half of the Iberian Peninsula and in North Africa (Fig. 2).

Reference individuals were chosen as follows: we selected specimens located well inside the recognized distribution range of each lineage (Carretero, 2008) and typical of these lineages morphologically according to the characters outlined by Geniez (2001), Sá-Sousa et al. (2002) and Geniez et al. (2007). Most of these specimens have already been submitted to multivariate morphological analyses (Geniez et al., 2007; Renoult et al., *in press*) and for many of them identification is backed by nuclear genetic data (Renoult et al., *in press*; unpublished data). Twenty-four morphological traits (the same as in Geniez et al., 2007) measured by the same observer (PG) were included in the analyses. Adults of both sexes were pooled together because they share the main traits that separate morphotypes in this group. Because of the weak morphological differentiation between the different lineages of the *P. hispanicus* species complex, DA performed well with no more than two or three reference morphotypes in each analysis.

Busack et al. (2005) reported a morphological differentiation between Iberian and North African *P. vaucheri* specimens and Geniez (2001) outlined some morphological specificity of the specimens corresponding to the “North Moroccan” haplogroup of Pinho et al. (2007) compared to other North African specimens. Since mitochondrial data indicated a north Moroccan *P. vaucheri* origin of the Alguazas population (see results), we first conducted a DA to classify the six individuals from this population to one of the three following reference morphotypes: Iberian *P. vaucheri* (n=265), Moroccan *P. vaucheri* (n=104, excluding specimens from northernmost Morocco) or *P. vaucheri* from the extreme north of Morocco (n=14). We performed a second DA including as reference morphotypes the one selected in the first DA and 174 *P. hispanicus* (sensu Geniez et al., 2007). Then, three additional DA were successively conducted, each time including the reference morphotype selected in the previous DA and one of the three following lineages: *P. hispanicus* “type 2” (n=105), *P. hispanicus* “type 3” (n=52) and *P. carbonelli* (n=17).

The investigated specimens belong to the collections of the Natural History Museum, London, UK (BMNH); Estación Biológica de Doñana, Seville, Spain (EBD); Naturhistorisches Museum, Bâle, Switzerland (NHMB); Muséum national d’Histoire naturelle, Paris, France (MNHN), Naturistorisches Museum Wien, Austria (NMW) and Collection BEV, CEFE-CNRS, Montpellier (France).

The first phylogenetic analysis based on the CR and ND4 fragments concatenated shows that four Alguazas specimens share the same and unique CR+ND4 haplotype (Fig. 3A). The two remaining individuals have the same CR haplotype than other Alguazas

specimens, but they were not successfully amplified for the ND4 fragment and are therefore missing in the phylogenetic analyses. The analysis unambiguously supports (BS: 100%) the branching of this haplotype within the *P. vaucheri* clade. As already reported previously (Harris et al., 2002; Busack et al., 2005), lizards from Iberia and from Morocco (except southernmost populations) form two distinct lineages into the *P. vaucheri* clade. Quite unexpectedly given their geographical origin, the four individuals from Alguazas branch within the clade that includes Moroccan specimens, a result well supported by bootstrap values (BS: 86%). Using the highly variable ND4 fragment alone, Pinho et al. (2007) identified nine geographic haplogroups in *P. vaucheri*. We therefore performed a second phylogenetic analysis with ND4 sequences selected to represent each of these haplogroups in order to refine the geographic origin of the Alguazas specimens (Fig. 3B). This analysis well supports (BS: 79%) the branching of the Alguazas haplotype with the “North Moroccan” haplogroup, constituted with haplotypes all originating from the extreme north of Morocco, near the Gibraltar Strait (Pinho et al., 2007).

Morphologically, Alguazas lizards (except for BEV 7321) fall within the score distribution of the Iberian *P. vaucheri*, but out of the distribution of both groups of *P. vaucheri* from North Africa (Fig. 4). In a second DA, we included *P. vaucheri* from Spain and *P. hispanicus* as reference morphotypes. All six individuals from Alguazas had classification scores that graphically fell within the score distribution of the former but out of the score distribution of the latter (Fig. 5A). For the third DA, we kept Iberian *P. vaucheri* as a reference morphotype and used *P. carbonelli* as a second reference morphotype: a result similar to the previous one was found (Fig. 5B). *P. carbonelli* was then replaced by *P. hispanicus* “type 2” (Fig. 5C). All the individuals from Alguazas but one fell within the distribution of the *P. vaucheri* scores, but out of the distribution of the *P. hispanicus* “type 2” scores. Finally, a similar result was found when *P. hispanicus* “type 2” was replaced by *P. hispanicus* “type 3” (Fig. 5D).

These results unambiguously identified the Alguazas lizards as *P. vaucheri* according to their morphology, adding further support to their identification. Moreover, because they fall within the middle of the graphical distribution of *P. vaucheri* and not particularly close to the distribution of *P. hispanicus*, individuals from Alguazas do not seem to present any traces of morphological introgression from surrounding *P. hispanicus* populations. This result is consistent with the absence of gene flow between *P. hispanicus* and *P. vaucheri* (Pinho et al. 2008) and highlights the cogency of their proposed specific status.

The current distribution of *P. vaucheri* in Iberia coincides with the Betic region, an area that was isolated from the rest of the Iberian Peninsula during the opening of the Gibraltar Strait, some 7 My ago (Busack et al., 2005). The Alguazas population is located well outside the Betic region and is located inside the range of *Podarcis hispanicus* sensu stricto. This is unusual as the only instances of range overlap in this complex involve the niche-divergent *Podarcis carbonelli* or *Podarcis bocagei* that occur in sympatry with “Type 2”, “Type 1” and *P. vaucheri* for the former and “Type 1” for the latter. These lines of evidence suggest that Alguazas is not part of the natural distribution of *P. vaucheri*. This is well supported by the mtDNA haplotype of the Alguazas specimens which is clearly of North African origin. Last, the complete lack of variation among the Alguazas specimens for the ND4 sequences is compatible with a recent founding by a small number of specimens. We therefore conclude that this population probably results from a human-mediated range expansion of North African lizards.

It is nevertheless surprising to find that Alguazas specimens are morphologically more similar to the Iberian *P. vaucheri* morphotype than to the two Moroccan morphotypes based on the characters we used. Three hypotheses could explain this result. First, the Alguazas population could be Iberian *P. vaucheri* with an introgressed north Moroccan mitochondrial haplotype. This hypothesis is hardly tenable given the location of Alguazas far from the natural range of *P. vaucheri* in Iberia and inside the range of *P. hispanicus*, and would still imply human-mediated introduction of Northern Moroccan lizards to explain the occurrence of a Northern Moroccan haplotype. Alternatively, rapid morphological change through local adaptation to a novel environment or through phenotypic plasticity could explain this result. Phenotypic plasticity of pholidosis or coloration has been seldom reported in lizards, but several studies in lizards documented instances where morphological divergence was more rapid than mitochondrial divergence (Losos et al., 1997; Vitt et al., 1997). Particularly, a human-mediated island colonization of five pairs of *Podarcis siculus* (Rafinesque-Schmaltz, 1810) in the eastern Mediterranean region leads to a pronounced behavioural and morphological local adaptation in less than 40 years, that is to say about 30 generations (Herrel et al., 2008). Introduction of North African *P. vaucheri* in southern Spain could potentially be much older: Iberian populations of *Chamaeleo chamaeleon* (Linnaeus, 1758) for example were probably introduced 4000-5000 years ago (Talavera and Sanchiz, 1985). There could thus have been ample time for morphological divergence from the ancestral population. A last explanation is that our sampling of North African specimens failed to capture the whole morphological variation of the source populations.

Confidently disentangling between these various hypotheses would necessitate an exhaustive investigation of the nuclear DNA variation in the Alguazas population and in the Iberomaghrebian lineages of *P. vaucheri*. We have nevertheless shown that a human-mediated range expansion is the most likely scenario to explain the presence of the *P. vaucheri* population in Alguazas. Human introductions have been well-documented in many small lacertids. Among other examples, *Teira dugesii* Milne-Edwards, 1829 has reached the Azores (Brehm et al., 2003) and mainland Portugal (Sá-Sousa, 1995) from Madeira, *P. pityusensis* (Boscá, 1883) has reached Mallorca and, most spectacularly *P. siculus* has established populations in Corsica, southern France, Spain (Podnar et al., 2005) and even North America (Oliverio et al., 2001). Although such a recent introduction may be regrettable in term of natural ecosystem conservation (Shea and Chesson, 2002), the comparison of a natural and ancient contact zone with an artificial and recent one may offer opportunities to study evolutionary phenomena such as the development of reinforcement, or to check for reproductive isolation. Eventually, a detailed investigation of the ecological factors characterizing each region that accommodates a distinct morphotype of *P. vaucheri* should help understanding the role of local adaptation or phenotypic plasticity in shaping the great morphological variability occurring among the Iberian Wall lizard species complex.

Acknowledgments

We thanks M.-A. Renoult, M.-P. Dubois and two anonymous referees for constructive comments on an earlier version of the manuscript. J.P. Renoult was supported by a grant from the French Ministry of Research.

Literature Cited

- Arévalo, E., Davis, S.K., Sites, J.W. (1994): Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. *Syst. Biol.* **43**: 387-418.
- Brehm, A., Jesus, J., Spinola, H., Alves, C., Vicente, L., Harris, D.J. (2003): Phylogeography of the Madeiran endemic lizard *Lacerta dugesii* inferred from mtDNA sequences. *Molecul. Phylogenet. Evol.* **26**: 222-230.
- Busack, S.D., Lawson, R., Arjo, W.M (2005): Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphibia-Reptilia.* **26**: 239-256.
- Carretero, M.A. (2008): An integrated assessment of a group with complex systematics: the iberomaghrebian lizard genus *Podarcis* (Squamata, Lacertidae). *Integr. Zool.* **4**: 247-266.
- Crochet, P.-A., Chaline, O., Surget-Groba, Y., Debain, C., Cheylan, M. (2004): Speciation in mountains: phylogeography and phylogeny of the rock lizard genus *Iberolacerta* (Reptilia: Lacertidae). *Mol. Phylogenet. Evol.* **30**: 860–866.
- Galán, M. (2003): Lagartija de Bocage *Podarcis bocagei* in *Enciclopedia Virtual de los Vertebrados Españoles* (L. M. Carrascal, and A. Salvador, eds.). Museo Nacional de Ciencias Naturales www.vertebradosibericos.org, Madrid.
- Geniez, P. (2001): Variation géographique des lézards du genre *Podarcis* (Reptilia, Sauria, Lacertidae) dans la péninsule Ibérique, l’Afrique du Nord et le sud de la France. Diplôme EPHE, Université Montpellier II, 338 p.
- Geniez, P., Cluchier, A., Sá-Sousa, P., Guillaume, C.P., Crochet, P.-A. (2007): Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetol. J.* **17**: 69-80.
- Guindon, S., Gascuel, O. (2003): A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**: 696-704.
- Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C., Ferrand, N. (2002): Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *J Biogeogr.* **29**: 1257-1262.
- Harris, D.J., Sá-Sousa, P. (2001): Species distinction and relationships of the western iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetol. J.* **11**: 129-136.
- Herrel, A., Huyghe, K., Vanhooydonck, B., Backeljau, T., Breugelmans, K., Grbac, I., Van

- Damme, R., Irschick, D.J. (2008): Rapid large-scale evolutionary divergence in morphology and performance associated with exploitation of a different dietary resource. *Proc. Nat. Acad. Sci. U.S.A.* **105**: 4792-4795.
- Hewitt, G.M. (1996): Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**: 247-276.
- Losos, J.B., Warheit, K.B., Schoener, T.W. (1997): Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature*. **387**: 70-73.
- Mallet, J. (2005): Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**: 229-237.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Kent, J. (2000): Biodiversity hotspots for conservation priorities. *Nature*. **403**: 853-858.
- Oliverio, M., Burke, R., Bologna, M.A., Wirz, A., Mariottini, P. (2001): Molecular characterization of native (Italy) and introduced (USA) *Podarcis sicula* populations (Reptilia, Lacertidae). *Ital. J. Zoolog.* **68**: 121-124.
- Pinho, C., Ferrand, N., Harris, D.J. (2006): Reexamination of the Iberian and North African *Podarcis* (Squamata : Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Mol. Phylogenet. Evol.* **38**: 266-273.
- Pinho, C., Harris, D.J., Ferrand, N. (2007): Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizards species inferred from mitochondrial DNA variation. *Mol. Ecol.* **16**: 1191-1205.
- Pinho, C., Harris, D.J., Ferrand, N. (2008): Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis spp.*) are an assemblage of incipient species. *Bmc Evol. Biol.* **8**.
- Podnar, M., Mayer, W., Tvrtkovic, N. (2005): Phylogeography of the Italian wall lizard, *Podarcis sicula*, as revealed by mitochondrial DNA sequences. *Mol. Ecol.* **14**: 575-588.
- Renoult, J.P., Geniez, P., Bacquet, P., Benoit, L., Crochet, P.-A: The Necessity of Multi-Marker Analyses for Systematic Studies: the Case of the Iberian Wall Lizard *Podarcis hispanicus* (Squamata, Lacertidae) Species Complex. *Mol. Ecol.* (*in press*).
- Sá-Sousa, P. (1995): The introduced Madeiran lizard, *Lacerta (Teira) dugesii* in Lisbon. *Amphibia-Reptilia.* **16**: 211-214.
- Sá-Sousa, P., Harris, D.J. (2002): *Podarcis carbonelli* Perez-Mellado, 1981 is a distinct species. *Amphibia-Reptilia.* **23**: 459-468.

- Sá-Sousa, P., Vicente, L., Crespo, E.G. (2002): Morphological variability of *Podarcis hispanica* (Sauria: Laccertidae) in Portugal. *Amphibia-Reptilia*. **23**: 55-69.
- Shea, K., Chesson, P. (2002): Community ecology theory as a framework for biological invasions. *Trends Ecol. Evol.* **17**: 170-176.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F. (1998): Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* **7**: 453-464.
- Talavera, R., Sanchiz, B. (1985). Restos holocénicos del camaleón común, *Chamaeleo chamaeleon* (L.) de Málaga. *Bol. R. Soc. Esp. Hist. Nat., Geol.* **81**: 81-84.
- Vitt L., Caldwell J., Zani P., Titus T. (1997): The role of habitat shift in the evolution of lizard morphology: Evidence from tropical *Tropidurus*. *Proc. Nat. Acad. Sci. U.S.A.* **94**: 3828-3832.

Figure 1. Dorso-lateral view of a male specimen (BEV 7323) from Alguazas (Spain) attributed to *Podarcis vaucheri*. Photo: PGe (19.04.2002).

Figure 2. Map of the locations included in the morphological analyses. White squares: *P. hispanicus* sensu stricto; black squares: Iberian *P. vaucheri*; crosses: Moroccan *P. vaucheri* (excluding extreme north Moroccan populations); white triangles: extreme north Moroccan *P. vaucheri*; white diamonds: *P. hispanicus* “type 2”; black circles: *P. hispanicus* “type 3”; black triangles: *P. carbonelli*; grey stars: Alguazas population.

Figure 3. Maximum Likelihood phylogenetic trees illustrating the relationships of the Alguazas haplotype with mitochondrial clades published previously. A. Control Region + DN4 phylogenetic tree including the major mitochondrial lineages of the genus *Podarcis* occurring in the Iberian Peninsula (Pinho et al., 2006: all sequences but those from Alguazas). Clade names are those used in Pinho et al. 2006 except “*P. hispanica* sensu stricto” converted in “*P. hispanicus* Valencia” and “*P. hispanica* Galera” converted in “*P. hispanicus* Valencia” in the present study. B. ND4 phylogenetic tree including the major haplogroups known in *P. vaucheri* (retrieved from Pinho et al. 2007). Values above branches refer to bootstrap supports of nodes; only values above 50% are indicated.

Figure 4. Morphological classification of the specimens from Alguazas to one of three main morphotypes of *P. vaucheri* using a Discriminant Analysis. Iberian (black squares), extreme north Moroccan (triangles) and Moroccan (crosses) specimens are reference groups while Alguazas specimens (grey stars) are supplementary individuals.

Figure 5. Morphological classification of Alguazas specimens to one of the *Podarcis* lineages occurring in the southern half of the Iberian Peninsula (A-D) using Discriminant Analyses. Axes represent the classification scores obtained with a Discriminant Analysis; intervals represent 0.5 score unit. Symbols provide the score for each reference individual (above and below axes) and for Alguazas individual (along axes). Symbols are similar to those used in Figure 2.

Figure 1.

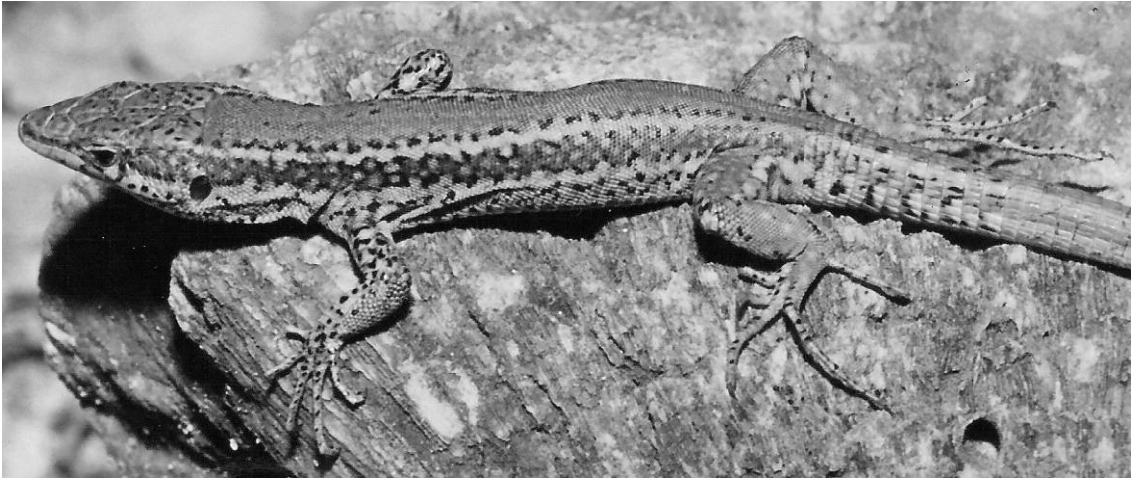


Figure 2.

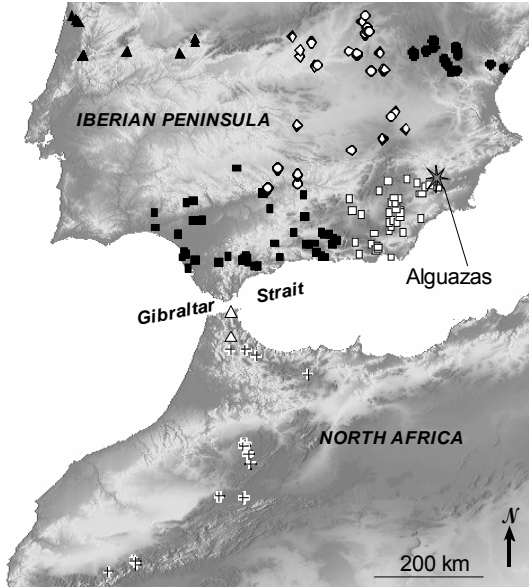


Figure 3.

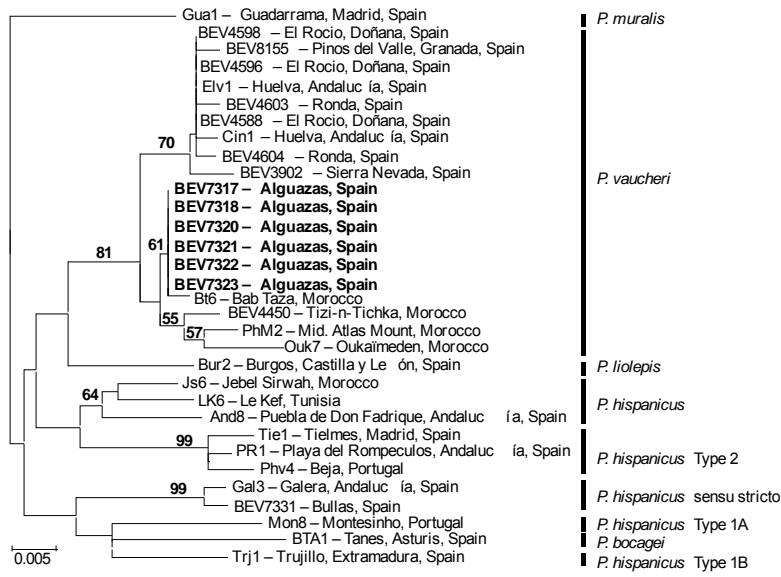


Figure 4.

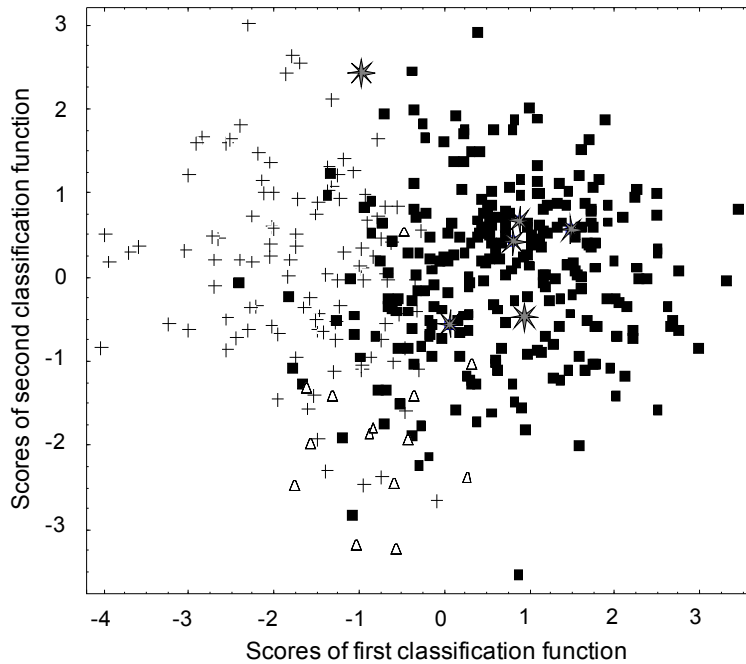
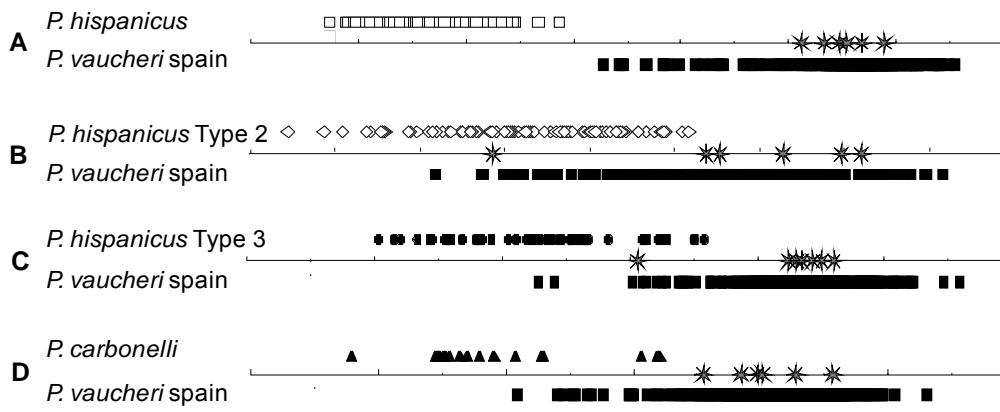


Figure 5.



***Typhlops rajeryi*, a new species of Malagasy blind snake
(Serpentes: Typhlopidae)**

JULIEN P. RENOULT¹ & ACHILLE P. RASELIMANANA²

¹*CNRS-UMR 5175, Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende,
34293 Montpellier cedex 5, France*

²*Département de Biologie Animale, Université d'Antananarivo, BP 906, Antananarivo (101),
Madagascar and Association Vahatra, BP 3972, Antananarivo (101), Madagascar*

Email: julien.renault@cefe.cnrs.fr & araselimanana@vahatra.mg

Running head: *Typhlops rajeryi*: a new species from Madagascar

Abstract

A new species of Typhlopidae is described from Madagascar (Ranomafana, Vatovavy Fitovinany Region, 21°15.388'S 47°25.308'E), based upon a single adult specimen. With the following character combination: 412 paravertebral scales along middorsal line, 24 midbody scale rows, T-V supralabial imbrication pattern, a round dorsal and lateral head shape and a pink coloration, *Typhlops rajeryi* differs from all described species of Typhlopidae. The specimen was opportunistically discovered in the eastern humid forest, close to a national park. We briefly discuss the distribution of other Malagasy *Typhlops* to stress the gap in the knowledge of blind snake diversity in Madagascar.

[*Typhlops rajeryi*, Typhlopidae, new species, blind snake, Distribution, Ranomafana NP, Vatovavy Fitovinany Region, Madagascar]

Madagascar is the third biodiversity hotspot in the world with regard to the reptile fauna, after the Caribbean and middle-America (Myers *et al.* 2000). Moreover, 92% of the 365 described species are endemic to the island (Glaw & Vences 2007). Malagasy snakes of the Typhlopidae family do not make exception to this pattern: except the cosmopolitan, widespread and introduced *Ramphotyphlops braminus* (Daudin, 1803), all the other species of the family encountered in Madagascar are endemic to the island. The species are allocated into two genera: *Xenotyphlops* (endemic to Madagascar) and *Typhlops* (with a worldwide distribution). However, when contrasted with other tropical forests worldwide, and given the extraordinarily high rates of micro-endemism characterizing several reptile lineages on Madagascar (Yoder *et al.* 1995), the current species richness in Malagasy blind snakes is relatively low: only eight species are recognized in the genus *Typhlops* and two in the genus *Xenotyphlops* (Glaw & Vences 2007). The opportunistic discovery of a *Typhlops*, that could not be assigned to any named taxon, close to a widely frequented research station of the eastern humid forest highlights the limited prospection efforts in this region. The purpose of this paper is to describe this new taxon and to stress the urge to increase field investigations in the eastern rain forest.

Material and Methods

In December 2007, a specimen of *Typhlops* was collected by JPR at 20 h 30 local time during heavy rain, on Route National 45, some 6 km west of Ranomafana (village), Vatovavy Fitovinany Region, Madagascar, and 100 m from the main entrance of the Ranomafana National Park. The specimen was lying on the tarmac road, unable to move normally, probably having been run over by a bicycle. The specimen naturally died soon after and was therefore collected and stored in 70% ethanol. On 7 March 2008, Frank Glaw exported the specimen to Germany with permit number 063N-EA03/MG08. The specimen is housed in the Zoologische Staatssammlung München (ZST), Munich, Germany. Muscle tissues were sampled and preserved in 98% ethanol.

Snout-vent length (SVL), tail length including apical spine (TL) and other measurements were performed with an electronic caliper to the nearest 0.1 mm. Mid-body and mid-tail widths were measured across a horizontal plane. Longitudinal scale rows were counted at the 10th scales posterior to the head, at mid-body (approximately), and at the 10th scales anterior to the vent; paravertebral scales were counted along the mid-dorsal longitudinal axis from snout tip to tail extremity with rostral and terminal spine excluded; subcaudal counts include the minimum number of mid-ventral scales between the vent and

terminal spine; dorso-caudal counts include dorsal scale above vent level to tail tip and do not include apical spine. Supralabial imbrication pattern follows the classification given in Wallach (1993). Vertebrae were counted from x-ray radiography. Geographical coordinates of localities were determined by global positioning receiver.

Description

Typhlops rajeryi n. sp.

Type— Holotype, ZSM 1589/2008, an adult collected by J.P. Renoult on 27 December 2007 in Madagascar: Vatovavy Fitovinany Region, near the entrance to the Ranomafana National Park, 21°15.388'S, 47°25.308'E, alt.: 918 m. Specimen is deposited at Zoologische Staatssammlung München (ZST), Munich, Germany and bears the field number FGZC 3165 (in Glaw's field number series) from April 2008. The specimen is in very good conservation state. It does not present any external injury, excepted one piece of epidermis and muscle removed for tissue collection. Radiography revealed a fractured pelvis.

Diagnosis.— T-V imbrication pattern of supralabial scales distinguishes *Typhlops rajeryi* from *Xenotyphlops grandidieri* (Mocquard, 1905) (T-0), *Ramphotyphlops braminus* (T-III), *Typhlops decorsei* Mocquard, 1901 (T-III), *T. ocellaris* (T-III) and the Comorian species *T. comorensis* Boulenger, 1889 (T-III). Compared to *T. rajeryi*, *T. microcephalus* (Werner, 1909) has fewer paravertebral scales (77-235 versus 412), a low number of scale rows (anterior: 20; mid-body: 20; posterior: 19-20 versus 22; 24; 22, respectively) and the pupil is not visible (distinct in *T. rajeryi*). *Typhlops domerguei* Roux-Estève, 1980 has fewer paravertebral scales (252-262), fewer subcaudals (6-7 versus 12) and has a short body (SVL: 150-176 versus 265). *Typhlops reuteri* Boettger, 1881 has also fewer paravertebral scales (345-359) and subcaudals (8-9), has a short (198-222) and thin body (mid-body width: 4.5-5.0 versus 7.4). The type of *T. madagascariensis* Boettger, 1877 is longer (SVL: 410), has more paravertebral scales (580) and has a striped pattern (uniform pattern in *T. rajeryi*). *Typhlops arenarius* (Grandidier, 1872) is shorter (SVL: 80-211), thinner (mid-body width: 4.5-5.0) and has ogival head shape in dorsal view and wedged shape in lateral view (rounded head shape in dorsal and lateral view in *T. rajeryi*). *Typhlops mucronatus* Boettger, 1880 has more scale rows (25-28; 24-28; 24-26), more paravertebral scales (488-577), a truncated dorsal head shape and a brown coloration (pink coloration on living specimen, yellow to beige in alcohol for *T. rajeryi*).

Description.— SVL 265 mm, TL 6.8 mm, total length 272 mm, head width (at the eye level) 4.9 mm, anterior body width 5.4 mm, mid-body width 7.4 mm, mid-tail width 5.4 mm, posterior body width 7.2 mm, length/width ratio 36.7, tail as percentage total length 2.5, tail length/width ratio 1.3, subcaudals 12, dorsocaudals 14, paravertebrals 412, anterior body scale rows 22, mid-body scale rows 24, posterior body scale rows 22, mid-body costal twice as broad as deep.

Snout rounded in dorsal and lateral view (Figs. 1B and 1C); rostral narrow (0.36 head width), oval and elongate, 1.3 times as long as broad, truncated posteriorly and not reaching interocular level, nasals nearly meeting at the median line behind rostral; frontal slightly larger than interparietal; supraoculars oblique, twice as wide as deep, as wide as frontal, contact with postnasal 0.3 supraocular wideness; parietals transversally oriented, 3.1 times as wide as deep and as wide as frontal; first occipital noticeably enlarged, 4.8 times as wide as deep; nostril directed laterally, its axis is on a 30° angle, equally distant to rostral and preocular; nasal semi-divided with shallow posterior concavity, inferior nasal suture contacting second supralabial, superior nasal suture extending nostril at a 30° angle across 0.75 of naso-rostral distance; four supralabials with T-V imbrication pattern (second and third supralabial overlap shields posterodorsally to them), first supralabial 0.7 the size of the second, second supralabial approximately as long as the third, third supralabial highest than the second, fourth supralabial the longest and the highest being as high as long, slightly more than twice the height and the length of the thirds. 210 vertebrae. Pink coloration on living specimen, yellow to beige in alcohol.

Etymology.— The species is dedicated to Emile Rajeryarison, an experimented naturalist working at Ranomafana, in honor to his attachment and his outstanding effort and contribution in the improvement of the knowledge on the herpetofauna of the National Park. He has collaborated with many scientific expeditions in the area and across Madagascar.

Ecology.— The type specimen was found close to the Ranomafana National Park and within an estimated 100 m of the entrance into the Talatakely forest. The type specimen was found at 20 h 30 (local time), e.g. two hours after sunset. It was crossing the tarmac road since it was not present on the road 15 minutes before it was collected. The weather was cool (approximately 20°C) and marked with heavy rain that had started for three hours. Secondary forest with dense understory is bordering the tarmac road. Embankment slope is very steep, almost vertical. The local submontane tropical moist forest (900-1200 m) is characterized by a

dense understory and a canopy of 18-35 m in height. The terrain is steep. The soil is notably unfertile and acid with high concentrations in iron and aluminum and low in phosphorus (Johnson 1994). Annual precipitation ranges from 2300-4000 mm, with the rainy season occurring from December to March (Overdorff 1993). Temperatures range from 4-12°C in June-September to 36-40°C in December-February (Wright & Andriamihaja 2003).

Discussion

The opportunistic discovery of this new species underlines the gap in knowledge of Malagasy blind snakes. This is mainly because Malagasy specimens are particularly rare in museums (Roux-Estève, 1980), the lack of field investigation for burrowing species having particular life style and the absence of a deep revisionary taxonomic investigation. According to V. Wallach (pers. com), several new species have been identified in museum collections and remain to be described. The sparse information on Malagasy Typhlopidae diversity can be highlighted by the distribution of the currently eight recognized species (Glaw and Vences 2007): two (*T. decorsei* and *T. arenarius*) have a broad distribution across the island, although they are absent from the eastern region; four species (*T. microcephalus*, *T. mucronatus*, *T. reuteri* and *T. madagascariensis*) occur in the extreme North, with apparent restricted distributions. One species (*T. ocellaris*) has been collected from the north-east and the extreme south-east, sites separated by 1700 km. Surprisingly, only one species (*T. domerguei*) has been described from the eastern humid forests of the island and this is the only species previously known from the well-studied site of Ranomafana.

Given the burrowing habit of these animals, they are particularly difficult to trap. Without any doubt, the heterogeneous distribution of the genus *Typhlops* in Madagascar is directly associated with their rarity in collections, perhaps the choice of sites that have been surveyed and an aspect of seasonality, rather than a true biogeographical pattern. Within the Ranomafana National park, a number of different areas have been extensively surveyed for the herpetofauna, including the use of pitfalls (M. Vences, pers. com.), and few examples of Typhlopidae are known from the park. In addition to the *T. rajeryi* n. sp. specimen, JPR observed two different individuals of *T. domerguei* dug up by workers repairing a trail in a conifer forest (21°16.125'S, 47°27.352'E).

Acknowledgment

We thank Van Wallach for providing data on Malagasy Typhlopidae and constructive comments on the manuscript, Miguel Vences and Frank Glaw for their remarks about the

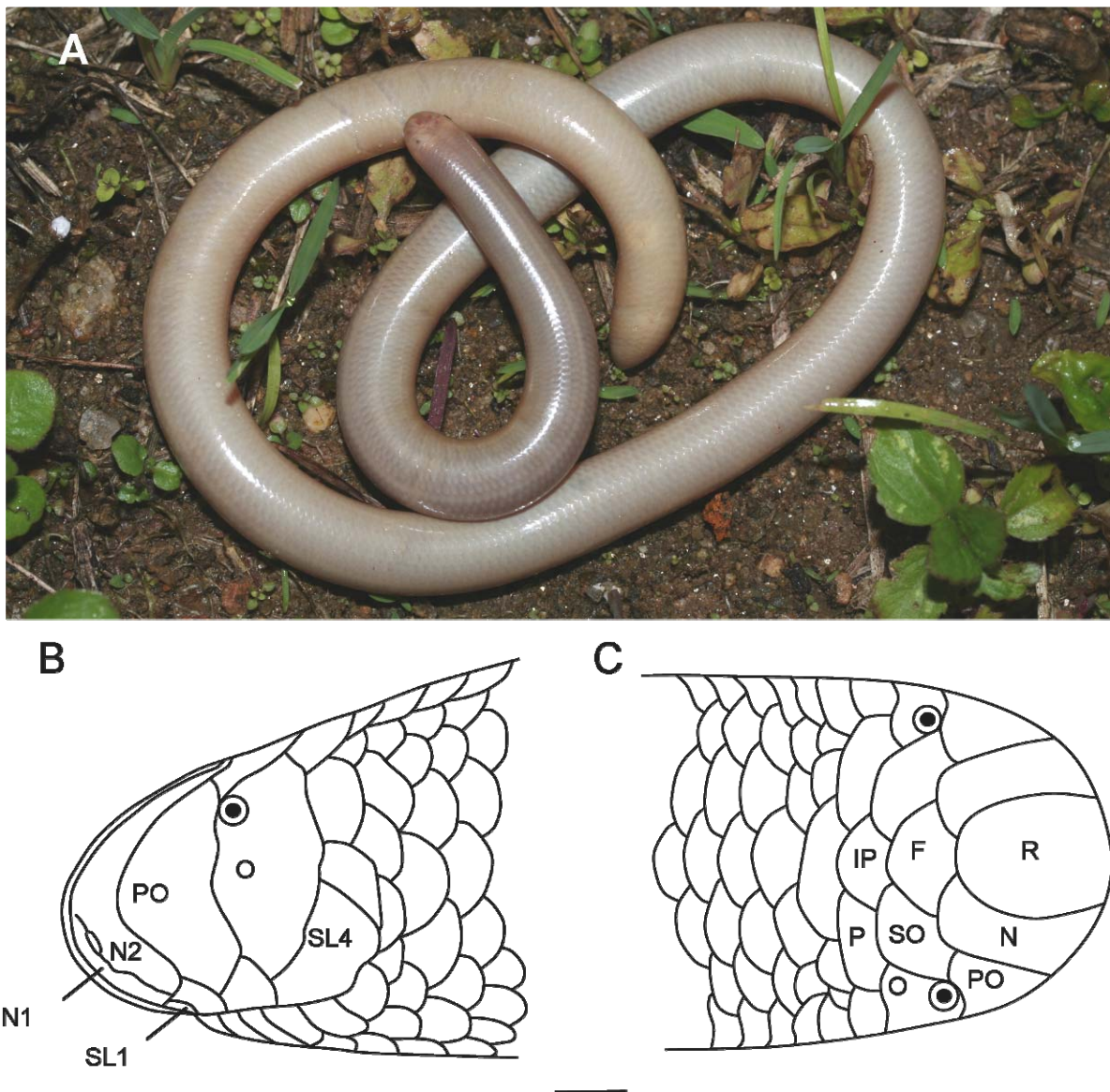
collected specimen, Michael Franzen and the Zoologische Staatssammlung München for loaning the specimen. Romain Cassaigne provided the radiography. Finn Kjellberg, Steve Goodman and Miguel Vences made valuable comments on the manuscript. Finally, we are grateful to the Malagasy authorities for export permits.

Literature cited

- Glaw, F. & Vences, M. (2007) *A field guide to the amphibians and reptiles of Madagascar*. Third edition. Cologne, Vences & Glaw Verlag, 496 pp.
- Johnson, B. (1994) *The soils of southeastern Madagascar*. Ph.D. dissertation. North Carolina State University, Raleigh, NC.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.
- Overdorff, D.J. (1993) Similarities, differences, and seasonal pattern in the diets of *Eulemur rubriventer* and *Eulemur fulvus rufus* in Ranomafana National Park, Madagascar. *International Journal of Primatology*, 14, 721-753.
- Raxworthy, C.J. (2003) Introduction to the reptiles. In: Goodman, S.M. & Benstead, J.P. (Eds), *The natural history of Madagascar*. The University of Chicago Press, Chicago, 934-949.
- Wallach, V. (1993) The supralabial imbrication pattern in the Typhlopoidea (Reptilia: Serpentes). *Journal of Herpetology*, 27, 214-218.
- Wright, P. & Andriamihaja, B. (2003) The conservation value of long-term research: a case study from the Parc National de Ranomafana. In: Goodman, S.M. & Benstead, J.P. (Eds), *The natural history of Madagascar*. The University of Chicago Press, Chicago, 1485-1488.
- Yoder, A.D., Olson, L.E., Hanley, C., Heckman, K.L., Rasoloarison, A.L., Russell, J., Ranivo, V., Soarimalala, V., Karanth, K.P. & Raselimanana, A.P., Goodman, S.M. (2005). A multidimensional approach for detecting species patterns in Malagasy vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 6587-6594.

Figure caption.

Figure 1. *Typhlops rajeryi*, n. sp. Holotype (ZSM1589/2008). (A) Photograph of the freshly dead specimen. Alive, the individual had a pink head coloration gradually lightening toward a whitish caudal coloration. After several months in alcohol, it turned beige-yellow. The red punctures on the rostral part correspond to petechia, probably of accidental origin. (B) Lateral view of the head. (C) Dorsal view of the head. F: frontal, IP: interparietal, N (N1+N2): nasals, P: parietal, PO: preocular, R: rostral, SL1 and SL4: first and fourth supralabial, SO: supraocular. Bar = 1 mm.



RESUME – Les marqueurs cytoplasmiques sont souvent utilisés seuls pour reconstruire des patrons phylogénétiques et phylogéographiques. Les processus évolutifs, démographiques et écologiques à l'origine de ces patrons sont le plus souvent étudiés à l'aide de marqueurs nucléaires. Cependant, certains mécanismes évolutifs, comme l'hybridation introgressive, peuvent rendre l'histoire des gènes discordante de l'histoire des espèces. En outre, les conséquences génétiques des hybridations introgressives peuvent être particulièrement prononcées pour les marqueurs cytoplasmiques par rapport aux marqueurs nucléaires. Pour cette raison, la fiabilité de l'utilisation des marqueurs cytoplasmiques pour la reconstruction des patrons est débattue. Afin d'évaluer l'intérêt respectif des marqueurs cytoplasmiques pour la reconstruction des patrons et pour l'étude des processus, nous avons comparé chez le complexe d'espèces du Lézard ibérique *Podarcis hispanicus* les histoires évolutives inférées avec un marqueur mitochondrial et avec des marqueurs nucléaires, ainsi que chez les figuiers de la section *Galoglychia*, les histoires évolutives inférées avec des marqueurs chloroplastiques et avec des marqueurs nucléaires. Dans les deux exemples, nous avons révélé plusieurs cas de discordances cyto-nucléaires imputables à des événements anciens d'hybridation introgressive. Dans le cas des lézards, une discordance est notamment causée par le remplacement complet du génome mitochondrial d'une espèce, dans une partie de son aire de distribution, par le génome mitochondrial d'une autre espèce aujourd'hui éteinte ou non décrite. Cette observation traduit des déplacements d'aires de distribution des espèces durant les oscillations climatiques du Quaternaire. Dans le cas des figuiers, la congruence entre la phylogénie chloroplastique des plantes et la phylogénie des pollinisateurs suggère que les changements d'hôtes sont la cause à la fois des discordances cyto-nucléaires chez les plantes et des incongruences entre phylogénie nucléaire des plantes et phylogénie des pollinisateurs. Ces deux cas d'études démontrent que les hybridations introgressives rendent limité l'intérêt des marqueurs cytoplasmiques pour la description précise des patrons phylogénétiques et phylogéographiques. En revanche, ces mêmes hybridations introgressives rendent les marqueurs cytoplasmiques particulièrement utiles pour l'étude des processus.

Mots-Clés : Lézard ibérique, *Podarcis hispanicus*, *Ficus*, pollinisateurs, *Galoglychia*, discordance cyto-nucléaire, phylogéographie, phylogénie, taxonomie, systématique, code-barres moléculaire, changement d'hôte, balayage sélectif, mitochondrie, chloroplaste

TITLE – *Using cytoplasmic markers and cyto-nuclear discordances for the study of evolutionary, demographic and ecological processes*

ABSTRACT – Cytoplasmic markers are often used alone to reconstruct phylogenetic and phylogeographic patterns. Evolutionary, demographic and ecological processes that cause these patterns are most often investigated using nuclear markers. However, some mechanisms, like introgressive hybridization, may result in gene histories differing from species histories. Further, genetic consequences of introgressive hybridization may be particularly marked for cytoplasmic markers comparatively to nuclear markers. This is why the reliability of cytoplasmic markers as a tool for pattern reconstruction is currently debated. To evaluate the value of cytoplasmic markers for pattern reconstructions and process studies, we compared, within the species complex of Iberian Wall Lizards *Podarcis hispanicus*, the evolutionary history inferred from a mitochondrial marker with the history inferred from nuclear markers, and in fig species of section *Galoglychia*, the history inferred from chloroplast markers with the history inferred from nuclear markers. In both cases, we evidenced several cases of cyto-nuclear discordances caused by ancient events of introgressive hybridization. In the lizard example, discordance was caused by the complete replacement of the mitochondrial lineage of one species, in a large part of its distribution range, by a mitochondrial lineage belonging to another species currently extinct or not described yet. In the *Ficus* example, congruence between plant chloroplast phylogeny and pollinator phylogeny supports the hypothesis that host shifts are responsible of both cyto-nuclear discordance in plant phylogenies and pollinator-plant phylogenetic incongruence. Both cases demonstrate that, because of introgressive hybridization, cytoplasmic markers are of limited interest to precisely describe phylogenetic and phylogeographic patterns. However, these introgressive hybridization events make cytoplasmic markers useful to investigate processes.

Keywords: Iberian Wall Lizard, *Podarcis hispanicus*, *Ficus*, pollinator, *Galoglychia*, cyto-nuclear discordance, phylogeography, phylogeny, taxonomy, systematics, DNA barcoding, host shift, selective sweep, mitochondria, chloroplast

Discipline : Biologie de l'Evolution et Ecologie

Adresse du laboratoire : CEFE, UMR5175, 1919 route de Mende 34293 Montpellier

Contact de l'auteur : jujurenoult@hotmail.com