



HAL
open science

Molecular Phylogeny of Melanoplinae (INSECTA: ORTHOPTERA: CAELIFERA: ACRIDIDAE)

I. Chintauan-Marquier

► **To cite this version:**

I. Chintauan-Marquier. Molecular Phylogeny of Melanoplinae (INSECTA: ORTHOPTERA: CAELIFERA: ACRIDIDAE). *Ecologie, Environnement*. Université de Grenoble, 2010. Français. NNT : . tel-00580813

HAL Id: tel-00580813

<https://theses.hal.science/tel-00580813>

Submitted on 29 Mar 2011

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

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

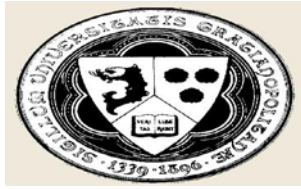


Phylogénie des Melanoplinae (Insecta: Orthoptera: Caelifera: Acrididae)

Présentée et soutenue publiquement par
CHINTAUAN-OANCEA MARQUIER IOANA-CRISTINA
le 15 décembre 2010

THÈSE
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE
Biodiversité - Écologie - Environnement





UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE
Spécialité **Biodiversité – Écologie - Environnement**

Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

CHINTAUAN-OANCEA MARQUIER IOANA-CRISTINA

le **15 décembre 2010**

PHYLOGÉNIE MOLÉCULAIRE DES MELANOPLINAE
(INSECTA: ORTHOPTERA: CAELIFERA: ACRIDIDAE)

Thèse dirigée par **Pompanon François**

Thèse préparée au sein du **Laboratoire d'Écologie Alpine UMR UJF-USavoie-**
CNRS 5553

dans l'**École Doctorale Chimie et Sciences du Vivant**

JURY

Civilité/Nom/Prénom	Fonction et lieu de la fonction	Rôle
Mme. Desutter-Grandcolas Laure	Professeur, Musée National d'Histoire Naturelle, Paris	Rapporteur
M. Douzery Emmanuel	Professeur, Université de Montpellier 2	Rapporteur
M. Nichols A. Richard	Professeur, Queen Mary University of London, UK	Examineur
M. Pérez Serge	Directeur de Recherche CNRS, ESRF Grenoble	Examineur
M. Pompanon François	Maître de Conférence, Université de Grenoble 1	Directeur de thèse

Résumé

La phylogénie moléculaire reconstruit des relations de parenté entre unités évolutives, en se basant sur des changements structurels au niveau moléculaire (ADN, protéines). Elle constitue donc un outil précieux pour déchiffrer l'évolution spatio-temporelle de la biodiversité.

Le présent travail examine l'histoire évolutive d'un groupe de criquets (Insecta: Orthoptera: Caelifera), par le biais de méthodes phylogénétiques (parcimonie, maximum de vraisemblance et bayésienne) et de datation, appliquées à l'étude de séquences d'ADN nucléaire et mitochondrial combinées. Dans un premier temps, nous étudions la sous-famille Melanoplineae (Orthoptera: Acrididae) et l'une de ses tribus, Podismini, pour éclaircir leur histoire évolutive, la resituer dans un contexte paléobiogéographique, et la mettre en relation avec la taxonomie existante. Dans un deuxième temps, les méthodes de reconstruction phylogénétiques et de datation sont appliquées à l'étude de la dynamique de l'évolution concertée au sein de l'espèce *Podisma pedestris*, en analysant le polymorphisme intra- et interindividuels de l'ADN ribosomal, i.e. gènes et pseudogènes d'ITS1.

Mots-clés: Acrididae, Melanoplineae, ADN ribosomal, pseudogène, phylogénie moléculaire, datation, biogéographie, taxonomie.

Abstract

Molecular phylogenies aim to build affiliations between evolutionary units by using their changes at the molecular level (DNA, proteins). Thus, they are a precious tool to understand the evolution biodiversity in space and time of.

The present work analyses the evolutionary history of a grasshopper group (Insecta: Orthoptera: Caelifera), using phylogenetic (parsimony, maximum likelihood and Bayesian) and dating methods on nuclear and mitochondrial DNA combined sequences. First, we study the subfamily Melanoplineae (Orthoptera: Acrididae) and one of its tribe (Podismini) in order to clarify their evolutionary history and taxonomy in a paleobiogeographic context. Then, we use the phylogenetic and dating methods in order to study the dynamics of the concerted evolution of rDNA (i.e. ITS1 genes and pseudogenes) inside the *Podisma pedestris* species.

Keywords: grasshopper, Acrididae, ribosomal DNA, pseudogene, molecular phylogeny, dating, biogeography, taxonomy.

"All observation must be FOR or AGAINST some view

if it is to be of any service."

Charles Darwin

À la Mémoire de Christiane Amédégnato...

In Memoria Christianei Amédégnato...

**À tous les gens qui aiment la Vie et La respectent dans
toutes ses formes...**

***Pentru toți oamenii care iubesc Viața și O respectă în
toate formele Ei ...***

**À ceux qui oublient jamais de rêver et gardent leurs
illusions...**

***Pentru cei care nu uită niciodată să viseze și își
păstrează iluziile...***

À ma chère famille...

Familiei mele dragi...

À mes chers amis ...

Prietenilor mei dragi ...

Remerciements

C'est le chapitre de la thèse le plus personnel, donc le plus difficile à écrire pour moi, d'autant plus que je ne le fais pas dans ma langue maternelle, le Roumain, malgré Chopin qui me soutient immuablement depuis l'âge de mes six ans...

Comment ne pas glisser d'un « Carpe Diem » à un « Memento mori » ? Comment garder l'équilibre entre le « trop » et le « pas assez » dit ? Comment concentrer dans ces quelques lignes tant de passion, d'amitié, de rencontres, de peurs, de connaissances et d'interrogations...qu'on accumule pendant une thèse ? Comment protéger l'unicité des certains gestes et vécus de la banalité d'une écriture dépourvue de talent comme la mienne ?

Je ne sais pas encore..., mais je vais me lancer dans ce récit ayant la conviction que la simple mention des personnes que je voudrais remercier suffit à elle seule pour anoblir ces lignes.

En premier lieu, je voudrais exprimer ma reconnaissance envers les membres du jury, Madame Laure Desutter - Grandcolas et Monsieur Emmanuel Douzery pour avoir accepté d'être les rapporteurs de ma thèse, et Messieurs Richard A. Nichols et Serge Pérez pour avoir accepté d'être présents pour juger mon travail et pour leur soutien sans faille...MERCI!

Mon travail de recherche moléculaire a été effectué principalement au sein du Laboratoire d'Ecologie Alpine (LECA) à Grenoble, sous la

direction des Messieurs Jean Bouvet et Pierre Taberlet à qui j'adresse ici mes sincères remerciements pour m'avoir accueilli.

Dans ce contexte, mes remerciements profonds vont aussi vers mon directeur de thèse, François Pompanon, qui est resté à mes côtés dans les épreuves, comme dans la réussite. On ne remercie pas simplement pour une amitié, on la partage. C'est pareil pour les connaissances scientifiques et autres, comme pour les expériences de vie... François, emprunt de discrétion et tact, m'a aidé à mûrir en étant le «compositeur» de ma vie, et pas seulement son simple «interprète»...MERCI!

Mes tendres pensées vont aussi vers sa famille: Odile et leurs quatre poussins. Merci à vous tous!

Pour rester à Grenoble, MERCI de tout cœur à mes amis et collègues du LECA sans lesquels cette thèse n'aurait pas vu le jour et sa soutenance aurait été dépourvue de tout sens:

Ludo, Christian, Aurélie, Delphine, Carole, Eva, Célia, Myriam, Olivier, Antoine, Christophe, Eric...

Merci de m'avoir accepté parmi vous. Avoir des amis comme vous a été essentiel pour combler le vide qu'accompagne habituellement l'exil...

L'Institut de Recherches Biologiques et l'Université « Babeş-Bolyai » de Cluj-Napoca, par le biais des professeurs Belá Kis, Gheorghe Coldea et Nicolae Tomescu, ainsi que le Laboratoire d'Entomologie du Muséum National d'Histoire Naturelle de Paris, sous la coordination de Madame Christiane Amédégnato et de Monsieur Simon Poulain, ont été les lieux

initiatiques de mes connaissances sur l'évolution morphologiques des acridiens. Mon cœur et ma reconnaissance sont avec eux, ici et ailleurs...

Merci Christiane de m'avoir appris, avec délicatesse et humour, que je ne suis pas seule, que je suis forte si je choisis de l'être. Merci de m'avoir donné la chance d'être à tes côtés pour m'initier, dans les limites de mes capacités, dans l'Acridologie et pas seulement...Tu resteras à jamais une source d'inspiration pour moi!

Meilleures pensées de remerciements à Monsieur Simon Poulain qui m'a guidé avec élégance, discrétion et rigueur scientifique dans le labyrinthe des collections entomologiques du Museum National d'Histoire Naturelle de Paris, comme dans la narration des inoubliables expéditions scientifiques faites avec Christiane dans l'Amérique du Sud...

Richard A. Nichols, Irène Keller, Paris Veltsos et Andrew Leitch m'ont guidé, avec générosité, ingéniosité et patience, vers les mystères de l'évolution de l'ADN ribosomal et de l'évolution « tout court »...

Mon affection va aussi envers Ellie et Joe Nichols qui m'ont adopté avec tant de générosité...

Merci à Raj, Heike, Lim, et Stephen pour leur soutien, scientifique et amical, dans la vie de laboratoire et de terrain...A eux tous, et à tous les autres collègues du « Fogg building » et « Sir John Vane Science Centre », merci de m'avoir honorée avec votre amitié pendant mon séjour au sein du « School of Biological and Chemical Sciences » à l'Université « Queen Mary » et au « Genome Centre at Barts and the London School of Medicine and Dentistry » de Londres.

Profond respect à tous ceux qui le revers de la vie a surpris le 7 juillet 2005 sur la ligne de métro Aldgate East - Liverpool Street à Londres.....

Ma reconnaissance va aussi vers Dr. Benzakin et Dr. G.... Elles m'ont aidé à revenir de loin et m'ont montré que l'Humanité existe toujours et qu'elle rime avec Compassion et Dévotion, malgré les dérives de violence et ignorance, généralisées dans une société présente dépourvue de contenu...

Dans le même esprit, je voudrais remercier les personnes de l'organisation humanitaire « Amnesty International » qui m'ont aidé à me rappeler combien les limites de la souffrance et du supportable sont malheureusement laxistes...Mais aussi, qu'un petit geste de soutien peut amener un grand réconfort et empêcher l'irréparable de se produire....

J'exprime ma tendre affection à mes amis de tous les jours et tous les continents, dans l'ordre et le désordre et sur les accords de la Polonaise en A Mineur Opus 53 de Chopin :

à Joëlle, Stéphanie, Nathalie, Cathy, Agnèse, Uria, Heike, Sylvie, Fabien, Xavier...

aux familles Grapini, Fiscutean et Mocanu (parents et enfants),

à Simona, Ella, Nicoleta, Jancsy, Monica, Florin....

à Thierry Noblecourt, Steve Jordan, Pierre Berthier, Paolo, Haruki, Bill Chapco, Maria-Martha...

et à tant d'autres Amis que je ne mentionne pas ici, par souci de discrétion, mais qui ne sont pas moins importants pour moi!

Mes remerciements les plus sincères s'envolent aussi vers toutes les personnes qui m'ont aidé sur le terrain en Roumanie, France, Italie et Espagne, ainsi vers ceux qui m'ont envoyé du matériel provenant des territoires et/ou collections inaccessibles pour moi. Vous rendre service à mon tour c'est un plaisant devoir...

La finalité temporaire de ce récit et le début perpétuel d'une autre histoire puisent leurs sources dans l'amour que je porte à mes deux pays, la Roumanie et la France, et à mes familles d'Aiud, de Limoux, de Cluj et d'ailleurs.

J'ai une pensée spéciale pour mes proches partis ailleurs: Firuca, Simion, Sever, Aurica, Ana, Maria, Aurel... «Odihnească-se în pace!»

Je remercie à mes Parents et à Luci pour m'avoir appris à bannir l'inertie en tout genre, mais aussi à toujours rêver et avancer.

Jérôme et Elisa, vous donnez le vrai sens de ma vie. Je vous aime et je vous en remercie pour tout ce que vous avez fait et senti pour moi pendant toutes ces années! Cette thèse est aussi la votre!

D...ajută-ne!

Merci à tous,

Ioana.

Fait à Grenoble, jour de Saint Nicolas, 2010.

TABLE DES MATIERES

Chapitre I.

Introduction Générale.....	3
-----------------------------------	----------

Chapitre II.

Matériel et Méthodes.....	8
----------------------------------	----------

II.1 Systématique phylogénétique – le concept d’homologie.....	9
--	---

II.2. Phylogénie moléculaire – notions préliminaires.....	10
---	----

Chapitre III.

Les Melanoplineae (Amédégato, 1974) - présentation du modèle d’étude.....	25
--	-----------

Chapitre IV.

Objectifs de la thèse.....	37
-----------------------------------	-----------

Chapitre V.

Utilisation des marqueurs moléculaires, nucléaires et mitochondriaux, dans l’étude de l’histoire évolutive de criquets (Orthoptera: Acrididae) à deux niveaux taxonomiques: la sous-famille (Melanoplineae) (article A) et la tribu (Podismini) (article B).....	40
---	-----------

V.1. Problématique et démarche scientifique.....	41
--	----

V.2. Histoire évolutive et taxonomie des Melanoplineae.....	43
---	----

V.3. Histoire évolutive des Podismini.....	56
--	----

Chapitre VI.

ADN ribosomal chez le criquet alpin <i>Podisma pedestris</i> (Linnaeus, 1758): pseudogènes et évolution (article C).....	87
---	-----------

VI.1. Problématique et démarche scientifique.....	88
---	----

VI.2. Contribution scientifique.....	89
--------------------------------------	----

Chapitre VII.

Discussion générale - Perspectives.....	104
--	------------

VII.1. La taille du génome et les pseudogènes chez les criquets	105
---	-----

VII.2. L’utilité de la phylogénie moléculaire dans la classification des Melanoplineae.....	106
---	-----

VII.3. Quelques perspectives immédiates	107
---	-----

Bibliographie sélective (hors articles).....	110
---	------------

Annexes.....	122
---------------------	------------



Chapitre I. Introduction générale

*Brâncuși,
La colonne infinie
1918*

Deux révolutions conceptuelles ont marqué la biologie évolutive durant le XX^e siècle: d'une part, le raisonnement en termes de population et d'autre part, le raisonnement en termes d'arbre phylogénétique (e.g. O'Hara 1997, Douzery et al. 2010).

La pensée «populationnelle» s'est largement développée à partir des années 1930, conduisant à l'essor de la biologie et de la génétique des populations. Dans ce cadre, chaque espèce est considérée comme un ensemble de populations changeantes comprenant des individus interféconds (e.g. Douzery et al. 2010).

La pensée «phylogénétique» trouve quant à elle ses racines dans les célèbres arbres évolutifs dressés dans la seconde moitié du XIX^e siècle par Darwin et Haeckel. Cependant, c'est seulement avec Hennig, dans les années 1950 - 1960, que la pensée «phylogénétique» s'est largement répandue, au travers de sa mise en œuvre du concept d'*homologie* dans l'interprétation des relations de parenté entre les taxons (e.g. Hennig 1966). Cette métamorphose conceptuelle touchant aux racines philosophiques mêmes de ce qui était autrefois «l'art des classifications» est désormais connue sous le nom de *systematique phylogénétique* ou *cladistique*.

Avec l'analyse cladistique on découvre ainsi avec effarement, dans les années soixante à présent, que nombreux groupes de la hiérarchie systématique traditionnelle, essentiellement morphologique, considérés jusque là comme naturels, apparaissent comme artificiels, dans la mesure où ils sont paraphylétiques. Dans cette lignée se placent aussi certaines unités taxonomiques d'orthoptères caelifères ou «criquets», comme le sous-ordre *Acridomorpha* (Dirsh 1956), la famille *Acrididae* (MacLeay 1821) ou les sous-familles *Catantopinae* (Brunner von Wattenwyl 1893), *Cyrtacanthacridinae* (Kirby 1902) et *Melanoplinae* (Amédégnato 1974).

Au cours de leur histoire, de Linnaeus (1758) à présent, ces groupes ont changé de composition et ont glissé sur l'échelle verticale de la classification traditionnelle, en épousant des rangs taxonomiques plus ou moins élevés. Dirsh (1956, 1975) expliquait cette instabilité hiérarchique, chez les acridiens, par l'absence de définitions non ambiguës pour chaque

niveau de classification supérieur à celui de l'espèce. A cette «carence théorique» se rajoutent des désavantages pratiques, comme (i) les descriptions de taxons basées exclusivement sur des caractères morphologiques externes, instables et convergents, et (ii) la rareté ou la manque de fossiles, témoins de l'évolution morphologique dans le temps.

Dans ce contexte, l'intégration des données moléculaires (e.g. séquences nucléotidiques et protéiques), dans les analyses phylogénétiques, s'est avérée essentielle pour l'éclaircissement de l'histoire évolutive des organismes étudiés.

Chez les insectes, les études phylogénétiques en cours, qui portent aussi bien sur la recherche de nouveaux marqueurs moléculaires (e.g. Walter et Winterton 2007, Blanchet et al. 2010), que sur les méthodes d'analyse (e.g. Whitfield et Kjer 2008, Ronquist et Deans, 2010 ; Schlick-Steiner et al. 2010), donnent lieu à des publications à un rythme rapide, mais la proportion de publications proposant des acridiens (Insecta: Orthoptera: Acrididae) comme modèle d'étude est assez faible. Il représente moins de 50 publications référencées dans le « Web of Knowledge » depuis 1995. Par exemple, les journaux «*Systematic Biology*» et «*Molecular Phylogenetic and Evolution*» proposent seulement un et, respectivement, huit articles sur la phylogénie des acridiens entre 1997 et 2010. Plusieurs raisons sont possibles, comme:

- la difficulté d'amplifier avec certitude les gènes-cible, et l'absence de fiabilité des résultats, dû au nombre élevé des pseudo-gènes dans un génome de grande taille (e.g., le génome du criquet *Podisma pedestris*, 18000 Mb, qui est cent fois plus grand que celui de *Drosophila* et quatre fois que celui de *Homo sapiens*) (Bensasson et al., 2000);
- l'accumulation du polymorphisme moléculaire ancestral chez les acridiens, en général, et chez les Melanoplinae, en particulier, qui semblent avoir divergé à la même période: fin du Crétacée – début du Tertiaire (e.g., Amédégato, 1977 ; Carbonell, 1977).

C'est dans ce contexte scientifique riche en réflexions méthodologiques, allant de pair avec des questionnements sur l'évolution des acridiens en particulier, et des insectes en général, que se situe mon travail de thèse.

Chapitre II. Matériel et Méthodes



*Picasso,
Sauterelle 1907*

II.1. Systématique phylogénétique - le concept d'homologie

L'apport original de Hennig aux principes de classifications va, paradoxalement, consister à démembrer le concept d'*homologie*, en apparence si efficace et si bien fondé (e.g. Darlu et Tassy 1993, De Ricqlès 1996). Au sein de ce concept, Hennig distingue en effet deux composantes. Dans une perspective évolutionniste, deux taxons quelconques peuvent présenter des caractères homologues pour deux raisons distinctes. Soit les caractères homologues proviennent d'un ancêtre commun lointain et non exclusif, soit, au contraire, ils proviennent d'un ancêtre commun proche et exclusif. Hennig démontre formellement que l'existence de *caractères dérivés partagés*, c'est-à-dire les *synapomorphies*, expriment l'apparement de façon précise et non ambiguë et définissent un *monophylum* ou group naturel au sens strict, c'est à dire cladistique. Il en découle qu'un groupe monophylétique proprement dit, quelle que soit sa richesse taxique et le rang formel qu'on lui reconnaisse, doit être défini comme composé par *l'ancêtre commun et tous ses descendants*.

En revanche, les taxons réunis par la possession de caractères homologues à l'état primitif ou généralisé, autrement dit par *symplesiomorphies*, constituent ce que Hennig distingue sous le nom de *groupes paraphylétiques*. Ces entités, où l'on range *un ancêtre commun mais pas tous ces descendants*, sont des regroupements artificiels, dépourvus de sens évolutif (Hennig 1966).

Cette métamorphose conceptuelle du principe d'homologie est appliquée dans la systématique phylogénétique pour interpréter les relations de parenté, entre les taxons, suggérées par une grande diversité de données morphologiques et moléculaires.

II.2. Phylogénie moléculaire – notions préliminaires

II.2.1 Introduction

En systématique moléculaire, les séquences d'ADN ou protéiniques sont la source principale de données pour estimer les relations évolutives entre organismes vivants. D'une part, les techniques de séquençage automatique permettent aux chercheurs d'accumuler rapidement un grand nombre de séquences d'ADN (plusieurs milliers de bases par heure). D'autre part, les progrès récents en analyse et traitement de séquences, notamment le développement d'algorithmes qui permettent l'alignement multiple des séquences (e.g. Thompson et al. 1997, Hickson et al. 2000) et la mise à disposition de modèles décrivant plus finement les substitutions nucléotidiques (e.g. Lio and Goldman 1998, Felsenstein 2004) permettent un traitement plus précis (e.g. identification des sites homologues, prédiction de la structure secondaire, établir et dater une phylogénie) des données. De plus, de nouvelles méthodes d'inférence phylogénétique (e.g. Lemmon and Milinkovitch 2002, Guindon and Gascuel 2003, Ronquist and Huelsenbeck 2003) permettent d'analyser des ensembles de données plus importants, jusqu'à quelques millions de caractères. Les séquences nucléotidiques sont donc devenues un des marqueurs phylogénétiques les plus utilisés dans la dernière décennie.

Néanmoins, l'inférence phylogénétique basée sur les séquences nucléotidiques n'est pas, malgré des avantages considérables, exempts de certaines difficultés (e.g. Naylor and Brown 1998). En effet, ces progrès permettent d'augmenter la taille des échantillons et le nombre de taxa ainsi que leur analyse plus fine, mais restent indéniablement dépendants du rapport signal/bruit des données. Certains paramètres, tels des taux de variation différents entre sites, la non indépendance des nucléotides ou d'autres contraintes fonctionnelles peuvent être prises en considération en les incorporant dans des modèles plus complexes,

mais l'homoplasie, de par sa nature stochastique, demeure un casse-tête pour l'analyse phylogénétique. Des évaluations statistiques des arbres inférés (e.g. Swofford et al. 1996, Felsenstein 2004) permettent d'estimer la qualité de l'information contenue dans les données (correspondant justement au rapport signal/bruit) pour un arbre donné, mais pas de quantifier la justesse des relations inférées dans l'absolu.

Récemment, l'utilisation d'autres types de marqueurs, des changements génomiques rares, semble être un bon complément pour l'analyse phylogénétique (e.g. Rokas and Holland 2000). Citons, à titre d'exemple, l'insertion ou délétion d'introns (e.g. Rokas et al. 1999), des changements dans le code de l'ADN mitochondrial (e.g. Castresana et al. 1998) ou l'ordre de gènes mitochondriaux (e.g. Dowton 1999), la duplication de gènes (e.g. Spaethe et Briscoe 2004, Peel et al. 2006) et le polymorphisme d'insertion d'éléments génétiques mobiles (e.g. Teruel et al. 2010).

Ce chapitre présente les notions fondamentales permettant de comprendre les fondements et les difficultés de la reconstruction phylogénétique. Il est constitué de trois parties. La première partie introduit les définitions et les propriétés des arbres qui sont utilisés pour représenter des histoires évolutives. La seconde partie rappelle les propriétés et les définitions importantes concernant les données moléculaires. La troisième partie présente brièvement les modèles d'évolution les plus couramment utilisés pour modéliser l'évolution des séquences nucléotidiques.

II.2.2 Arbres – phylogénies

Depuis Darwin, les arbres sont utilisés comme support pour représenter l'aspect temporel de l'évolution et les regroupements d'espèces qui en découlent. Cette partie présente la terminologie relative à cette représentation et précise les principales propriétés des phylogénies évaluées.

L'hypothèse fondamentale de la reconstruction phylogénétique est que l'histoire évolutive se déroule par cladogènes successives. Suivant cette hypothèse, une lignée ancestrale peut, par spéciation, donner le jour à deux nouvelles lignées, et une phylogénie peut être représentée par un arbre similaire à ceux de la Figure 1. Ce paragraphe introduit plusieurs concepts qui sont fréquemment utilisés tout au long de cette thèse.

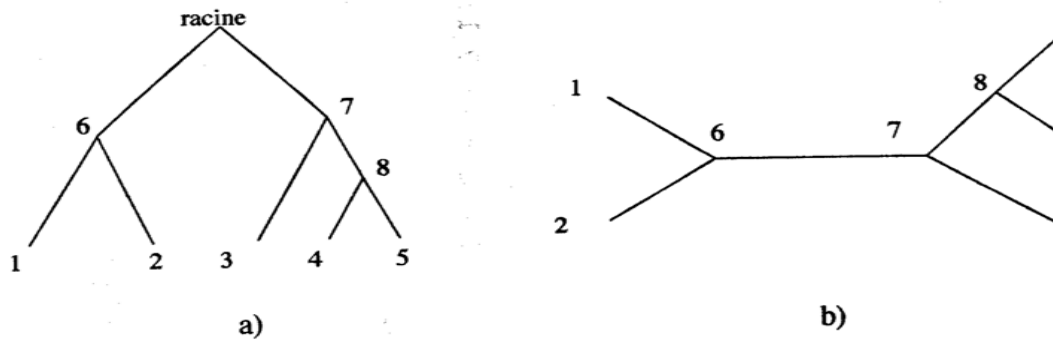


Figure 1. Représentations arborées d'une phylogénie: a) Arbre enraciné à 5 feuilles; b) version non enracinée de a).

Certaines des notions évoquées ci-dessous s'appuient sur des notions élémentaires de la théorie des graphes. Une phylogénie enracinée T pour un ensemble de n séquences $E = \{1, 2, \dots, n\}$ est un arbre dont les feuilles sont bijectivement associées aux séquences de E , et qui possède un seul sommet de degré deux qui en est la racine (e.g. Darlu et Tassy 1993). Les feuilles de la phylogénie représentant les séquences contemporaines étudiées sont parfois appelées nœuds externes. Les autres nœuds de l'arbre, qui correspondent à des hypothétiques séquences ancestrales des séquences étudiées, sont parfois appelés nœud internes. Un nœud interne peut aussi être vu comme un événement de spéciation. La racine de la phylogénie est le seul nœud interne de degré deux qui représente une spéciation. Les nœuds internes de degré supérieur à trois peuvent représenter le fait que plusieurs spéciations ont eu lieu et que l'on ne

sait pas dans quel ordre elles se sont produites. Ils peuvent également représenter le fait qu'une seule spéciation a engendré plus de deux nouvelles espèces. On considère généralement que ce dernier phénomène est peu probable, d'où la définition suivante : une phylogénie enracinée est dite complètement résolue si tous ses nœuds internes, autres que sa racine, sont de degrés trois. Une phylogénie qui ne possède pas de nœud de degré deux est dite non- enracinée.

Les méthodes de reconstruction phylogénétique reconstruisent généralement des phylogénies non-enracinées. Dans la suite de cette thèse, nous parlerons toujours, sauf mention contraire, de phylogénie enracinée avec des «groupes externes» (outgroups).

II.2.3 Données moléculaires

On sait aujourd'hui que l'ADN (l'Acide DésoxyriboNucléique), est le support de l'information génétique et de la transmission héréditaire. L'ADN est donc, à l'évidence, une source privilégiée d'information pour la reconstruction phylogénétique. De manière très simplifiée, on peut considérer que l'information génétique contenue dans l'ADN est transcrite en ARN (AcideRiboNucléique) puis en une chaîne d'acides aminés qui constitue une protéine. Il est possible d'utiliser ces séquences d'ADN, d'ARN ou d'acides aminés pour reconstruire des phylogénies.

Dans cette partie, nous introduisons la terminologie relative aux séquences moléculaires. Nous expliquons ensuite les raisons pour lesquelles ces séquences n'évoluent pas toutes à la même vitesse, et l'importance de ce phénomène en reconstruction phylogénétique. Puis, nous présentons les raisons pour lesquelles la phylogénie des séquences ne correspond pas de manière systématique à la phylogénie des espèces dont elles sont issues.

II.2.3.1 Vitesses d'évolution des séquences

Pour pouvoir reconstruire la phylogénie d'un ensemble de séquences moléculaires, il faut que ces séquences soient homologues, c'est-à-dire, issues d'une même séquence ancestrale. Lorsque l'on étudie un ensemble de séquences homologues, l'importance de la similitude entre ces séquences est un facteur important qui conditionne la qualité de la reconstruction phylogénétique. En effet, il est impossible de reconstruire la phylogénie de ces séquences, si elles sont toutes semblables ou si elles ne partagent aucun point commun. Or cette similitude dépend à la fois du temps depuis lequel les séquences ont divergé et de la vitesse d'évolution des séquences. On ne peut pas modifier l'époque à laquelle les événements évolutifs ont eu lieu, par contre on peut choisir des séquences qui évoluent plus ou moins rapidement suivant que l'on cherche à reconstruire une phylogénie portant sur des événements récents ou anciens.

La vitesse d'évolution correspond au pourcentage d'éléments de la séquence qui sont modifiés en une unité de temps. Pour qu'une modification apparaisse sur une séquence et soit transmise, il faut que cette transformation n'empêche pas le bon fonctionnement de l'organisme. Plus une séquence peut subir de modifications neutres plus cette séquence va pouvoir évoluer rapidement. Il est facile de voir qu'en fonction du type de données moléculaires que l'on considère, le nombre d'événements neutres possibles, et donc la vitesse d'évolution, varie. Par exemple, des modifications de l'ARN messenger n'affectent pas forcément l'ARN messenger mature, puisqu'elles peuvent être éliminées lors de l'épissage. De même, des modifications de l'ARN messenger mature n'affectent pas forcément la séquence d'acides aminés qu'il engendre, puisque plusieurs codons représentent le même nucléotide.

Pour des raisons similaires à celles évoquées ci-dessus, la vitesse d'évolution d'une séquence dépend également de "l'importance de ce qu'elle code", c'est-à-dire des contraintes fonctionnelles qui pèsent sur elle. Il est donc également important de choisir des séquences dont les contraintes fonctionnelles correspondent à l'histoire que l'on souhaite reconstruire.

Pour reconstruire la phylogénie de populations d'une même espèce, on peut, par exemple, étudier les séquences de **pseudo-gènes or NUMTs** (« NUclear MiTochondrial-like Sequences »), *i.e.* "des séquences d'ADN ressemblant à des gènes mitochondriaux et qui mutent rapidement" (e.g. Bensasson 2001). Pour étudier des histoires plus anciennes on peut, par exemple, considérer les séquences codant le cytochrome C, impliqué dans la respiration cellulaire; ces séquences sont donc fortement contraintes et évoluent lentement.

La vitesse d'évolution des séquences étant liée aux contraintes fonctionnelles qui portent sur elles, il est raisonnable de penser que des séquences homologues, qui subissent des contraintes fonctionnelles comparables, évoluent approximativement à la même vitesse et que cette vitesse est constante au cours du temps. Cette hypothèse, dite de l'horloge moléculaire, est basée sur les observations du taux d'évolution des protéines (Zuckerandl et Pauling 1962). En acceptant cette hypothèse, il est possible de dater la divergence entre deux espèces en divisant la quantité d'événements évolutifs qui séparent leurs séquences moléculaires par la vitesse d'évolution de ces séquences. En effet, sous cette hypothèse, la vitesse d'évolution de ces séquences, peut être estimée à partir de séquences homologues présentes chez des espèces pour lesquelles la date de divergence est déjà connue (par des données fossiles par exemple).

On sait aujourd'hui que cette "horloge moléculaire" est rarement fiable. Les mutations ne se produisent pas de manière indépendante, créant ainsi des épisodes d'accumulation de mutations suivis d'arrêts évolutifs qui font varier la vitesse d'évolution au cours du temps (Gillespie 1984). Cette hypothèse de l'horloge moléculaire reste cependant un outil efficace pour certaines études phylogénétiques. Sous l'hypothèse de l'horloge moléculaire, la racine d'une phylogénie évaluée doit être équidistante de toutes ses feuilles. Cela fournit un critère permettant d'enraciner une telle phylogénie.

Une autre manière d'enraciner une phylogénie consiste à ajouter une séquence dont on sait que sa divergence par rapport aux autres séquences est antérieure aux divergences propres au groupe étudié. La racine de la phylogénie initiale correspond alors à l'endroit où cette

dernière séquence se greffe. Cette méthode d'enracinement par groupe externe ("*outgroup*") est applicable, que l'hypothèse de l'horloge moléculaire soit ou non vérifiée. La phylogénie n'est généralement pas directement reconstruite en incluant la séquence du groupe externe, car la présence d'une séquence très différente des autres peut fausser la reconstruction.

Une discussion récurrente lors de l'analyse phylogénétique de séquences codantes concerne la manière de prendre en compte les troisièmes positions du codon. Dans les modèles ML, ces dernières ne satisfont généralement pas à l'hypothèse d'homogénéité de la composition en bases sur l'ensemble des séquences comparées et de comportement stationnaire. De plus, les substitutions en troisième position du codon saturent plus rapidement au cours du temps que celles aux deux premières positions. Cependant, ces troisièmes positions sont aussi celles qui apportent dans l'alignement de séquences le plus de sites variables, et il a déjà été noté sur les génomes mitochondriaux que les troisièmes positions contiennent un signal phylogénétique non négligeable (Yoder et al. 1996 ; Yoder & Yang 2000). Il semble que l'ajout des troisièmes positions du codon contribue à vieillir les estimations des âges de divergence. Ceci peut s'expliquer par le fait que la grande variabilité des troisièmes positions conduit à l'estimation de plus grandes longueurs de branches quand ces caractères sont inclus: les dates de divergence estimées sont alors plus profondes (Douzery 2002).

II.2.3.2 Phylogénies de séquences et phylogénies de taxons

A travers l'histoire des séquences moléculaires, on s'intéresse généralement à l'histoire des organismes d'où sont issues ces séquences, et à travers l'histoire de ces organismes c'est l'histoire des "groupes" auxquels ces organismes appartiennent que l'on cherche à reconstruire.

Pour pouvoir reconstruire la phylogénie d'un ensemble de séquences moléculaires, il faut que les séquences soient issues d'une même séquence ancestrale ; on dit qu'elles sont homologues. La phylogénie d'un ensemble de séquences moléculaires peut refléter l'histoire des gènes qui contiennent ces séquences ou celle des êtres vivants qui possèdent ces gènes. Ces deux histoires ne sont pas forcément identiques (par exemple dans le cas de duplications géniques) il faut donc s'assurer que les séquences utilisées permettent bien de reconstruire l'histoire à laquelle on s'intéresse.

Au cours de leur évolution les gènes peuvent être modifiés localement ou dupliques. Pour reconstruire la phylogénie d'un ensemble d'espèces on considère des gènes orthologues, c'est-à-dire des gènes homologues dont l'histoire ne contient pas de duplication.

Les phénomènes de duplication ne sont pas les seuls à compliquer l'analyse phylogénétique. Il arrive, par exemple, que des espèces échangent une partie de leur matériel génétique, on parle alors de transferts horizontaux. Lors de la reconstruction phylogénétique, on suppose que la similitude entre les séquences est due au fait qu'elles possèdent un ancêtre commun proche, or les transferts horizontaux augmentent la similitude entre des séquences qui n'ont pas forcément de parents communs proches. Les transferts horizontaux peuvent jouer un rôle considérable dans l'évolution des séquences et fausser complètement les reconstructions phylogénétiques issues d'analyses de séquences moléculaires.

Comme nous venons de le voir, il existe plusieurs phénomènes biologiques qui peuvent perturber la reconstruction phylogénétique, et à cause desquels la phylogénie reconstruite pour les séquences ne reflète pas forcément la phylogénie des organismes contenant ce patrimoine génétique. De plus on ne s'intéresse généralement pas à l'histoire de ces organismes particuliers mais plutôt à l'histoire des taxons auxquels ils appartiennent. Un taxon est un ensemble d'organismes dont l'unité est admise et qui a des relations de parenté avec d'autres ensembles comparables. On parle de taxon pour désigner indifféremment les criquets, les chiens ou une souche particulière d'un virus. Chaque organisme ayant un

patrimoine génétique qui lui est propre, les phylogénies reconstruites dépendent en partie des représentants que l'on utilise pour chaque taxon (Philippe et Douzery, 1994).

Pour reconstruire une phylogénie à partir de séquences moléculaires, il est nécessaire de choisir avec soin ces séquences. C'est une étape fondamentale qui ne peut être effectuée correctement qu'au cas par cas et avec de solides connaissances en biologie moléculaire et en taxonomie.

II.2.3.3 Alignement des séquences

A partir des années 70, les techniques de séquençage (permettant d'obtenir ces données moléculaires) ont très rapidement évolué. Les études ont d'abord concerné des fragments d'ADN, puis des gènes entiers et enfin, l'intégralité d'un chromosome. Le séquençage de l'ADN a notamment beaucoup progressé grâce à la technique "d'amplification en chaîne par polymérase", PCR («*Polymerase Chain Reaction*») (Mullis et Faloona 1987). Il est maintenant devenu plus facile d'obtenir des séquences d'ADN correspondant aux taxons que l'on souhaite étudier, soit par séquençage, soit en utilisant les banques de données qui mettent gratuitement à disposition les résultats de très nombreux séquençages déjà effectués.

De même qu'il faut s'assurer que les séquences étudiées sont orthologues, il faut aussi s'assurer que les nucléotides ayant la même position dans les différentes séquences sont issus d'un même nucléotide ancestral. Cette étape d'alignement des séquences permet par la suite de comparer des nucléotides qui sont effectivement comparables. Aligner les séquences est une étape difficile. En effet, au cours de l'évolution, des nucléotides ont pu être modifiés ou même supprimés tandis que d'autres ont pu être insérés au sein de la séquence. Plusieurs explications sont possibles pour un alignement, mais l'explication la plus parcimonieuse est celle que l'on retiendra. De la même manière, parmi tous les alignements existants, on tend toujours à conserver celui qui s'explique par un nombre minimum d'événements évolutifs.

D'un point de vue informatique, ce problème s'apparente à la recherche de la "distance d'édition" entre les séquences (Levenshtein, 1966). Il s'agit de trouver l'alignement qui implique un nombre minimal d'opérations d'édition nécessaires pour transformer une séquence en un autre, donc qui minimise le nombre d'événements mutationnels (nombre de substitutions + nombre de délétions + nombre d'insertions). Généralement, en biologie, la probabilité d'une substitution est supérieure à celle d'une délétion suivie d'une insertion (e.g. Felsenstein 2004). Cela est pris en compte lors de l'alignement en attribuant des coûts distincts aux différents événements mutationnels. Mais, en phylogénétique on ne cherche pas exactement l'alignement de coût le plus faible. L'objectif est avant tout de mettre en correspondance des positions pour lesquelles on est sûrs qu'il y a homologie.

Dans les régions stables/conservées des séquences, *i.e.* les régions dans lesquelles on peut proposer un alignement sans insertion ni délétion, il est très probable que les nucléotides, alignés sur un même site, sont effectivement homologues. Dans ce cas, les biologistes utilisent généralement un algorithme heuristique afin d'obtenir un premier alignement raisonnable qui privilégie les grandes régions stables. Puis, ils affinent cet alignement à la main en se servant de logiciels interactifs (e.g. MESQUITE 2.72, Maddison et Maddison 2006). Finalement, les positions pour lesquelles il n'y a eu ni insertion ni délétion sont utilisées comme repères pour délimiter les régions très variables, riches en insertions - délétions ("indels") pour lesquelles l'homologie est impossible à déterminer avec certitude. Ces régions hypervariables sont soit éliminées de l'analyse phylogénétique, soit sont codées, comme caractères avec multiples états (voir le programme INAASE, Lutzoni et al. 2000), en vue de leur utilisation dans l'analyse de parcimonie.

II.2.4 Modèles de l'évolution moléculaire

La modélisation de l'évolution moléculaire consiste à exprimer en terme probabiliste la transformation d'une séquence (e.g. Page et Holmes 2001). Ainsi, le calcul de la fonction de

vraisemblance dans sa forme la plus simple nécessite un modèle décrivant les probabilités de substitution d'un nucléotide par un autre. Cependant, pour être statistiquement robustes et performantes, les méthodes probabilistes nécessitent l'incorporation de modèles qui décrivent de la façon la plus réaliste possible les processus biologiques d'évolution des séquences. Ces modèles incorporent ainsi généralement des paramètres prenant en compte les fréquences nucléotidiques, les taux de substitutions, la présence d'une fraction de sites invariables et l'hétérogénéité des taux de substitution entre sites (Figure 2) (e.g. Posada et Crandall 1998, Delsuc et Douzery 2004a).

Les principaux modèles utilisés à présent dans les analyses probabilistes (maximum likelihood et bayésienne) sont présentés brièvement dans la figure 2. Aucun des modèles couramment employés ne traite les insertions et les délétions, c'est une des hypothèses faites par les modèles d'évolution. La plupart de ces modèles sont "réversibles dans le temps", fait qui rend le résultat de l'analyse phylogénétique indépendant de l'enracinement de l'arbre. Quelques " *a priori*" communs peuvent être nommés pour les modèles présentés dans la Figure 2:

- les séquences évoluent exclusivement par le mécanisme de substitution nucléotidique;
- les sites sont indépendants : les modifications qui affectent un site au cours du temps, ne sont affectées ni par l'état du reste de la séquence ni par les modifications qui affectent les autres sites. Sous cette hypothèse, il suffit de connaître le processus d'évolution des sites pour connaître celui des séquences;
- les sites sont identiquement distribués : le processus d'évolution est le même pour tous les sites et ne dépend donc pas de la position du site dans la séquence. Le modèle d'évolution et ses paramètres sont donc les mêmes pour tous les sites;
- le processus d'évolution est markovien ou "sans mémoire": l'évolution d'une séquence ne dépend que de l'état actuel de cette séquence et non du processus évolutif qui a conduit à cette séquence;

- le processus d'évolution est homogène: il est le même pour toutes les branches de la phylogénie, le processus reste le même au cours du temps;
- le processus d'évolution est stationnaire : la probabilité d'observer une base particulière ne dépend pas du moment de l'observation, et cette probabilité est la même pour toutes les séquences considérées;
- il y a au plus une mutation par unité de temps (e.g. Swofford et al. 1996, Nei et Kumar 2000, Felsenstein 2004).

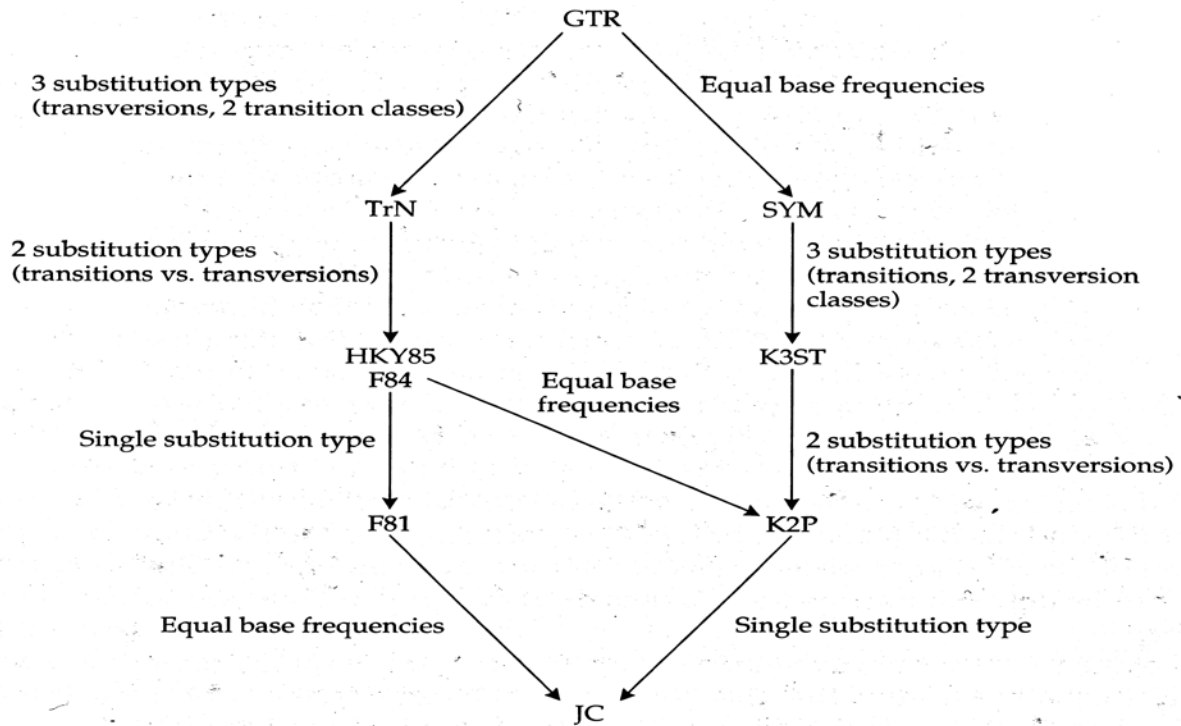


Figure 2. Principaux modèles d'évolution (d'après Swofford et al. 1996). Chaque flèche indique les restrictions nécessaires pour passer d'un modèle à un cas particulier de ce modèle. Le nombre de paramètres est indiqué entre parenthèses. Les abréviations utilisées pour représenter les différents modèles sont les suivants: **GTR** - "general time-reversible model" (Lanave et al. 1984) ; **TrN** - modèle de Tamura et Nei (1993) ; **SYM** - modèle de Zarkikh et Li (1995) ; **F84** - modèle de Felsenstein (1993) dans "PHYLIP" ; **HKY85** - modèle de Hasegawa Kishino et Yano (1985) ; **K3ST** - modèle de Kimura (1981) 3 types de substitutions ; **K2P** - modèle de Kimura (1980) 2 types de substitutions ; **F81** - modèle de Felsenstein (1981), Tajima-Nei (1982) ; **JC** - modèle de Jukes et Cantor (1969).

II.2.5 Méthodes de reconstruction phylogénétique utilisées

La phylogénie consiste à construire des arbres représentant des relations de parenté putatives entre êtres vivants (actuels ou fossiles) à partir d'observations sur un échantillonnage des taxons, par exemple des individus représentant leur espèce ou leur genre.

Morphologie, anatomie, protéines, ADN ou comportements sont des sources de caractères pouvant servir à comparer les taxons. L'utilisation des séquences d'ADN, de par la relative facilité (e.g. PCR, séquençage automatisé) avec laquelle on peut obtenir un grand nombre de caractères homologues, est très répandue en phylogénie, quand les taxons étudiés sont actuels ou quand on dispose de spécimens anciens dans un état de conservation suffisamment bon pour permettre l'extraction d'ADN. On parle alors de «phylogénie moléculaire».

La « phylogénie moléculaire», en synergie avec l'accessibilité croissante des moyens informatiques, a suscité le développement de nouvelles méthodes d'analyses ; des modèles d'évolution de séquences ont permis l'amélioration des méthodes de distance et ont été à l'origine du développement spectaculaire de l'application à la phylogénie des méthodes probabilistes.

Les trois grandes familles de méthodes de construction des phylogénies utilisées actuellement sont donc (i) la méthode cladistique (maximum de parcimonie) (e.g. Edwards et Cavalli-Sforza 1963), (ii) les méthodes phénétiques (méthodes de distance) (Fitch et Margoliash 1967) et (iii) les méthodes probabilistes : de maximum de vraisemblance (e.g. Felsenstein 1973) et bayésiennes (e.g. Rannala and Yang 1996, Huelsenbeck et Ronquist 2001, Delsuc et Douzery 2004b).

Toutes les méthodes évoquées ci-dessus reposent sur trois points :

(i). Un critère d'optimalité, qui mesure l'adéquation des données à un arbre phylogénétique donné (par exemple : la parcimonie, la vraisemblance, etc) ;

(ii). Une stratégie de recherche pour identifier l'arbre optimal (par exemple, recherche heuristique, exacte etc.)

(iii). Des hypothèses sur le mécanisme d'évolution des données.

A l'antipode de ces points communs se trouvent les critères et techniques d'optimalité, spécifiques pour chaque méthode. Ainsi, le but de la technique de parcimonie est de rechercher, par le biais d'algorithmes exacts ou heuristiques, l'arbre phylogénétique impliquant le moins de mutations, le plus parcimonieux. Certains logiciels font explicitement référence à l'usage de la parcimonie comme critère d'optimalité, comme PAUP (Swofford 1995, abréviation de « Phylogenetic Analysis Using Parsimony »).

Bien que fondées toutes les deux sur la fonction de vraisemblance utilisant les mêmes modèles d'évolution des séquences (voir Figure 2), les méthodes de maximum de vraisemblance et bayésienne diffèrent par le concept de probabilité qu'elles emploient. Ainsi, la méthode du maximum de vraisemblance sélectionne l'arbre qui maximise la vraisemblance (celui ayant la plus forte probabilité d'avoir conduit aux données), alors que l'approche bayésienne assigne une distribution de probabilités *a priori* aux différents arbres et réalise des inférences à partir de la distribution des probabilités *a posteriori* de ces arbres (Delsuc et Douzery 2004a).

Les performances des différentes méthodes dépendent, entre autres, des différences structurales (e.g. composition en bases nucléotidiques ou acides aminés) et de la différence de vitesse d'évolution entre les sites nucléotidique d'une même séquence et entre des séquences différentes pour un même site. Il n'existe pas de méthode supérieure à toutes les autres, chacune à ses forces et ses faiblesses et le débat sur les mérites comparés de deux méthodes n'est pas clos (e.g. DeBry et Abele, 1995 ; Douady et al., 2003 ; Holder et al., 2010).

Dans ce contexte, nous avons choisi d'illustrer l'analyse de l'évolution des Melanoplinae par le biais des méthodes de maximum parcimonie, de maximum de vraisemblance et bayésienne. Nous avons employé des techniques différentes dans le souci d'adopter une

stratégie comparative des résultats, que nous avons ainsi interprété sur des principes d'optimalité distincts, fait qui nous a permis d'observer la fiabilité de nos résultats phylogénétiques.

II.2.6 Stratégie d'échantillonnage utilisée

Le schéma de la stratégie d'échantillonnage utilisée dans le présent travail est présenté dans la figure 3.

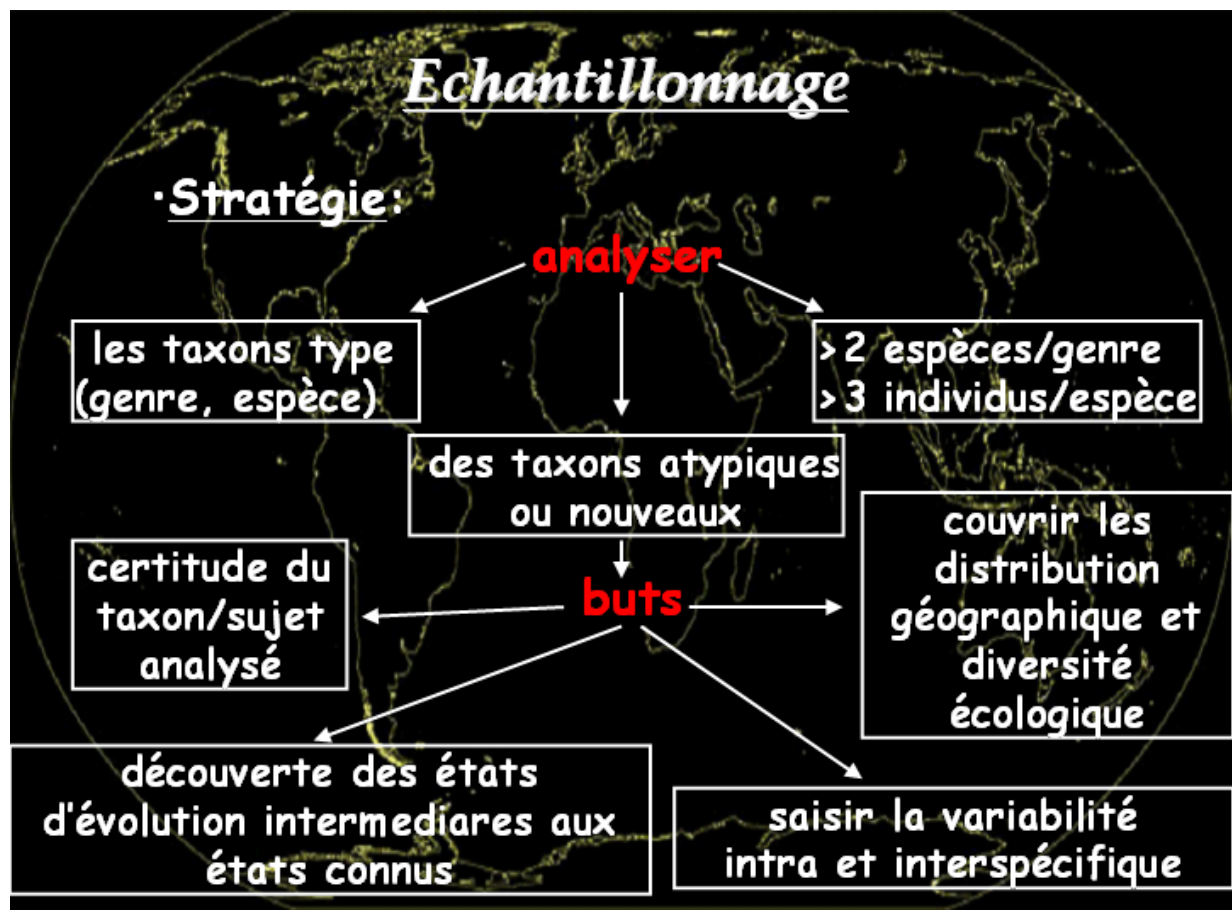


Figure 3. Stratégie d'échantillonnage utilisé dans l'étude des Melanoplinae.

III. Les Melanoplinae (Amédégnato, 1974) - présentation du modèle d'étude



Parmi les insectes, les acridiens (Insecta: Orthoptera: Caelifera: Acrididae) constituent un modèle intéressant pour les investigations phylogénétiques et biogéographiques grâce à leurs ancienneté, distribution géographique, caractéristiques écologiques, et à leur impact économique.

III.1. Morphologie - caractères utilisés dans la détermination des espèces

Membres de la famille Acrididae (Insecta, Orthoptera, Caelifera), les Melanoplineae ou « miramelles » sont des criquets de taille moyenne à petite, de formes diverses, souvent trapues, couramment brachyptères (chez les Dichroplini), microptères (Jivarini) ou aptères (chez les Podismini); les formes ailées (chez les Melanoplini et Dichroplini) sont peu nombreuses. Les téguments sont lisses, la coloration généralement vert ou brune, la face ventrale et la plaque sous génitale du mâle, ainsi que les tibias postérieurs fréquemment vivement colorées (Figure 4).

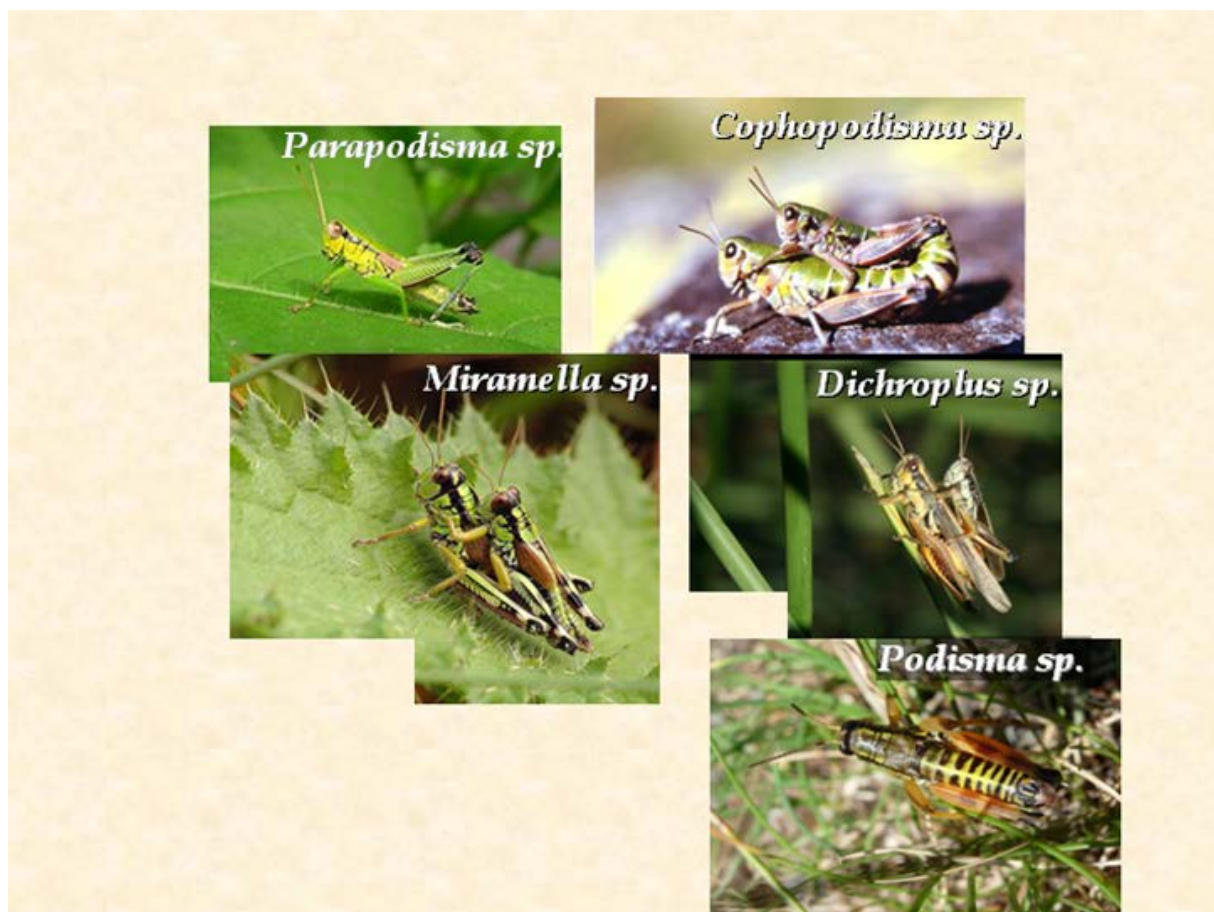
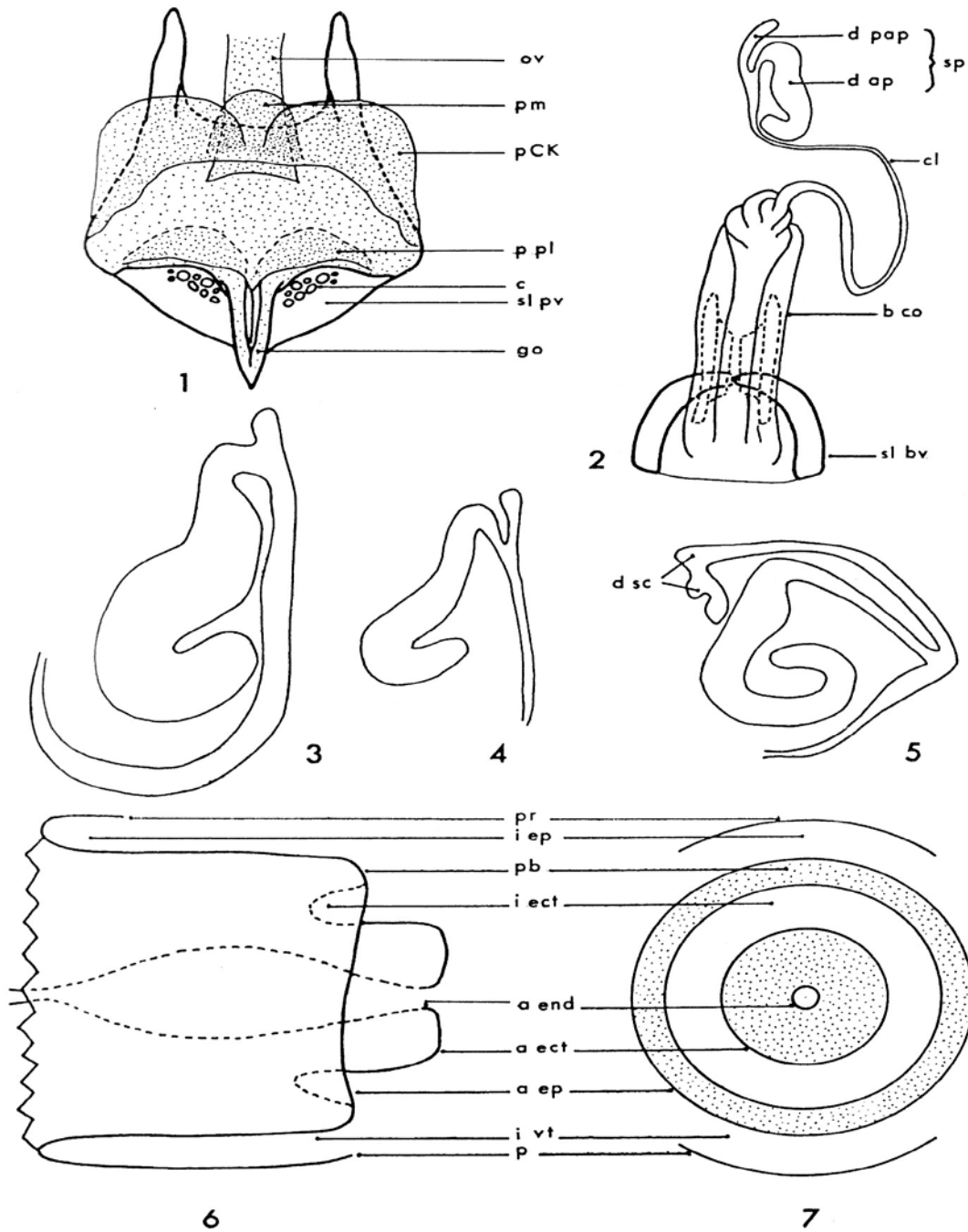


Figure 4. Genres des criquets de la sous-famille Melanoplineae (France, 2010; Argentine, 2003).

La tête est orthognate ou légèrement opistognathe, le front bombé; les yeux sont petits, ovales ou arrondis; les antennes sont composées de 18 à 24 articles. Le prosternum a une saillie à tubercule le plus souvent spiniforme. Le pronotum est généralement sans carènes latérales (sauf les Jivarini) et avec une carène médiane peu visible ou absente. Le tympan abdominal est présent. Les pattes postérieures sont à femurs bien développés, les tibias sont dépourvus d'épine apicale externe (exception: les Jivari ou Jivarini), et avec 8 à 14 épines du côté externe. L'extrémité abdominale mâle est large, volumineuse, la plaque sous-génitale et les derniers segments plus ou moins recourbés vers le haut ; le dernier tergite abdominal à petite furcula souvent présente; la plaque sous-génitale large, cupuliforme arrondie ou légèrement allongée à l'apex, dépassant plus ou moins la plaque sous-anale (Fig. 5), laissant

apparaître un pallium épaissi et ridé; plaque sous-anale subtriangulaire, sans callosités, à dépression longitudinale plus ou moins marquée, parfois légèrement divisée transversalement; les cerques mâles ont des formes variées, courts coniques à longs et sinueux ou larges et aplatis, parfois bifurqués; les valves de l'oviscapte sont courtes et robustes généralement.



LISTE DES ABRÉVIATIONS

Genitalia ♀			
b.co.	bourse copulatrice	g.	gaine
c.	columelles	i.ect.	invagination ectophallique
cl.	canal	i.ep.	invagination épiphallique
d.ap.	diverticule apical	i.vt.	invagination ventrale
d.pap.	diverticule préapical	m.c.cg. ..	membrane centrale du cingulum
g.o.	guide-œuf	m.ect.	membrane ectophallique
ov.	oviducte	m.ep.	membrane épiphallique
p.C.K. ..	poches de Comstock Kellog	p.	pallium
p.m.	poche médiane	p.ant.	partie antérieure du sclérite endophallique
p.pl.	poches du plancher	p.b.	pli basal = pli dorsal
sl.bv.	sclérites basivalvaires	p. moy. .	partie moyenne du sclérite endophallique
sl.pv.	sclérites postvaginaux	p.post. ..	partie postérieure du sclérite endophallique
sp.	spermathèque	pr.	paraproctes
		pr.g.	processus du gonopore
		p.sz.	plaque supra zygomale
		pt.	pont
		r.cg.	rami du cingulum
		s.ej.	sac éjaculateur
		s.sp.	sac spermatophore
		s.r.	supra rami
		sl.i.e.	sclérites inférieurs de l'édéage
		sl.end.	sclérites endophalliques
		sl.lv.	sclérites latéroventraux
		sl.s.e.	sclérites supérieurs de l'édéage
		v.i.e.	valves inférieures de l'édéage
		v.s.e.	valves supérieures de l'édéage
		z.	zygoma
Genitalia ♂			
a.cg.	arche du cingulum		
ap.cg.	apodèmes du cingulum		
ap.end. ..	apodèmes endophalliques		
a.ect.	assise ectophallique		
a.end.	assise endophallique		
a.ep.	assise épiphallique		
d.sc.	diverticules secondaires		
ect.	ectophalle		
end.	endophalle		
ep.	épiphalle		
fl.	flexure		

Figure 5. Genitalia femelle et mâle chez les acridiens (d'après Amédégnato, 1977): 1 – plaque sous-génitale (♀) dorsalement ; 2 – réceptacle séminal et sclérites basivalvaires dorsalement ; 3-5 spermathèque (♂); complexe phallique, structure générale.

Au point de vue genitalia, les mâles de Melanoplinae sont caractérisés par la possession de trois assises épi-, ecto- et endo-phallique (Fig. 5). Les sclérites endophallique sont aplatis latéralement, à apodèmes généralement courts et avec la partie moyenne largement étalée sur le sac spermatophore; processus du gonopore digitiforme, dirigé vers l'arrière; partie moyenne séparée de la partie postérieure (sclérite inférieur de l'édéage) par une fracture nette (parfois moins nette chez certains Podismini); sclérites supérieurs et inférieurs de l'édéage de configuration variable selon les groupes (Amédégnato 1977).

Les caractéristiques du genitalia femelle sont un réceptacle séminal, renflé en une bourse copulatrice à la base et terminé par une spermathèque munie d'un diverticule préapical très long, simple ou sinueux, irrégulier et avec des diverticules secondaires (Fig. 5). La plaque

sous-génitale à des poches de Comstock-Kellog; les sclérites post-vaginaux sont à columelles en nombre et disposition variables selon les groupes (Amédégnato 1977). Ce sont donc les caractères des appareils génitaux, mâles et femelles, qui constituent les éléments essentiels dans la classification et la phylogénie morphologique des espèces des acridiens, en général.

III.2 Habitat – répartition

Les Melanoplinae sont des insectes de milieux ouverts, que ce soit de biotopes purement herbacés ou plus ou moins buissonneux, souvent situés dans les étages sous alpin et/ou alpin des régions tempérées et tropicales. Leur distribution géographique actuelle (Fig. 6) est issue d'une interaction entre l'âge du groupe, sa capacité de dispersion et ses exigences écologiques. Avec plus de 100 genres et 900 espèces (Otte 1995, Litzemberger and Chapco 2003), les Melanoplinae constituent un élément essentiel de la faune acridienne de l'Eurasie et des Amériques; ils représentent l'élément dominant de l'entomofaune d'altitude dans les Andes.

Outre sa valeur faunistique, ce groupe possède une importance économique par ses espèces nuisibles pour les productions agricoles ou pastorales, tels que *Melanoplus bivittatus*, *M. sanguinipes*, *Dichroplus elongatus* ou *D. maculipennis* dans les Amériques (e.g. Knowles and Otte 2000, Torrussio et al. 2002).

• Habitat et répartition des Melanoplinae

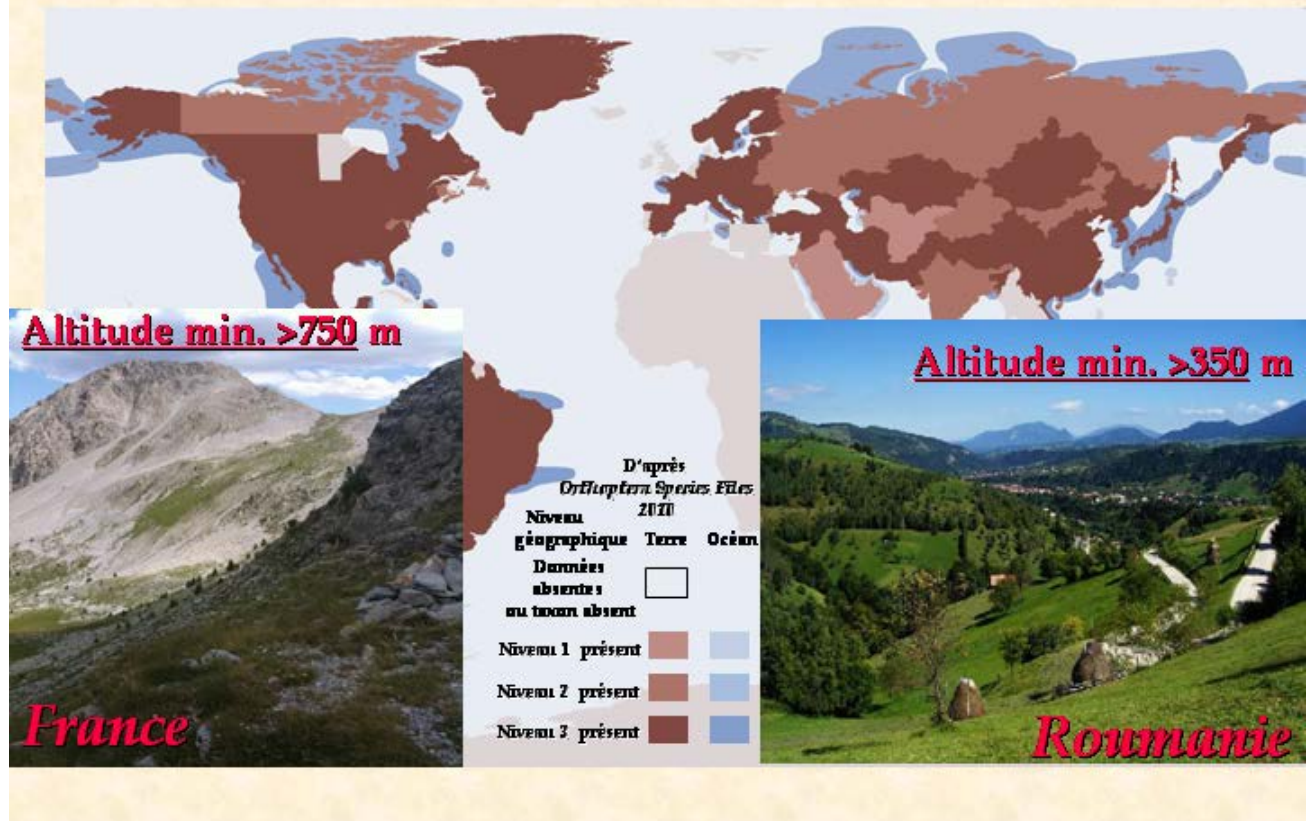


Figure 6. Répartition géographique actuelle des Melanoplinae.

III.3 Contexte paléontologique

Dans un souci de repérage temporel dans l'évolution des acridiens, nous présentons quelques notions de paléontologie des Caelifères, compte tenu de l'absence de fossiles pour le groupe de Melanoplinae *sensu stricto*.

L'ordre ORTHOPTERA (Arthropoda, Hexapoda, Insecta, Pterygota, Neoptera) est un des groupes d'insectes les plus anciens (Zeuner 1939, Sharov 1968, 1971, Kukalova-Peck 1991, Carpenter 1992, Gorokhov 1995). L'opinion de Ander (1939) qui découpe cet ordre en

deux sous-ordres, ENSIFERA et CAELIFERA, est la plus acceptée parmi les scientifiques à présent. L'ordre regroupe plus de 20 000 espèces d'Ensifères (sauterelles et grillons) et de Caelifères (criquets) (Kevan 1982, Flook et Rowell 1997). Avec les Phasmes, les Orthoptères seraient issues du groupe des Protorthoptères qui existait au Carbonifère et s'est éteint au Permien. De ce groupe auraient entre autre dérivé les Locustopseidae (Locustopsoidea - Gorokhov 1995) qui auraient permis la radiation depuis le Jurassique des groupes de Caelifères actuels (Fig. 7) (Storozhenko 1997).

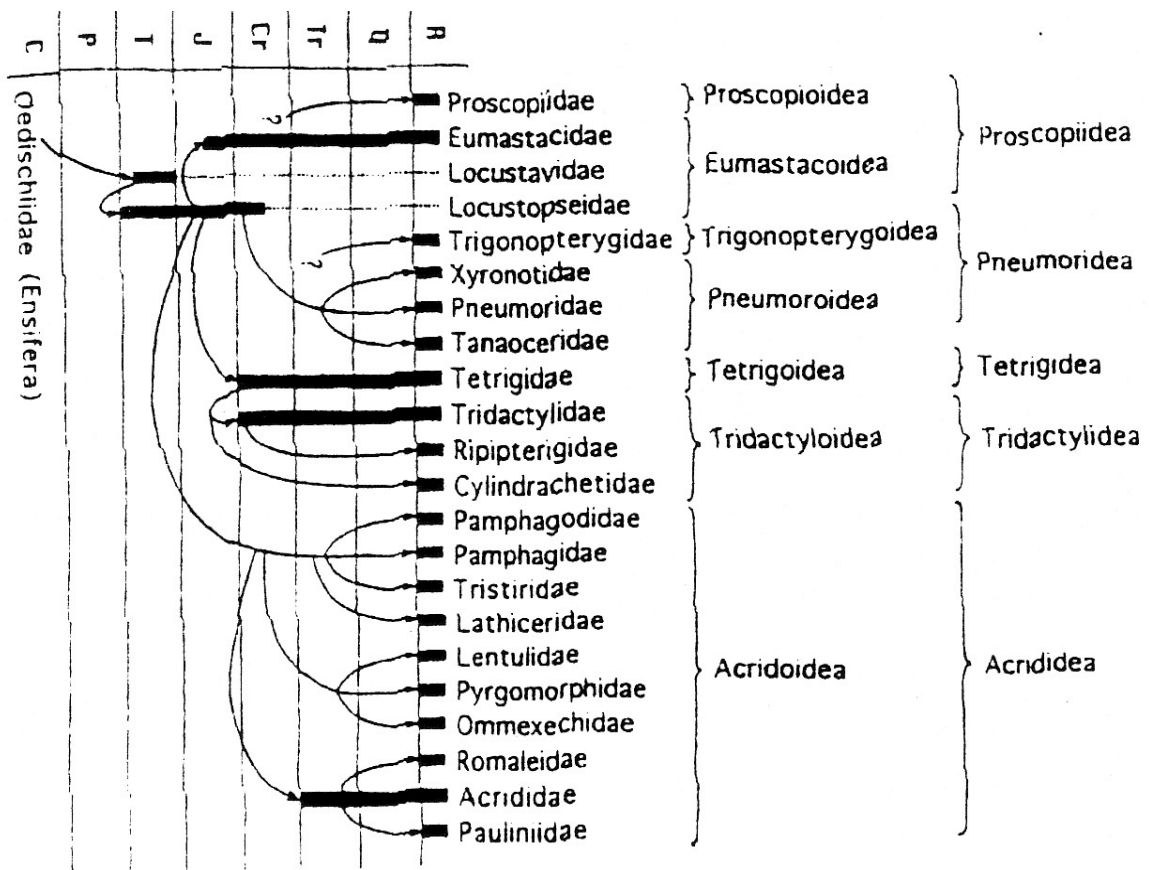


Figure 7. Arbre phylogénétique pour le sous-ordre Caelifera (d'après Storozhenko, 1997) ; C – Carbonifère, P – Permian, T – Triasique, J – Jurassique, Cr – Crétacique, Tr – Tertiaire, Q – Quaternaire, R – Récent.

La plasticité morphologique des Caelifera en général, et des ACRIDOMORPHA (Dirsh 1966) en particulier, étant très grande, ceci engendre une diversité phénotypique considérable, des formes semblables étant généralement en relation avec le type de formation

végétale préférentiel. De plus, nombreux sont les groupes qui fréquentent différents habitats, ainsi que les habitats occupés par des espèces appartenant aux différentes sous-familles, familles ou superfamille d'où la fréquence des convergences morphologiques.

Cette instabilité morphologique, jointe au manque de caractères externes constants, est responsable des contradictions qui rendent obscures les classifications au sein de criquets (Insecta : Orthoptera), en général (Amédégnato 1977, 1993 ; Otte 1995 ; Vickery 1997).

Dans ce contexte, pour l'identification des espèces (quand nécessaire), nous sommes essentiellement attachée à l'anatomie des genitalia, seul moyen d'approche, devant la carence de la morphologie externe (Amédégnato, 1977). Au sein de ces taxa, et généralement des ACRIDOMORPHA, la diversité de l'anatomie des genitalia est énorme allant de formes archaïques à d'autres, très récentes (Dirsh 1956, 1975 ; Amédégnato 1977). Mais, cette diversité des forme représente des variations d'un même schéma général : (1) très uniforme à travers les groupes pour le genitalia femelle, et (2) avec trois assises élaborées progressivement au cours du développement larvaire, pour le genitalia mâle (Fig. 5) (Amédégnato 1977).

En se basant sur la structure du complexe phallique, Roberts (1941) a proposé l'organisation des Acridomorpha en deux divisions : (I) les Chasmosacci (sac éjaculatoire ouvert) et (II) les Cryptosacci (sac éjaculatoire fermé) (Fig. 8). Les genitalia femelles n'apportent aucun élément à ce sujet (Amédégnato, 1977).

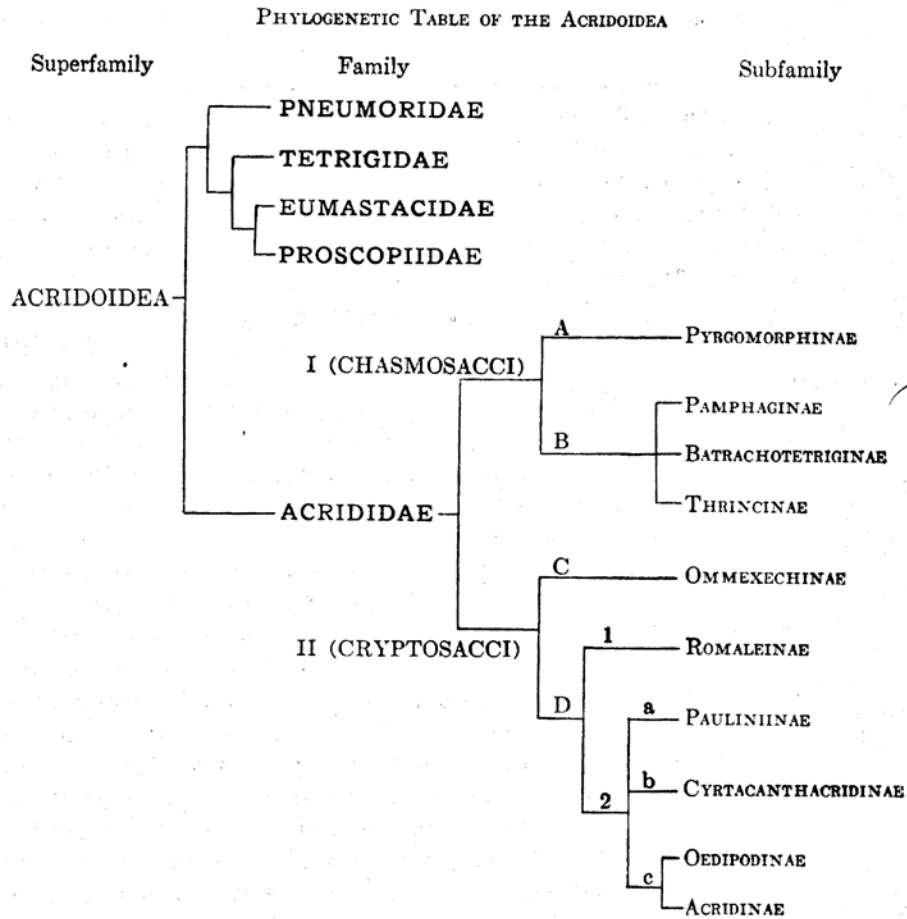


Figure 8. Classification morphologique des Acridoidea (d'après Roberts, 1941).

Ultérieurement, de son étude comparative, des genitalia mâle et femelle des Acridoidea du Nouveau Monde, Amédégnato (1976) soulignait que :

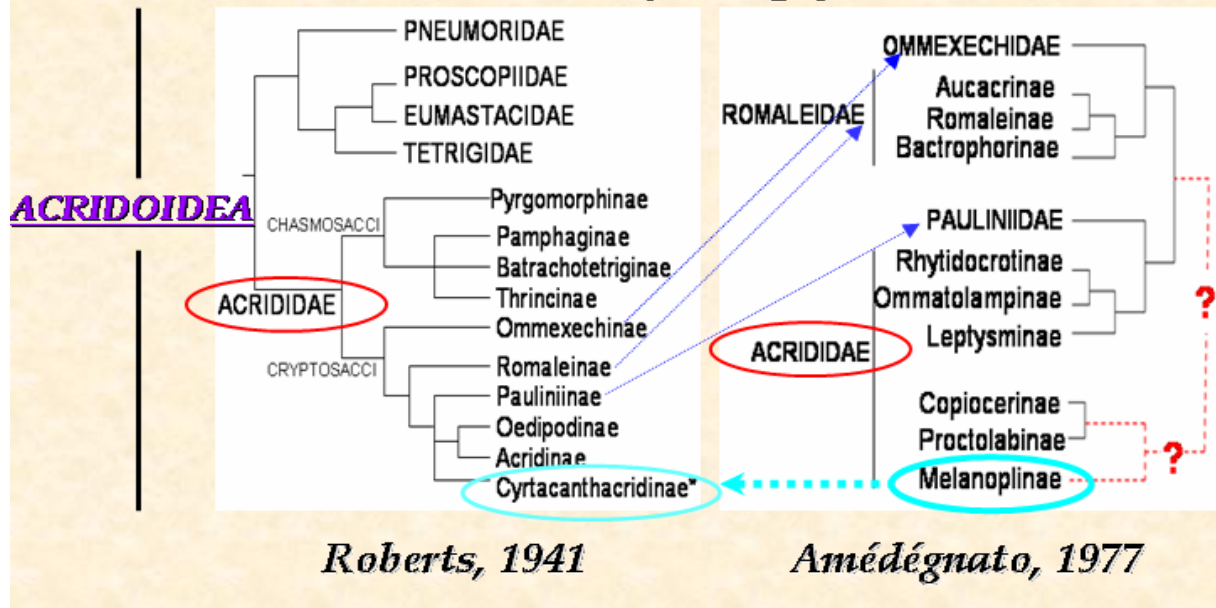
(1) les Romaleidae, Ommexechidae, Acrididae et Pauliniidae constituent un ensemble homogène s'opposant aux autres familles actuellement encore incluses dans les Acridoidea ;

(2) Acridoidea représente le stade évolutif le plus avancé à l'intérieur du sous-ordre des Acridomorpha (Dirsh, 1966), lui-même parfaitement cohérent avec un complexe phallique toujours tri stratifié et une direction évolutive constante (élaboration progressive des trois assises en commençant par la plus externe) depuis la superfamille la plus primitive, Eumastacoidea, jusqu'à la plus évoluée.

L'utilisation extensive de certains caractères morphologiques, comme l'anatomie des genitalia mâle et femelle, a introduit dans les classifications actuelles une pléthore de rangs taxonomiques (sous-familles, tribus, sous-tribus, « groupes »), ainsi que de taxons qui ne sont pas basés sur de vraies synapomorphies. Cet « excès » taxinomique ne fait que masquer les relations d'apparentement entre taxa. L'élévation des taxa aux rangs supérieurs (e.g. Dirsh 1975), a eu le même effet de « brouillard » (e.g. Jago 1979, Vickery 1997). Un état de l'art des connaissances actuelles sur la systématique des acridiens, aux niveaux taxonomiques « Acridoidea » et « Acrididae », est proposé dans la figure 9.

Systematique et phylogenie: niveau Acridoidea

• cladogrammes caracteres morphologiques :



Systematique et phylogenie: niveau Acrididae

• cladogrammes caracteres:

• morphologiques

• moléculaires

ADN mitochondrial 12S+16S

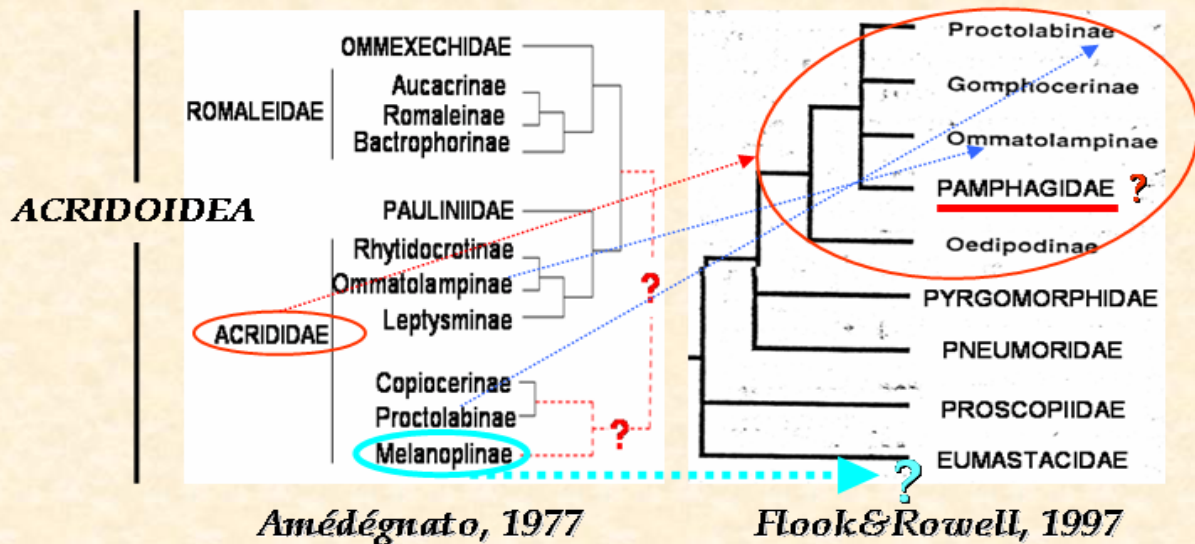
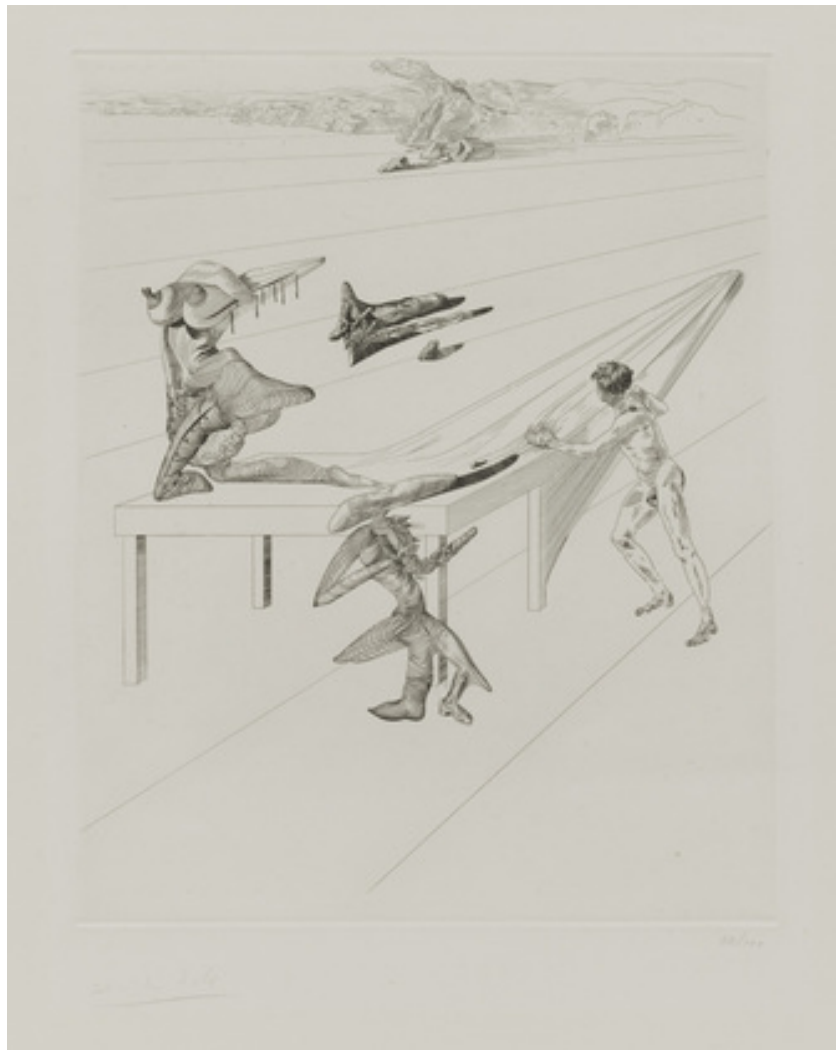


Figure 9. Systematique et phylogenie des acridiens aux niveaux taxonomiques de la « superfamille » (Acridoidea) et à celui de la « famille » (Acrididae).

IV. Objectifs de la thèse



Avec cet état de l'art des connaissances sur la systématique et l'évolution des Melanoplinae comme point de départ, nous nous sommes proposés, dans le cadre de notre thèse, d'atteindre les suivants objectifs:

(i) tester les hypothèses préexistantes d'évolution spatio-temporelle aux deux niveaux taxonomiques:

(1) de la sous-famille Melanoplinae,

(2) de la tribu Podismini,

(ii) proposer des nouveaux scénarios pour l'évolution des lignées ci-dessus mentionnées, ci c'est le cas;

(iii) proposer des affiliations évolutionnistes axées sur l'interprétation synthétique des données moléculaires, biogéographiques et de géologie tectonique globale, pour des taxons nouveaux (*status incerta*) dans les classifications morphologiques courantes);

(iv) à identifier et dater les pseudogènes d'ADN ribosomal chez *P. pedestris* et à proposer des scénarios concernant le rôle des pseudogènes dans l'évolution du génome.

L'ensemble de ces objectifs a été atteint en générant des données moléculaires pour des spécimens préservés dans la collection des Caelifères du MNHN. Malgré les difficultés et lourdeurs techniques que cela impose, l'intérêt est de pouvoir étudier des taxons difficiles à échantillonner, parfois disparus de leur aire naturelle ou, par contre nouvellement décrits.

Chapitre V - Utilisation des marqueurs moléculaires, nucléaires et mitochondriaux, dans l'étude de l'histoire évolutive de criquets (Orthoptera: Acrididae) à deux niveaux taxonomiques: la sous-famille (Melanoplinae) et la tribu (Podismini).



**Chapitre V - Utilisation des marqueurs moléculaires, nucléaires et mitochondriaux,
dans l'étude de l'histoire évolutive de criquets (Orthoptera: Acrididae) à deux niveaux
taxonomiques: la sous-famille (Melanoplinae) et la tribu (Podismini).**

V.1. Problématique et démarche scientifique

L'évolution et la systématique morphologique de la **sous-famille** Melanoplinae et de ses sous-unités (e.g., la **tribu** Podismini) demeurent un sujet controversé et les contours des taxons, ainsi que leurs relations d'apparement, demeurent flous. La plasticité morphologique du sous-ordre *Acridomorpha* (Dirsh, 1966) se retrouve au niveau de la sous-famille Melanoplinae qui présente une diversité phénotypique considérable, et une étonnante convergence des formes, en relation avec le type préférentiel de formation végétale. En conséquence, différents auteurs (e.g. Rehn and Randell 1963, Amédégnato 1974, Otte 1995, Vickery 1987, 1997) ont divisé les Melanoplinae en deux, trois ou plusieurs tribus, en utilisant comme critère essentiel la structure du complexe génital mâle (voir paragraphe II.1, figure 4). Le point commun de ses classifications morphologiques est le regroupement des espèces Paléarctiques dans une seule unité, les Podismini (Rehn and Randell 1963, Amédégnato 1974, Vickery 1997) ou les Podisminae (Dirsh 1975, Otte 1995). A son tour, cette unité a été rattachée à différentes entités taxonomiques telles que la super-tribu (Melanoplini, Rehn and Randell 1963), la sous-famille (Cyrtacanthacridinae, Rehn and Randell 1963; Melanoplinae, Amédégnato 1974, Catantopinae, Harz 1975) ou la famille (Acrididae, Otte 1995).

Lors de la dernière décennie, les classifications ont tenté de prendre en compte les relations évolutives entre taxons, suite aux recherches moléculaires. C'est le cas pour les Melanoplinae. Le séquençage de l'ADN et son analyse par le biais de méthodes phylogénétiques (parcimonie, maximum de vraisemblance et méthode bayésienne) a éclairci les relations phylogénétiques au sein du groupe et a permis d'affiner sa systématique

(Amédégnato et al. 2003, Chapco et al. 2001, Litzemberger et Chapco 2001). Mais, malgré les avancés mentionnées, l'histoire évolutive des Melanoplineae demeure en grande partie énigmatique. Trois scénarios ont été proposés à cet égard. Un premier plaide pour une origine Paléarctique suivie, au début du Tertiaire, par des migrations vers l'Amérique du Nord puis vers le Sud (Dirsh 1975). Un autre propose la région Néarctique comme le berceau des Melanoplineae, d'où ils se seraient dispersés vers les régions Paléarctique et Néotropeale (Rehn and Randell 1963, Amédégnato 1977, Carbonell 1977, Rowel and Carbonell 1977, Vickery 1987, Amédégnato 1993). Une troisième hypothèse, issue des recherches moléculaires, propose l'Amérique du Sud comme le berceau de multiples migrations des Melanoplineae vers le Nord (Amédégnato et al. 2003, Chapco et al. 2001, Chapco 2006).

Dans ce contexte, nous nous sommes proposés de clarifier l'histoire évolutive des Melanoplineae et d'évaluer la concordance entre les relations phylogénétiques suggérées par les données moléculaires et les affinités proposées par la taxonomie morphologique couramment acceptée (Rehn et Randell 1963, Amédégnato 1977) et/ou les compilations subséquentes (e.g., Otte 1995, Vickery 1997). Plus exactement, nous avons (i) évalué la monophylie de la sous-famille Melanoplineae et de ses tribus, en relation avec des ancêtres présumés d'Afrique, Eurasie et/ou des Amériques, (ii) évalué le scénario biogéographique le plus probable pour leurs évolution, (iii) daté les événements suggérés par les phylogénies, et (iv) apprécié la cohérence des classifications morphologiques et proposé des affiliations pour les taxons morphologiques « atypiques », « non classés » ou nouveaux.

Pour atteindre nos objectifs, nous avons analysé des séquences d'ADN combinés, de trois régions nucléaires, i.e. 18S, ITS1 et 28S, et une mitochondriale, i.e. 12S, provenant des 19 genres de Melanoplineae représentatifs de toutes les tribus Holarctiques et de la majorité des tribus Néotropicales. Parmi les genres analysés, 17 ont été séquencés et utilisés *pour la première fois*, pour les loci mentionnés ci-dessus, dans des analyses de phylogénie moléculaire. L'utilisation des spécimens des collections du muséum, comme source pour nos études moléculaires, nous a permis d'analyser (i) des taxons décrits *en première* par la

morphologie et « non classés » au début de nos recherches, et (ii) des taxons provenant des régions difficilement accessibles d'Amérique du Sud et Centrale.

V.2. Histoire évolutive et taxonomie des Melanoplinae

Contribution scientifique correspondant à l'article (A):

Chintauan-Marquier, I.C., Jordan, S., Berthier, B., Amédégnato, C., Pompanon, F. 2010.
Evolutionary history and taxonomy of a short-horned grasshopper subfamily: the Melanoplinae (Orthoptera: Acrididae). *Molecular Phylogenetics and Evolution*: **in press**.

V.2.1. Principaux résultats

La sous-famille Melanoplinae, en tant que groupe taxonomique actuellement accepté (e.g., Rehn et Randell 1963, Amédégnato 1977) apparaît comme polyphylétique, de par la position du genre Nord-Centre Américain, *Netrosoma*. D'ailleurs, nos résultats phylogénétiques sont le reflet de la distribution géographique actuelle des taxons analysés plutôt que celui des classifications proposées sur la base des caractères morphologiques (e.g. Rehn et Randell 1963, Otte 1995), à une exception près. Cette exception est représentée par la tribu Eurasiatique, Podismini, qui est proposée comme unité monophylétique par cette étude.

Nos résultats démontrent que, si l'on enlève le genre *Netrosoma*, les autres Melanoplinae constituent une unité monophylétique qui se divise en deux groupes vicariants, Holarctique et Néotropical.

Concernant le premier groupe – celui de l'hémisphère du Nord - nos résultats suggèrent une clade dichotomique, avec un groupe Néarctique et un Holarctique, mélangeant des Melanoplini et Dactylotini, pour le premier, et Melanoplini, Dactylotini et Podismini, pour le deuxième. A leur tour, les Melanoplinae Centre et Nord Américains (Néarctique) apparaissent

comme unité monophylétique des tribus morphologiques Dactyloini et Melanopliini (Rehn et Randell 1963 ; Amédégnato 1977).

Nous proposons aussi d'enlever les espèces sud-américaines de Melanopliinae du groupe des Dactyloini nord-américains (voir Dactyloini *sensu* Rehn and Randell 1963), et de les placer dans une tribu distincte, les Dichropli ou Dichropliini *sensu stricto*, en accord avec Amédégnato (1977).

Pour la faune Néotropicale, nos résultats étayent l'hypothèse de deux groupes monophylétiques : un qui regroupe les Dichropliini *sensu stricto*, et un autre qui réunit les taxons « atypiques » et « non-classés », principalement distribués dans les Andes, en accord avec Amédégnato (1977).

L'application de l'horloge moléculaire Bayésienne (programme Multidivtime de Thorne et Kishino 2002), dans la datation de la phylogénie, soutient que la faune actuelle des Melanopliinae a commencé à se différencier à la fin du Crétacé – début du Tertiaire, avec la plus récente séparation entre les faunes Nord Américaine et Eurasiatique datant de 65 à 27 million d'années.

V.2.2. Discussion

Biogéographie et datation. Nos résultats proposent une origine Américaine, du Paléocène Inférieur – Eocène Moyen, pour les Melanopliinae, sans pouvoir différencier avec certitude le rôle de chaque faune, Néarctique et Néotropicale, dans les migrations intercontinentales qui ont mené à la différenciation de ce groupe. Ces échanges faunistiques s'avèrent difficile à évaluer faute de preuves scientifiques pour l'existence d'un contact physique entre les Amériques au début du Tertiaire. Si nos résultats sont évasifs sur l'origine des Melanopliinae, ils sont plus clairs pour interpréter la différenciation interne du groupe, en relation avec la naissance des faunes actuelles, Holarctique et Néotropicale. D'après nos résultats, cette différenciation supposerait plusieurs phénomènes de dispersion et diversification,

allopatriques ou sympatriques, qui peuvent s'expliquer par l'évolution climatique, biotique et géotectonique du Tertiaire. Un des résultats des phénomènes d'évolution ci-dessus mentionnés serait la faune de Melanoplineae de l'Eurasie, les Podismini *sensu* Rehn and Randell (1963), qui aurait une origine Ouest Nearctique.

Correspondance entre les données morphologiques et moléculaires dans la taxonomie. Nos résultats montrent que les caractères moléculaires ne soutiennent pas les taxons définis sur des bases morphologiques, en accord avec Rehn et Randell qui soutenaient, en 1963, qu'il n'existe pas de réelle synapomorphie pour les divisions internes des Melanoplineae. Concernant la classification des taxons définis morphologiquement et bénéficiant d'un *status incerta* dans la systématique actuelle, nos résultats proposent des affiliations et fournissent une interprétation des données dans le contexte climatique et géomorphologique des Amériques pendant le Tertiaire.

V.2.3. Article - **Chintauan-Marquier, I.C.**, Jordan, S., Berthier, B., Amédégnato, C., Pompanon, F. **2010. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: the Melanoplineae (Orthoptera: Acrididae).** *Molecular Phylogenetics and Evolution: in press.*



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae)

Ioana C. Chintauan-Marquier^{a,*}, Steve Jordan^b, Pierre Berthier^c, Christiane Amédégno^d,
François Pompanon^a

^a Laboratoire d'Ecologie Alpine, Université Joseph Fourier UMR CNRS 5553, 38041 Grenoble Cedex 09, France

^b Department of Biology, Bucknell University, Lewisburg, PA 17837, USA

^c Institute of Neuroinformatics, UNI – ETH Zurich Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

^d Muséum National d'Histoire Naturelle, Département Systématique et Evolution, UMR CNRS 7205 OSEB, Entomologie, Case 50, 45, rue Buffon, 75005 Paris, France

article info

Article history:
Received 16 December 2009
Revised 29 June 2010
Accepted 7 July 2010
Available online xxx

Keywords:
Grasshopper
Acrididae
Melanoplinae
Combined phylogeny
rDNA
Molecular dating
Biogeography
Taxonomy

abstract

The Melanoplinae is one of the largest subfamilies of the Acrididae grasshoppers, with a Holarctic–Neotropical distribution. To investigate its controversial taxonomy and evolutionary history, we studied 19 species representative of its main tribes, and 7 Acridoidea outgroup species. More than 1650 base pairs of three regions of nuclear rDNA (18S, ITS1, 28S) and one mitochondrial rDNA (12S) were combined and used to construct parsimony, maximum likelihood and Bayesian phylogenies. Our results correspond with the present geographical distribution of the taxa rather than the existing taxonomy based on morphological characters. The morphologically unclassified and atypical taxa group with the Neotropical Melanoplinae. The currently recognized Melanoplinae appear to be polyphyletic due to inclusion of the Mexican genus *Netrosoma*. The distribution of the American and Eurasiatic Melanoplinae fauna can be explained by climatic and geological events, such as the Andean uplift, that would have affected the diversification and migration of Neotropical taxa.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The Melanoplinae form one of the largest subfamilies of the Acrididae, with more than 900 species distributed throughout Eurasia and America (Otte, 1995; Litzenger and Chapco, 2003). They are one of the main components of the Acrididae fauna in North America (e.g., Amédégno et al., 2003) and in the high Andes of South America (e.g., Rowell and Carbonell, 1977). They include several of the worst orthopteroid crop and grassland pests in America, such as the “two-striped grasshopper” (*Melanoplus bivittatus*), the “migratory grasshopper” (*Melanoplus sanguinipes*) and several *Dichroplini* grasshoppers (e.g., *Dichroplus elongatus*, *D. maculipennis*). The systematics of this group is controversial, with different authors recognising different numbers of tribes (e.g., Rehn and Randell, 1963; Amédégno, 1977; Vickery, 1997), and its evolutionary history remains unclear. Three scenarios have been proposed to explain the early biogeographic history of the Melanoplinae. One assumes an Eurasian origin of the group followed by migrations to North America and then South America

(Amédégno, 1977). Another posits an ancient Laurentian (southern Quebec) origin with the Nearctic region as centre from which migrations began towards the Palaearctic and the Neotropic regions (Vickery, 1987). A third alternative, based on recent molecular works (Chapco et al., 2001; Amédégno et al., 2003), suggests a South American origin followed by a substantial number of northward, eastward and/or westward migrations.

Given these diverse proposals, we aimed to clarify the evolutionary history of melanoploid grasshoppers and to evaluate the concordance between the evolutionary relationships inferred from molecular data and the affinities suggested by current morphotaxonomies (Rehn and Randell, 1963; Amédégno, 1977) and/or subsequent compilations (Otte, 1995; Vickery, 1997). More specifically, (i) we evaluated the monophyly of the Melanoplinae and of its tribes in relation to possible ancestors from Africa, Eurasia and/or the Americas, (ii) we assessed the most probable biogeographic scenario, (iii) we dated the events inferred from the phylogenies, and (iv) we checked the coherence of the current classifications and proposed an affiliation for “atypical”, “unclassified” or new morphologically described taxa.

To this end, the analysis combined nuclear and mitochondrial rDNA sequences for 19 genera representative of all the Holarctic

* Corresponding author.
E-mail address: iochintauan.marquier@gmail.com (I.C. Chintauan-Marquier).

2

I.C. Chintauan-Marquier et al. / Molecular Phylogenetics and Evolution xxx (2010) xxx–xxx

83 and most of the Neotropical tribes, of which 17 have not been previously analysed in molecular phylogenies. We also included seven closely related outgroup genera. By making use of museum specimens it was possible to include in the analysis new morphologically described taxa and taxa from relatively inaccessible regions.

88 2. Materials and methods

89 2.1. Taxon sampling

90 DNA sequences were obtained from samples of the Caelifera (Insecta, Orthoptera) collection of the Museum National d'Histoire Naturelle (MNHN) in Paris and from samples collected for this study (Table 1). To organize our sampling, we followed the classification recognising five main tribes of Melanoplinae (Amédégno, 1974, 1977; Rowell and Carbonell, 1977; see Table 1). When possible, we included a species of the type genus for each tribe (i.e., *D. elongatus*, *Melanoplus frigidus*, and *Podisma pedestris* for Dichroplini, Melanoplini and Podismini, respectively), following the classifications of Vickery (1997) for the Northern melanoplinae, and of Amédégno (1977) for the Southern tribes. The 19 species covered the present worldwide distribution and ecological diversity of the Melanoplinae. Seven other Acridoidea were included as outgroup species (Table 1).

104 2.2. DNA extraction, amplification and sequencing

105 Total genomic DNA was extracted from femoral muscles using the Qiagen DNeasy tissue kit, according to the manufacturer's instructions. The remainder of the specimen was vouchered, and recorded in a database maintained at MNHN.

109 We sequenced a total of more than 1430 base pairs (bp) of three regions of nuclear rRNA (about 508 bp of the 18S region, 256 bp of ITS1, and 319 bp of 28S) and about 350 bp of a mitochondrial gene (12S). We used the same primers for both DNA amplification and sequencing (Table S1, Supplementary online). Multiple independent PCRs were performed for each set of primers. The PCR reactions were performed in a final volume of 25 l containing 2–5 l of DNA, and 20–23 l buffer solution with 10 l M of each primer, 10 mM of each dNTP, 25 mM MgCl₂, and 5u/l of AmpliTaq Gold polymerase (Applied Biosystems, Foster City, California). PCR was performed according to the following protocol: initial activation at 95 °C for 10 min; then for 35–45 cycles: denaturation at 95 °C for 30 s, annealing (temperatures in Table S1) for 60 s, elongation at 72 °C for 2 min, and a final elongation step at 72 °C for 10 min. PCR products corresponding to a single fragment were purified using the Qiagen QIAquick PCR purification kit. Otherwise, we used a combination of gel separation and subsequent purification with the Qiagen QIAEX II gel extraction kit. To increase the quantity of target DNA for the most degraded samples, PCR products were ligated into a pGEM-T Easy vector (Promega) overnight. Then 2 l of the ligation reaction were used to transform JM109 High Efficiency Competent Cells (Promega) according to the manufacturer's recommendations. Positive clones were then used as a template for a second PCR amplification.

133 Purified PCR products were used as templates in 20 l BigDye Terminator Cycle Sequencing kit version 2 (Applied Biosystems) reactions, with 25 cycles: 96 °C for 30 s, the annealing temperature for 30 s, and 60 °C for 4 min. Excess dye terminators were removed by sephacryl/Sephadex column purification. Sequencing reactions were electrophoresed on an ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems). Sequencing reactions were carried out on both DNA strands and SeqScape 2.5 (Applied Biosystems) was used to reconcile chromatograms of complementary strands. Autapomorphies were verified on the original chromatograms. Ambiguous

143 results were checked by sequencing either a different DNA extraction from the same individual, or an extraction from another conspecific individual. However, the quality of museum-preserved specimens varied considerably, and DNA degradation did not allow the amplification of target sequences for all species (see Table 1).

2.3. Sequences analysis

2.3.1. Alignment and gap treatment

150 We generated 111 sequences that were edited in Sequence Navigator 1.01 (PE Applied Biosystems, Inc.). Multiple alignments were generated for each locus using ClustalX 1.81 (Thompson et al., 1997) with default options, and then reviewed and corrected by eye. Both the primary sequence and the secondary structure of rDNA were taken into account when aligning the sequences. Positional homology of rDNA sequence alignments was constrained manually using the structural method described in Kjer (1995), and information about the secondary structure of rDNA on insects (e.g., Kjer, 2004; Page, 2000). Ambiguous aligned regions were identified by following the method proposed by Lutzoni et al. (2000), and coded as single multistate characters with the INAAASE package. The implementation of this coding method was not problematic when using PAUP 4.0b8 (Swofford, 1995) because we identified no more than 20 states for each ambiguous aligned region. The alignments are available as Supplementary files.

2.3.2. Characteristics of molecular data

167 Base composition, substitution rates, and rate heterogeneity across sites were estimated under the most likely model of evolution suggested by the Akaike Information Criterion (AIC; Akaike, 1974) in Modeltest 3.06 (Posada and Crandall, 1998) for each dataset (Table 2). The constancy of base frequencies across taxa was evaluated with a chi-square (χ^2) "goodness-of-fit" test using PAUP[®]. Tests were performed with all sites and variable sites only (Table 2) in order to assess the potential confounding effects of invariant sites (e.g., Buckley et al., 2001).

2.4. Phylogenetic analysis

177 We conducted maximum parsimony (MP), maximum likelihood (ML) and Bayesian analyses on separate and combined data sets to address questions regarding the levels of homoplasy, incongruence, and information content within and among loci. Independent analyses were conducted on the ingroup alone to determine the effect of distant outgroups (e.g., Lyons-Weiler et al., 1998; Collin, 2003) not only on the rooting of the ingroup but also on the recovered topology within the ingroup. The trees were edited using TREEVIEW 1.3 (Page, 1996).

2.4.1. Outgroup selection

187 In order to avoid the possible effects of a distant outgroup, preliminary analyses were performed to choose the outgroups among the taxa supposed to be close relatives of Melanoplinae according to morphological criteria (mainly genitalia), including taxa from the subfamilies Catantopinae, Cyrtacanthacridinae, Proctolabinae, and Copiocerinae (see Table 1) (cf. Amédégno, 1977; Rehn and Randell, 1963). Our choice was based on tests for phylogenetic signal and preliminary MP and ML ratchet searches in PAUP[®] by means of PAUPRat v.01b (Skjes and Lewis, 2001); 10% 15% and 20% of equally weighted characters were perturbed, and 20 independent ratchet searches of 200 iterations each were performed. A consensus tree from the combined results of all these searches was then created. Moreover, in order to test the stability of the relationships between the Melanoplinae and their closely related taxa, and to increase the number of node time constraints in the dating procedures, we added three outgroups external to Acridoidea

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

Table 1
Origin of study material and references of sequence data. LECA: Laboratoire d'Ecologie Alpine, UMR CNRS-UJF 5553 Grenoble, France; MNHN: MNHN Cœlifera collection.

Taxa	Authority	Sampling place, date	Collection nos.	GenBank accession nos.			
				18S	ITS1	28S	12S
Acridoidea							
Pamphagidae							
<i>Prionotropis hystrix</i>	Uvarov, 1923	France, Bouches du Rhône, 1991	MNHN P91-01	GQ354782	GQ354705	GQ354732	GQ354758
Romaleidae, Phaenopariini							
<i>Maouliparia rotundata</i>	Hebard, 1923	Ecuador, Pichincha, 1997	MNHN E97-8	GQ354772	GQ354695	GQ354721	GQ354747
Paulinidae, Marellinae							
<i>Marellia geyskesia</i>	Willemse, 1948	Venezuela, Guarico, 1981	MNHN Ve81-01	GQ354773	GQ354696	GQ354722	GQ354748
Acrididae							
Catantopinae s.s.							
Catantopini							
<i>Oxyantops spissus</i>	Walker, 1870	Guinea, Nimba, 1984	MNHN Gu84-01	GQ354778	GQ354701	GQ354727	GQ354753
Pezotettigini							
<i>Pezotettix giornae</i>	Rossi, 1794	France, Dordogne, 2000	MNHN Fr00-01	GQ354780	GQ354702	GQ354729	GQ354755
Copiocerinae							
<i>Copioera austera</i>	Gerstaecker, 1889	Peru, Huanuco, 2001	MNHN Pe01-F111	GQ354763	GQ354686	GQ354712	GQ354738
Proctolabinae							
Coscinutini							
<i>Coscinuta</i> n. sp.	Amedegnato, det.	Ecuador, Sucumbios, 1997	MNHN E97-26	GQ354764	GQ354687	GQ354713	GQ354739
Melanoplineae							
Podismini							
<i>Italopodisma acuminata</i>	La Greca, 1969	Italy, Abruzzo, 2000	LECA IAIA01-00	GQ354770	GQ354693	GQ354719	GQ354745
<i>Podisma pedestris</i>	Linnaeus, 1758	France, Balldonne, 2000	LECA PFFB01-00	—	GQ354704	GQ354731	GQ354757
<i>Snopodisma punctata</i>	Mistshenko, 1954	Japan, Tatsugou, 1998	LECA SPUT01-98	—	GQ354708	GQ354735	GQ354760
Melanoplini							
<i>Aidemona azteca</i>	Brunner von Wattenwyl, 1893	Nicaragua, Nueva Segouia, 1986	MNHN Ni86-01	GQ354761	—	GQ354710	GQ354736
<i>Melanoplus frigidus</i>	Boheman, 1846	France, Lautaret, 1999	LECA MFFL01-99	GQ354774	GQ354697	GQ354723	GQ354749
<i>Netrosoma xanthops</i>	Scudder, 1897	Mexico, Hidalgo, 1986	MNHN Me86-01	GQ354777	GQ354700	GQ354726	GQ354752
<i>Phaulotettix compressus</i>	Scudder, 1897	Mexico, Hidalgo, 1986	MNHN Me86-02	GQ354781	GQ354703	GQ354730	GQ354756
Dactylotini							
<i>Dactylotum bicolor bicolor</i>	Charpentier, 1843	Mexico, Hidalgo, 1986	MNHN Me86-04	GQ354765	GQ354688	GQ354714	GQ354740
<i>Hesperotettix viridis viridis</i>	Thomas, 1872	Mexico, San Luis Potosi, 1986	MNHN Me86-05	GQ354769	GQ354692	GQ354718	GQ354744
Jvarini							
n. gen n. sp. 1	Amedegnato, det.	Peru, Cajamarca, 2001	MNHN Pe01-18	GQ354771	GQ354694	GQ354720	GQ354746
Dichropilini							
<i>Chibchacris bordoni</i>	Hebard, 1923	Venezuela, Merida, 1984	MNHN Ve84-01	GQ354762	GQ354709	GQ354711	GQ354737
<i>Dichropilus (Baecris) punctulatus</i>	Thunberg, 1824	Ecuador, Tungurahua, 1997	MNHN E97-4	GQ354766	GQ354689	GQ354715	GQ354741
<i>Hazelacris nigrovittatus</i>	Ronderos, 1981	Ecuador, Manabi, 1995	MNHN E95-01	GQ354768	GQ354691	GQ354717	GQ354743
<i>Neopodis brunneri</i>	Hebard, 1931	Argentina, El Chaco, 1972	MNHN Ar72-02	GQ354776	GQ354699	GQ354725	GQ354751
<i>Scotussa impudica</i>	Giglio-Tos, 1894	Brazil, Parana, 1972	MNHN Br72-02	GQ354784	GQ354707	GQ354734	—
«Atypical» or unclassified							
<i>Meridacris subaptera</i>	Roberts, 1937	Venezuela, Merida, 1985	MNHN Ve85-01	GQ354775	GQ354698	GQ354724	GQ354750
New genus	Amedegnato, det.	Mexico, Guerrero, 1986	MNHN Me86-06	GQ354767	GQ354690	GQ354716	GQ354742
<i>Pediella</i> n. sp.	Amedegnato, det.	Peru, Cajamarca, 2001	MNHN Pe01-01	GQ354779	—	GQ354728	GQ354754
<i>Propedies</i> n. sp. 2	Amedegnato, det.	Brazil, Mato Grosso, 2000	MNHN Br00-01	GQ354783	GQ354706	GQ354733	GQ354759

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplineae (Orthoptera: Acridoidea). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

4

I.C. Chintauan-Marquier et al. / Molecular Phylogenetics and Evolution xxx (2010) xxx–xxx

Table 2
Descriptive statistics for the different sequence datasets with the Acrididae outgroups included.

Loci	No. taxa	No. characters	% variable sites	No. Parsimony Informative sites	Model of evolution ^b	AIC	(Å) In L	Average frequencies				Constancy of base frequencies (V ² and P values for df = 69)	
								A	C	G	T	All sites	Variable sites
Ingroups													
18S + ITS1 + 28S	19												
AARs + UARs ^a		1319	28.58	243								32.5 ; P = 0.99	34.6 ; P = 0.98
UARs		1030	8.93	24	K81uf + G + I	4196.9	2091.45	0.21	0.26	0.3	0.21	17.8 ; P = 0.99	19.8 ; P = 0.99
12S	18	350	26.86	52	TIM + G + I	2799.54	1391.77	0.42	0.16	0.09	0.31	8.2 ; P = 1	32.2 ; P = 0.98
Combined data	19												
AARs + UARs ^a		1669	28.22	295								48.4 ; P = 0.68	58.4 ; P = 0.31
UARs		1380	13.48	76	SYM + G + I	7419.69	3702.84	Equal frequencies				37.2 ; P = 0.96	45.5 ; P = 0.79
Ingroups + Outgroups													
18S + ITS1 + 28S	24												
AARs + UARs ^a		1319	29.42	266								41.4 ; P = 0.99	53.4 ; P = 0.91
UARs		1030	9.61	32	HKY + G + I	4476.84	2232.42	0.22	0.26	0.3	0.22	20.5 ; P = 1	26.9 ; P = 0.99
12S	23	350	39.43	69	GTR + G + I	3851.45	1915.72	0.44	0.14	0.08	0.34	14.9 ; P = 1	39.2 ; P = 0.99
Combined data	24												
AARs + UARs ^a		1669	31.02	335								63.8 ; P = 0.65	108.3 ; P < 0.05
UARs		1380	17.17	101	SYM + G + I	8948.29	4467.15	Equal frequencies				43.6 ; P = 0.99	57.5 ; P = 0.83

^a After removal of primer sequences. AAR: ambiguous aligned region, UAR: Unambiguous aligned region.

^b Selected with the AIC criterion in Modeltest 3.0 (Posada and Crandall, 1998); df = degrees of freedom; g1 = skewness of tree-length distribution.

203 but within the Acridoidea: Maculiparia, Marellia and Prionotropis
204 (Table 1).

205 2.4.2. Maximum parsimony

206 MP analyses were carried out on the ingroups alone and on the
207 ingroups + outgroups. Heuristic searches on nuclear and mitochon-
208 drial data, separated or combined, were implemented in PAUP^r
209 under the MP criteria with accelerated transformation (ACCTRAN)
210 option to optimize the state of unordered (Fitch) characters, 100
211 random sequence addition replicates, TBR branch swapping and
212 gaps treated as a fifth character. Potentially parsimony-informative
213 indel regions were coded as separate multistate characters at the
214 end of the data matrix. The unambiguously aligned regions were
215 analysed separately or in combination with ambiguously aligned
216 regions. Equally weighted analyses were conducted for 12S and
217 the combined data. Strict consensus trees were then constructed.
218 When the equally parsimonious trees had very different topologies
219 and their strict consensus tree had a low resolution, we opted for
220 the equally parsimonious tree with the highest negative log-likeli-
221 hood (Å In L) using the Shimodaira–Hasegawa test (Shimodaira and
222 Hasegawa, 1999) implemented in CONSEL 0.1i (Shimodaira and
223 Hasegawa, 2001) with ten sets of 100,000 bootstrap replicates.
224 To explore “tree islands” in a multidimensional scaling space,
225 based on tree-to-tree distance metrics, we used the “Tree Set Visu-
226 alization 2.1” module (Amenta and Klingner, 2002) in “Mesquite”
227 package (Maddison and Maddison, 2006) as suggested by Hillis
228 et al. (2005).

229 2.4.3. Maximum likelihood

230 ML searches were conducted with PAUP^r. We first evaluated the
231 models of DNA substitution that best fit each data partition (i.e.,
232 separated nuclear and mitochondrial or combined datasets) fol-

233 lowing the AIC implemented in Modeltest 3.06 (Posada and
234 Crandall, 1998) with the default settings and $\alpha = 0.01$. In all ML
235 searches, we estimated the appropriate ML model parameters
236 using the data partition of interest and a neighbour-joining (NJ)
237 topology. These estimated values were fixed, and a full heuristic
238 ML search was conducted with 10 random addition sequence re-
239 plicates, retaining all minimal trees, and TBR branch swapping. Then
240 we re-estimated model parameters on the ML tree and used these
241 new values for a new search. We repeated this process until the
242 new search found the same topology as the previous search and
243 all parameter estimates were identical.

244 2.4.4. Bayesian analyses

245 Bayesian Markov chain Monte Carlo (MCMC) analyses were per-
246 formed on unambiguously aligned regions with MrBayes 3.0b4
247 (Ronquist and Huelsenbeck, 2003). We performed two simulta-
248 neous independent Metropolis-coupled MCMC runs starting from
249 different random points. We set each run to have one cold and five
250 heated chains to allow better mixing of the MCMC chain and mini-
251 mize the chance of being trapped in local optima (Pereira et al.,
252 2007). We performed runs assuming the same topology to be
253 shared among all partitions, and each partition allowed to have
254 its own model of DNA evolution according to the AIC. MCMC sam-
255 ples were taken in every 500th cycle. We plotted the log-likelihood
256 of sampled topologies to determine the burn-in period in which
257 the MCMC had reached a stationary status (Huelsenbeck and
258 Ronquist, 2001). Post-burn-in samples from the independent runs
259 were used to construct a 50% majority rule consensus trees. The
260 proportion of trees in which nodes were recovered after the
261 burn-in period was interpreted as the posterior probability of that
262 node, or the probability that node was true (Ronquist and
263 Huelsenbeck, 2003).

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

264 2.4.5. Node support

265 Support for internal branches was assessed in PAUP^{*} by non-
266 parametric bootstrapping (Felsenstein, 1985; Efron et al., 1996)
267 with 1000–2000 pseudoreplicates for MP and 100–500 pseudore-
268 plicates for ML, using full heuristic searches with 10–20 random
269 addition sequence replicates; TBR branch swapping was imple-
270 mented and one tree held at each step during stepwise addition
271 (e.g., DeBry and Olmstead, 2000). Posterior probabilities for indi-
272 vidual nodes were estimated in a Bayesian MCMC framework as
273 presented above. For ease of visual comparison to nonparametric
274 bootstrap values we present these probabilities as numbers rang-
275 ing from 0 to 100 (Fig. 1).

276 2.4.6. Evaluation of alternative phylogenetic hypotheses

277 Three tests were performed to assess the extent to which our
278 data conflict with a priori hypotheses represented by candidate
279 trees estimated under ML using conventional, backbone, or re-
280 versed constraints in PAUP^{*} and AutoDecay programs (Eriksson,
281 1998). The weighted Shimodaira–Hasegawa test (WSH; Shimoda-
282 ira and Hasegawa, 1999), the approximately unbiased test (AU;
283 Shimodaira, 2002) and the Kishino–Hasegawa test (KH; Hasegawa
284 and Kishino, 1989) were implemented in CONSEL 0.1i (Shimodaira

and Hasegawa, 2001) with ten sets of 100,000 bootstrap replicates
285 each.

287 2.5. Biogeographic framework

288 The history of the distribution of Melanoplinae was recon-
289 structed applying an event-based parsimony method using DIVA
290 1.1 (Ronquist, 1996). An exact search, according to the dispersal-
291 vicariance optimization proposed by Ronquist (1997) and the max-
292 imum number of unit areas allowed in ancestral distributions set to
293 seven was applied to the data with all outgroups included in order to
294 restrict the ancestral distribution at the basal node (Sanmartin,
295 2003; Donato, 2006). DIVA 1.1 assumes that speciation is caused
296 by vicariance and only imposes a cost for a dispersal or extinction
297 event. The areas considered in the interpretation corresponded to
298 the Ethiopian, Palaearctic, Nearctic, and Neotropical regions accord-
299 ing to the zoogeographical reconstructions of Gentry and Sutcliffe
300 (1981). The Holarctic was divided into four infraregions, the Eastern
301 and Western Nearctic and the Eastern and Western Palaearctic. We
302 followed Sanmartin et al. (2001), especially in considering the tem-
303 perate Mexican distributions as Western Nearctic. The Neotropical
304 region was divided into South America (i.e., Brazilian shield) and

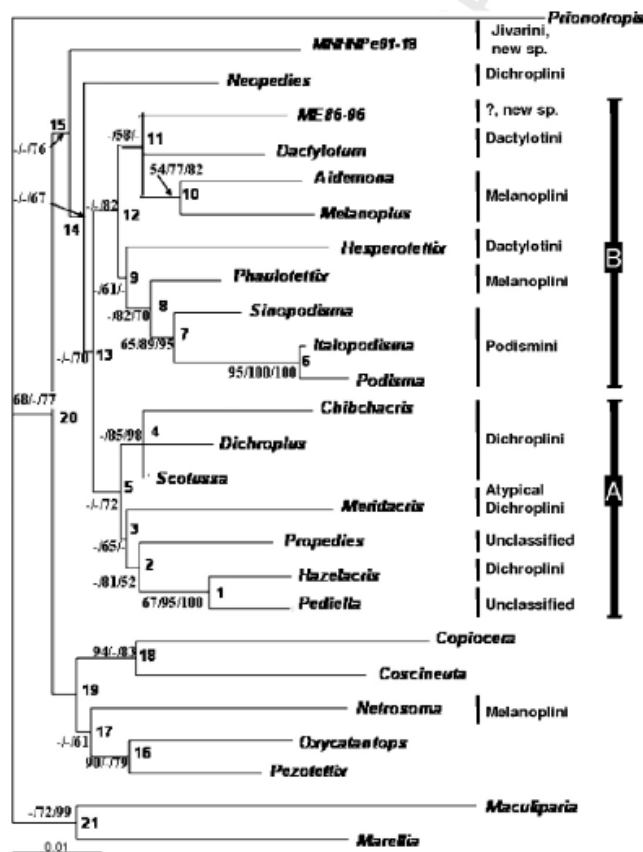


Fig. 1. Maximum likelihood tree for combined data. Numbers on branches correspond to the nonparametric bootstrap values and posterior probabilities for the MP, ML and Bayesian analyses, respectively. A and B corresponds to the Neotropical (node 5) and Holarctic (node 12) clades, respectively.

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

6

I.C. Chintauan-Marquier et al. / Molecular Phylogenetics and Evolution xxx (2010) xxx–xxx

305 Northern South America (i.e., the northern Andes range extended
306 over Ecuador, Colombia, and Venezuela). The present geographical
307 distributions of the selected genera of Melanoplinae (Table 1) are
308 available as Supplementary file (Table S2).

309 2.6. Temporal diversification

310 Dating was based on the ML analysis, which was the best a posteriori
311 hypothesis selected by the weighted Shimodaira–Hasegawa
312 test (Shimodaira and Hasegawa, 1999) in CONSEL 0.1. Thus, we
313 used the topology produced by the ML analysis with all outgroups
314 included, and the ML topology with the Acrididae outgroups imposed
315 as constraint. A Likelihood Ratio Test revealed that none of the data
316 partitions conformed to the molecular clock hypothesis (in all the cases
317 $P < 0.05$ and $\chi^2 > 33.92$ for 22° of freedom. Therefore, we estimated
318 the age of nodes and the 95% credibility intervals from our combined
319 DNA sequence data using the Bayesian molecular clock method
320 implemented in MultiDivTime (Thorne and Kishino, 2002). The
321 assumption of constant rates over time is not required for this method,
322 and multiple constraints on node ages are allowed. We extended our
323 data set to obtain additional calibration points (see “Outgroup selection”) and
324 we used time constraints inferred by previous morphological and
325 molecular studies. We considered as time constraints (Fig. 1): (1) the
326 emergence of the modern acridian fauna (node 22) between the Late
327 Palaeogene (Carbonell, 1977; Amédégno, 1993; Jørgo, 1979) 23
328 MYA, and the Upper Cretaceous (Storozhenko in Gangwere et al.,
329 1997) 99 MYA, (2) the divergence between the Dichroplini and
330 Jvarini tribes (node 15) 66–99 MYA (Amédégno et al., 2003),
331 (3) the possible incursion of the South-American melanoplinines into
332 North-America (node 13) during the Late Cretaceous–Early Palaeocene
333 61–99 MYA (Chapco et al., 2001; Amédégno et al., 2003),
334 and (4) the incursion of the North-American melanoplinines into
335 Eurasia (node 8) during the Palaeocene 54.8–65 MYA (Chapco
336 et al., 2001; Amédégno et al., 2003). The ages of the geological
337 periods and chronostratigraphic references were taken from the
338 1999 Geological Time Scale of the Geological Society of America
339 (www.geosociety.org/science/timescale/timescl.pdf).

340 We did not use fossils for imposing time constraints because of
341 the absence of fossils for certain Melanoplinae and of the questionable
342 interpretation of Acridian fossils (e.g., Arillo and Ortuño, 1997).
343 We followed a conservative approach in using calibration references
344 (e.g., Bossuyt et al., 2006). We repeated the analyses for three
345 combinations of calibration points: (1) all calibrations points, (2)
346 calibration points one to three, and (3) calibration point four. All
347 analyses were duplicated to confirm successful convergence for the
348 estimation of divergence ages.
349

350 3. Results

351 3.1. Sequence analyses

352 When aligning the sequences (with or without outgroups), we
353 could not determine the homology for six indel-rich regions inside
354 the ITS1 fragment (positions on the combined sequence: 538–633,
355 669–695, 718–767, 792–805, 826–871 and 903–964). The
356 exclusion of these ambiguously aligned regions resulted in the loss
357 of 17.02% of the sites (283 over 1663 bp), and of 69.8% of the
358 potentially informative sites for the combined data set (Table 2).
359 After the removal of the ambiguous aligned regions, the mitochondrial
360 DNA remained more variable than the nuclear DNA (39.43% and
361 9.61% variable sites respectively, Table 2). A high content of A/T
362 (73–78%) for the mitochondrial DNA was noted for each taxon.
363 The proportion of transitions was typically high among closely
364 related taxa, decreasing progressively with increasing genetic diver-

gence (not shown). Globally, base frequencies were constant for
either all sites or variable sites only (Table 2). Models of evolution
and associated parameters estimates for the ML analysis are given
in Table 2.

3.2. Phylogenetic analyses

For all phylogenetic methods, combining the nuclear and mitochondrial data improved the resolution and nodal support. For MP analyses, adding the Acrididae outgroups (i.e., Palaearctic Catantopinae, Neotropical Proctolabinae and Copiocerinae) or non-Acrididae outgroups (i.e., Pamphagidae, Pauliniidae and Romaleidae) did not modify the basic ingroup topologies for the Holarctic and Neotropical faunas. Moreover, varying factors such as the presence of autapomorphies or of invariant sites had no significant effect on tree topologies. For both datasets (i.e., including or excluding outgroups) appending the ambiguously aligned regions to the unambiguously aligned regions modified the MP topologies. The trees became significantly longer (mean of the D tree length = 1205.70 ± 232.18 ; $P = 0.05$) with a significantly lower likelihood (mean of the $D(\Delta \ln L) = 25.82 \pm 4.15$; $P = 0.05$). A weighted Shimodaira–Hasegawa test, using the combined data set and the corresponding model selected by the AIC, showed that the best a posteriori topology was the ML one (Fig. 1). Similar topologies were obtained with the MP and Bayesian analyses (data not shown).

All the 18 nodes connecting Melanoplinae taxa were supported by posterior probabilities and/or bootstrap values higher than 50% (Fig. 1) with one exception (node 19). The Bayesian posterior probabilities for the nodes were generally higher than the bootstrap values. The Bayesian analysis gives strong supports for basal (i.e., nodes 15 and 13) and middle level (i.e., nodes 12 and 5) nodes that are not supported in the ML analysis. We tested one possible source of these discrepancies, i.e., model misspecification (c.f., Buckley et al., 2001; Wróbel, 2008), by using different models of evolution in the ML analyses (data not shown). This did not reconcile the differences between the two methods in supporting the same topology. Thus, we considered the posterior probabilities and bootstrap values as potential upper and lower bounds of node support, as suggested by Douady et al. (2003).

The monophyly of morphologically defined taxa (mainly tribes) and of groups based on the geographic distribution of the genera were tested given the combined phylogeny (Table 4). The congruence of the different tests was good since they gave comparable results for all hypotheses but two. The Melanoplinae, in its current taxonomic sense (Otte, 1995), appears to be polyphyletic because the genus *Netrosoma* seems to be closer to taxa from the subfamilies Catantopinae sensu stricto (Fig. 1, node 17), Copiocerinae and Proctolabinae (Fig. 1, node 19). However, the support for the basal branching of the tree is low hence unequivocal conclusions cannot be drawn. Our results (Fig. 1) support the monophyly of the Podismini (Rehn and Randell, 1963) (node 7) but not of the Melanoplini (Rehn and Randell, 1963), Dactyloini (Rehn and Randell, 1963; Otte, 1995), and Dichroplini (Amédégno, 1977). Our results also suggest a high divergence between the Jvarini (MNHNPe01-18, node 15) and Neopedies (node 14) on one hand, and the other Dichroplini (Amédégno et al., 2003) on the other hand. The existence of a monophyletic group including “atypical” and unclassified taxa (i.e., *Meridacris*, *Pediella* and *Propedies*; node 2) in the Neotropical fauna is also supported (Fig. 1). Additional analyses (data not shown) involving the gradual exclusion of “unclassified” taxa suggest that the Andean genus *Pediella* is more related to the Nearctic melanoplinines than to some Dichroplini sensu stricto (i.e., *Dichroplus*, *Scotussa* and *Chibchacris*), and that the Brazilian genus *Propedies* is closest to Dichroplini s.s.

Thus it appears that the clades supported by the phylogenies are not in accordance with the existing classifications, but

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

429 instead are congruent with the geographic distribution of these
430 taxa. Our results (Fig. 1) support the existence of a South Amer-
431 ican clade (A: node 5) composed of two monophyletic groups.
432 The first one (i.e., *Chibchacris*, *Dichroplus* and *Scotussa*; node 4)
433 is strongly supported by the ML and Bayesian methods. The second
434 includes *Hazelacris* and the “atypical” and “unclassified” taxa
435 (node 3).

436 The existence of a monophyletic Northern (i.e., Nearctic and
437 Palaearctic) group is also supported (clade B: node 12). It is divided
438 into a Nearctic clade (*Me86-06*, *Dactyloptum*, *Aidemona* and *Melanoplus*;
439 node 11), and a Holarctic clade (*Hesperotettix*, *Phaulotettix*,
440 *Snopodisma*, *Italopodisma* and *Podisma*; node 9). The Eurasian
441 *Melanoplinae* make a monophyletic group (node 7) highly
442 supported in all analyses, with the genus *Phaulotettix* as a sister
443 taxon (node 8). The genus *Aidemona* that was previously proposed
444 to have a basal position (Amédégno et al., 2003) appears in a ter-
445 minal position.

3.3. Biogeographic analysis

447 A DIVA exact search resulted in 180 equally parsimonious
448 reconstructions that differ only at two nodes (8 and 20; Fig. 2).
449 The unambiguous DIVA optimal area reconstructions at each
450 ancestral node are summarized in Fig. 2. DIVA suggested six vicar-
451 iance events (allopatric speciations, nodes 2, 4, 5, 7, 13, 15) and
452 seven duplication events that mainly affect terminal branches
453 (sympatric speciations, nodes 1, 4, 6, 10–12, 14). Eight nodes
454 (3–5, 9–11, 13) correspond to possible dispersals and a single
455 extinction would have occurred (node 13).

3.4. Temporal diversification

456 The three combinations of the calibration points used give compar-
457 able results (Table 3). The most recent common ancestor
458 between some of the most divergent *Melanoplinae* that are studied
459

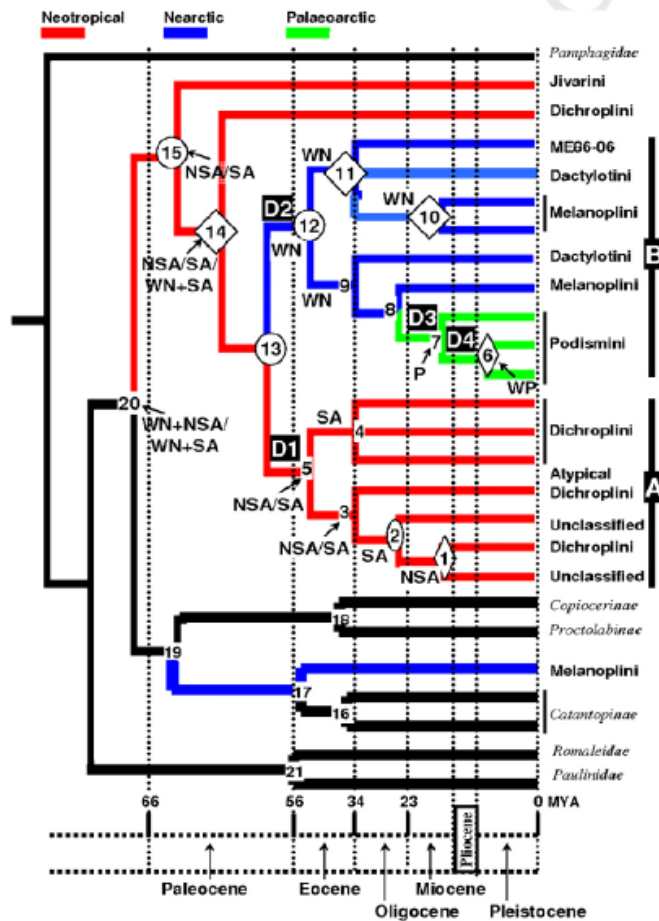


Fig. 2. Chronogram of *Melanoplinae* and reconstruction of ancestral areas as suggested by DIVA. Western (WP) and Eastern (EP) Palaearctic, Western (WN) and Eastern (EN) Nearctic, Northern South America (NSA) and South America (SA). Clades A and B as in Fig. 1. Branch colors indicate the biogeographical distribution of the taxa and their hypothetical ancestors: red – Neotropical, blue – Nearctic, green – Palaearctic. Numbers indicate nodes. Letters to nodes (D1–D4) indicate diversification events as represented in Fig. 3. Symbols: circle: vicariance; rhomb: duplication (sympatric speciation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The *Melanoplinae* (Orthoptera: Acrididae). *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.ympev.2010.07.003

Table 3
Bayesian estimation of divergence times using the MCMC method in the "Multidivtime" package of Thorne and Kishino (2002). Node numbers correspond to Figs. 1 and 2. 95% CI = 95% confidence interval. The three estimates represent use of the following calibrations: (1) all calibration points; (2) calibration points 1–3; (3) calibration point 4.

Node	Age (million of years)					
	Estimate 1		Estimate 2		Estimate 3	
	Mean	95%CI	Mean	95%CI	Mean	95%CI
1	28.84	12.02–48.42	26.32	11.11–43.42	30.80	12.37–52.96
2	51.19	35.85–66.77	46.89	33.86–60.51	54.88	37.61–75.02
3	57.82	43.69–71.86	53.02	40.81–65.71	62.03	45.42–82.76
4	49.7	34.36–65.55	45.38	32.12–59.53	53.16	35.93–73.92
5	62.42	48.93–76.02	57.26	46.29–69.44	66.97	50.78–87.74
6	8.1	0.54–20.95	6.18	0.36–16.33	8.36	0.54–21.62
7	44.93	29.44–57.47	32.07	17.65–47.98	45.29	29.59–58.12
8	57.66	54.88–63.54	40.73	27.30–55.76	58.21	54.90–64.21
9	63.12	56.43–72.61	51.15	38.05–64.42	65.76	56.59–80.09
10	44.43	26.10–62.26	38.55	22.59–54.71	46.83	27.53–67.76
11	59.28	47.16–71.84	51.49	39.27–64.44	62.71	48.40–80.26
12	66.88	58.80–77.28	57.96	47.40–69.92	70.69	59.36–87.85
13	71.45	62.47–82.64	65.67	61.14–75.98	76.59	62.40–96.67
14	74.88	64.80–86.55	69.31	62.44–80.69	81	65.26–102.86
15	79.82	68.33–91.74	74.74	66.65–86.90	87.43	69.41–112.66
16	54.87	36.68–73.22	51.42	35.08–69.11	60.59	39.54–86.54
17	70.43	54.68–85.36	66.39	52.05–81.76	78.11	57.14–105.01
18	62.62	43.75–81.30	59.47	42.26–78.10	69.54	46.08–98.15
19	77.46	63.42–90.91	73.23	60.35–87.48	86.00	65.01–113.34
20	84.49	72.09–95.84	79.87	69.35–92.72	93.79	73.29–121.41
21	74.59	55.30–92.61	72.58	53.40–90.98	86.38	58.99–120.39
22	89.99	76.53–98.59	86.87	73.67–98.20	103.36	78.98–136.37

here (node 20) is likely to have existed about 69 million years ago (MYA) (Fig. 2). The present fauna would have evolved from a common ancestor of Melanoplinae sensu Rehn and Randell, 1963 (node 15), Copiocerinae and Proctolabinae (node 18), Catantopinae sensu stricto (node 16) and perhaps Netrosoina (node 17). This would have begun in the late Cretaceous–early Tertiary, with the most recent split between the North American and Eurasian faunas (node 8) occurring 64–27 MYA. Melanoploid lineages within the zoogeographical regions have been diverging for at least 17 MYA for Eurasia (node 7), 47 MYA for North America (node 12), and 46 MYA for South America (node 5).

4. Discussion

4.1. Biogeography and dating

The polyphyly of the Melanoplinae in the molecular phylogeny is due to the inclusion of the Nearctic genus *Netrosoina* (node 17: 105–52 MYA, Table 3). The molecular data indicate that *Netrosoina* is actually related to the Catantopinae sensu stricto (Amédégno, 1977) (node 16, Fig. 2) from the Ethiopian (Oxyacantops) and Eurasian (Pezotettix) regions. Thus, this genus appears to be a descendent of a tropical acridian ancestral fauna, which evolved, in the late Cretaceous–Early Palaeocene, in the landmasses connected to the South of the North America. However, the position of *Netrosoina* is not strongly supported and these conclusions require confirmation by further studies. In this context, we will focus now on the monophyletic Melanoplinae (node 15).

Three scenarios have previously been proposed to explain the origin of the Melanoplinae. They are: (i) an ancient Palaeotropical origin followed by migrations to North America and then to South America (Vickery, 1987), (ii) an ancient Nearctic origin followed by migrations towards the Palaearctic and Neotropical regions (Rehn and Randell, 1963; Amédégno, 1977, 1993; Carbonell, 1977; Rowell and Carbonell, 1977), (iii) a South American origin followed by several northward and eastward migrations (Chapco et al., 2001; Amédégno et al., 2003; Chapco, 2006). Our results (Fig. 2) sup-

port an American origin (scenarios (ii) and (iii)) for the two clades of vicariant South American (clade A) and North-American Melanoplinae (clade B). Their ancestor could have migrated either from South America (NSA/SA) to the North or from the western North America (WN) to the South (Fig. 3). Both scenarios involve inter-continental dispersal events (nodes 15 ? 14 ? 13 levels) during the early Paleocene–mid Eocene. During this period, the evolution of the American Melanoplinae would involve vicariant diversification events (duplication) in situ (diversification centers D1 and D2, Fig. 3), with at least three lineage-splittings (nodes 15, 14 and 13, Fig. 2) in less than 20 millions years. These scenarios are concordant with the DIVA analysis but raise the perennial problem of explaining proposed ancient exchanges between North and South America. They would appear unlikely given the paucity of direct evidence of a physical connection between the Americas in this period. This discrepancy could result from the low reliability of the DIVA analysis for basal nodes (Ronquist, 1996), and/or imprecise estimates of the timing of these events.

If the exact roles of the Nearctic and Neotropical fauna in the origin of the Melanoplinae remain questionable, our results give much more robust insights into the subsequent diversification of this subfamily. Several independent diversifications (Fig. 3) would have affected the ancestors of the Melanoplinae from each part of the American continent (likelihood tests 5–7, Table 4), accounting for eight cladogenetic events equally distributed between the South (nodes 5, 4, 3, 2) and North America (nodes 12, 11, 9, 8). These divergences would have lasted less than 15 and 25 MYA for the South and North America, respectively (Fig. 2 and Table 3). The present distribution of several genera would reflect recent dispersals from the North to Southern Central America (i.e., *Melanoplus*, *Aidemona*) and from the South to Northern Central America (i.e., *Dichroplus*). Considering the Palaearctic fauna, the parphyly between the Eurasian (node 7) and Nearctic taxa such as *Hesperotettix* (node 9: 80–38 MYA) and *Phaulotettix* (node 8: 64–27 MYA) suggests that the Western Nearctic fauna were at the origin of the Eurasian fauna (node 8, Table 4, likelihood tests 4–8) about 64–27 MYA (Fig. 3). This explanation requires dispersal events that would have been possible over known contemporary land bridges (Thulean and Bering) between Eurasia and North America. *M. frigidus* (node 10: 68–22 MYA, Fig. 2) was unique in having a Holarctic

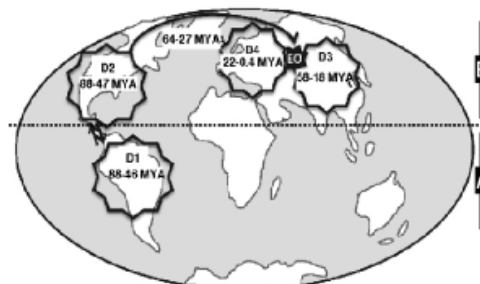


Fig. 3. Biogeographic scenario for the evolution of Melanoplinae. The diagram represents the main Cenozoic migration and diversification events inferred from the combined rDNA phylogeny (Fig. 1) and an event-based parsimony method using DIVA 1.1. Datations result from the analysis of combined DNA sequences using the Bayesian molecular clock method implemented in MultiDivtime (see Table 3). A and B refer to the origin and diversification events affecting the clades A and B as in Fig. 1. D1 and D2: diversification in situ of the descendants of South- and North-American ancestors of the Melanoplinae, followed by northward or southward migrations through Central America. EO: Origin of the Eurasian Melanoplinae from a West-Nearctic ancestor. D3 and D4: diversification of the Asian and respectively European descendants from EO.

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.07.003

10

I.C. Chintauan-Marquier et al. / Molecular Phylogenetics and Evolution xxx (2010) xxx–xxx

614 ancestor close to Propedies. The constant phylogenetic association
615 between the "unclassified" *Pediella* and the "atypical *Dichropli*"
616 *Hazelacris* could be an artefact induced by the absence in our anal-
617 yses of the mitochondrial fragment for the former taxon, hence fur-
618 ther investigations are necessary to confirm these relationships.
619 They include the oligonucleotides used for sequencing (Table
620 S1), the alignment files in nexus format for the four sequenced
621 regions, and the geographical distribution matrix used for DIVA
622 (Table S2).

623 5. Uncited references

624 Brower and DeSalle (1994), Buckley and Cunningham (2002),
625 Chapco et al. (1999), Crisci et al. (1991), Cunningham (1997), Farris
626 (1969), Farris et al. (1994), Flook and Rowell (1997), Flook et al.
627 (1995), Fries et al. (2007), Hipp (2004), Sanmartin and Ronquist
628 (2004), Schmidt et al. (2002), Simon et al. (1994, 2006), Strimmer
629 and von Haeseler (1997) and Whitfield and Kjer (2008).

630 Acknowledgments

631 We thank all those who collected or sent specimens for the
632 present study: Simon Poulain, William Chapco, Paolo Fontana, Haru-
633 ki Tatsuta, A. Foucart, Stefan Kis, Ioan Chintauan, Jérôme and Elisa
634 C. Marquier, Florin Cosoiu, Alexandru and Ionel Aluas, Eugen Smol-
635 ka, Olimpiu Ionasca. We are very grateful to Ludovic Gielly, Chris-
636 tian Miquel and Delphine Roux for their precious technical
637 support and to Richard A. Nichols for his critical reading of the
638 manuscript and help with the English. ICCM was supported by
639 grants from the French Embassy in Bucharest and the INCA pro-
640 gram of the region Rhône-Alpes (France).

641 Appendix A. Supplementary material

642 Supplementary data associated with this article can be found, in
643 the online version, at doi:10.1016/j.ympev.2010.07.003.

644 References

645 Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans.*
646 *Automat. Control* 19, 716–723.
647 Amédégato, C., 1974. Les genres d'Acridiens neotropicaux, leur classification par
648 familles, sous familles et tribus. *Acrida* 3, 193–204.
649 Amédégato, C., 1977. Etude des Acridoidea centre et sud américains (Catantopinae
650 sensu lato). Anatomie des genitalia, classification, répartition, phylogénie.
651 Unpublished Thesis, University of Paris VI, 385 pp.
652 Amédégato, C., 1993. African–American relationships in the acridians (Insecta,
653 Orthoptera). In: George, W., Lavocat, R. (Eds.), *The Africa–South America*
654 *Connection*. Clarendon Press, Oxford, pp. 59–75.
655 Amédégato, C., Chapco, W., Litzenberger, G., 2003. Out of South America?
656 Additional evidence for a southern origin of melanopline grasshoppers. *Mol.*
657 *Phylogenet. Evol.* 29, 115–119.
658 Amenta, N., Klingner, J., 2002. Case study: visualising sets of evolutionary trees. In:
659 8th IEEE Symposium on Information Visualization 2002, pp. 71–74.
660 Arillo, A., Ortuño, V.M., 1997. The fossil Acrididae from the Oligocene of Izarra
661 (Alava, Spain). The antiquity of gregarious behavior (Orthoptera, Caelifera).
662 *Geobios* 30, 231–234.
663 Bossuyt, F., Brown, R.M., Hillis, D.M., Cannatella, D.C., Milinkovitch, M.C., 2006.
664 Phylogeny and biogeography of a cosmopolitan frog radiation: late Cretaceous
665 diversification resulted in continent-scale endemism in the family Ranidae.
666 *Syst. Biol.* 55, 579–594.
667 Brower, A.V., DeSalle, R., 1994. Practical and theoretical considerations for choice of
668 a DNA sequence region in insect molecular systematics, with a short review of
669 published studies using nuclear gene regions. *Ann. Entomol. Soc. Am.* 87, 702–
670 716.
671 Buckley, T.R., Cunningham, C.W., 2002. The effects of nucleotide substitution model
672 assumptions on estimates of nonparametric bootstrap support. *Mol. Biol. Evol.*
673 19, 394–405.
674 Buckley, T.R., Simon, C., Chambers, G.K., 2001. Exploring among-site rate variation
675 models in a maximum likelihood framework using empirical data: effects of
676 model assumptions on estimates of topology, branch lengths, and bootstrap
677 support. *Syst. Biol.* 50, 67–86.

678 Caronell, C.S., 1977. Origin, evolution, and distribution of the Neotropical
679 acridomorph fauna (Orthoptera): a preliminary hypothesis. *Rev. Soc. Entomol.*
680 *Argentina* 36, 153–175.
681 Chapco, W., 2006. A note on the phylogenetic position of *Duartetix montanus*
682 within the subfamily Melanoplineae. *J. Orthoptera Res.* 15, 59–63.
683 Chapco, W., Kuperus, W.R., Litzenberger, G., 1999. Molecular phylogeny of
684 Melanopline grasshoppers (Orthoptera: Acrididae): the genus *Melanoplus*.
685 *Ann. Entomol. Soc. Am.* 92, 617–623.
686 Chapco, W., Litzenberger, G., Kuperus, W.R., 2001. A molecular biogeographic
687 analysis of the relationships between North American melanopline
688 grasshoppers and their Eurasian and South American relatives. *Mol.*
689 *Phylogenet. Evol.* 18, 460–466.
690 Collin, R., 2003. Phylogenetic relationships among Calyptraeidae gastropods and their
691 implications for the biogeography of marine speciation. *Syst. Biol.* 52, 618–640.
692 Crisci, J., Cigliano, M.M., Morrone, J.J., Roig-Juñent, S., 1991. Historical biogeography
693 of southern South America. *Syst. Zool.* 40, 152–171.
694 Cunningham, C.W., 1997. Is congruence between data partitions a reliable predictor
695 of phylogenetic accuracy? Empirically testing an iterative procedure for
696 choosing among phylogenetic methods. *Syst. Biol.* 46, 464–478.
697 DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort
698 in bootstrap resampling analysis. *Syst. Biol.* 49, 171–179.
699 Dirsh, V.M., 1975. Classification of the Acridomorphoid Insects. *Gasssey Ltd.*,
700 Faringdon, Oxon.
701 Donato, M., 2006. Historical biogeography of the family Tristiridae (Orthoptera:
702 Acridomorpha) applying dispersal–vicariance analysis. *J. Arid Environ.* 66, 421–
703 434.
704 Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzy, E.J.P., 2003. Comparison
705 of Bayesian and maximum likelihood bootstrap measures of phylogenetic
706 reliability. *Mol. Biol. Evol.* 20, 248–254.
707 Efron, B., Tibshirani, R., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic
708 trees. *Proc. Natl. Acad. Sci. USA* 93, 13429–13434.
709 Eriksson, T., 1996. AutoDecay version 4.0. Program distributed by the author.
710 Department of Botany, Stockholm University, Stockholm, Sweden.
711 Farris, J.S., 1969. A successive approximations approach to character weighting.
712 *Syst. Zool.* 18, 374–385.
713 Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of
714 incongruence. *Cladistics* 10, 315–319.
715 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the
716 bootstrap. *Evolution* 39, 783–791.
717 Flook, P.K., Rowell, C.H.F., 1997. The effectiveness of mitochondrial rRNA gene
718 sequences for the reconstruction of the phylogeny of an insect order
719 (Orthoptera). *Mol. Phyl. Evol.* 8, 177–192.
720 Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995. The sequence, organisation and
721 evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* 41, 928–
722 941.
723 Fries, M., Chapco, W., Contreras, D., 2007. A molecular phylogenetic analysis of the
724 Oedipodinae and their intercontinental relationships. *J. Orthoptera Res.* 16,
725 115–125.
726 Gangwere, S.K., Muralirangan, M.C., Muralirangan, M., 1997. *The Bionomics*
727 *Grasshoppers, Katydid, and their Kin*. CAB International, New York.
728 Gentry, A.W., Sutcliffe, A.J., 1981. Pleistocene geography and mammal faunas. In:
729 Cocks, L.R.M. (Ed.), *The Evolving Earth*. British Museum (Natural History) and
730 Cambridge University Press, London and Cambridge, pp. 237–251.
731 Hasegawa, M., Kishino, H., 1989. Confidence limits on the maximum-likelihood
732 estimate of the hominoid tree from mitochondrial-DNA sequences. *Evolution*
733 43, 672–677.
734 Hillis, D.M., Heath, T.A., Jahn, K.S.T., 2005. Analysis and visualization of tree space.
735 *Syst. Biol.* 54, 471–482.
736 Hipp, A.L., 2004. Congruence versus phylogenetic accuracy: revisiting the
737 incongruence length difference test. *Syst. Biol.* 53, 81–89.
738 Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: bayesian inference of phylogeny.
739 *Bioinformatics* 17, 754–755.
740 Jago, N.D., 1979. The breakup of Pangaea and prediction of numbers of subfamily
741 elements, living and extinct in Acridoidea. In: *Proceedings of the 2nd Triennial*
742 *Meeting of the Pan American Acridological Society, Montana, USA*, pp. 139–162.
743 Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify
744 homologous positions: an example of alignment and data presentation from the
745 frogs. *Mol. Phylogenet. Evol.* 4, 314–330.
746 Kjer, K.M., 2004. Aligned 18S and insect phylogeny. *Syst. Biol.* 53, 506–514.
747 Litzenberger, G., Chapco, W., 2003. The North American Melanoplineae (Orthoptera:
748 Acrididae): a molecular phylogenetic study of their origins and taxonomic
749 relationships. *Ann. Entomol. Soc. Am.* 96, 491–497.
750 Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned
751 regions of DNA sequences in phylogenetic analyses without violating positional
752 homology. *Syst. Biol.* 49, 628–651.
753 Lyons-Weiler, J., Hoelzer, G.A., Tausch, R.J., 1998. Optimal outgroup analysis. *Biol. J.*
754 *Linn. Soc.* 64, 493–511.
755 Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary
756 analysis. Version 1.11. <<http://mesquiteproject.org>>.
757 Olte, D., 1995. Orthoptera species file 4. The Orthopterists' Society and the Academy
758 of Natural Sciences of Philadelphia, Philadelphia.
759 Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on
760 personal computers. *Comput. Appl. Biosci.* 12, 357–358.
761 Page, R.D.M., 2000. Comparative analysis of secondary structure of insect
762 mitochondrial small subunit ribosomal RNA using maximum weighted
763 matching. *Nucleic Acids Res.* 28, 3839–3845.

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplineae (Orthoptera: Acrididae). *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.ympev.2010.07.003

- 764 Pereira, S.L., Johnson, K.P., Clayton, D.H., Baker, A.J., 2007. Mitochondrial and nuclear
765 DNA sequences support a Cretaceous origin of Columbigrasshoppers and a dispersal-
766 driven radiation in the Paleogene. *Syst. Biol.* 56, 656–672.
- 767 Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution.
768 *Bioinformatics* 14, 817–818.
- 769 Rehn, J.A.G., Randell, R.L., 1963. A preliminary analysis of the lines of the supertribe
770 Melanoplina Orthoptera: Acrididae: Cyrtacanthacridinae. *Proc. Acad. Natl. Sci.*
771 (Phila.) 115, 1–32.
- 772 Rehn, J.A.G., Rehn, J.W.H., 1939. Studies of certain cyrtacanthacridoid genera
773 (Orthoptera: Acrididae) Part I. The Podisma complex. *Trans. Am. Entomol. Soc.*
774 65, 61–96.
- 775 Ronquist, F., 1996. DIVA version 1.1. Computer program and manual available by
776 anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se).
- 777 Ronquist, F., 1997. Dispersal-variance analysis: a new approach to the
778 quantification of historical biogeography. *Syst. Biol.* 46, 195–203.
- 779 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: bayesian phylogenetic inference
780 under mixed models. *Bioinformatics* 19, 1572–1574.
- 781 Rowell, C.H.F., Carbonell, C.S., 1977. *Baeacris talamancensis* gen. and sp. Nov.
782 (Acrididae, Melanoplinae), a Neotropical montane grasshopper; its
783 implications for the origin of the Dichroplini and the Costa Rican paramo.
784 *Acrida* 6, 55–74.
- 785 Rowell, C.H.F., Flook, P.K., 2004. A dated molecular phylogeny of the Proctolabinae
786 (Orthoptera, Acrididae), especially the Lithosiriinae, and the evolution of their
787 adaptive traits and present biogeography. *J. Orthoptera Res.* 13, 35–56.
- 788 Sanmartin, I., 2003. Dispersal vs. vicariance in the Mediterranean: historical
789 biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *J.*
790 *Biogeogr.* 30, 1883–1897.
- 791 Sanmartin, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by
792 event-based models: plant versus animal patterns. *Syst. Biol.* 53, 216–243.
- 793 Sanmartin, I., Engghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance
794 and diversification in the Holarctic. *Biol. J. Linn. Soc.* 73, 345–390.
- 795 Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE:
796 maximum likelihood phylogenetic analysis using quartets and parallel
797 computing. *Bioinformatics* 18, 502–504.
- 798 Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree
799 selection. *Syst. Biol.* 51, 492–508.
- 800 Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with
801 applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- 802 Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of
803 phylogenetic tree selection. *Bioinformatics* 17, 1246–1247.
- 804 Skes, D.S., Lewis, P.O., 2001. Beta software, version 1. PAUPRat: PAUP
805 implementation of the parsimony ratchet. Distributed by the authors.
806 Department of Ecology and Evolutionary Biology, University of Connecticut,
807 Storrs, USA.
- 808 Smon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution,
809 weighting, and phylogenetic utility of mitochondrial gene sequences and a
810 compilation of conserved PCR primers. *Ann. Entomol. Soc. Am.* 87, 651–701.
- 811 Smon, C., Buckley, T.R., Frati, F., Stewart, J., Beckenbach, A., 2006. Incorporating
812 molecular evolution into phylogenetic analysis, and a new compilation of
813 conserved polymerase chain reaction primers for animal mitochondrial DNA.
814 *Annu. Rev. Ecol. Syst.* 37, 545–579.
- 815 Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to
816 visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci.*
817 *USA* 94, 6815–6819.
- 818 Swofford, D.L., 1995. PAUP². Phylogenetic Analysis Using Parsimony (and Other
819 Methods). Sinauer Associates, Sunderland, Massachusetts.
- 820 Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The
821 ClustalX windows interface: flexible strategies for multiple sequence alignment
822 aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- 823 Thorne, J.L., Kishino, H., 2002. Divergence time estimation and rate evolution with
824 multilocus data sets. *Syst. Biol.* 51, 689–702.
- 825 Vickery, V.R., 1987. The northern Nearctic Orthoptera: their origins and survival. In:
826 Bacetti, B.C. (Ed.), *Evolutionary Biology of Orthopteroid Insects*. Ellis Harwood,
827 West Sussex, England, pp. 581–591.
- 828 Vickery, V.R., 1997. Classification of orthoptera (sensu stricto) or caelifera. In:
829 Gangwere, S.K., Muralirangan, M.C., Muralirangan, M. (Eds.), *The Bionomics of*
830 *Grasshoppers, Katydid and their Kin*. CAB International, New York, pp. 5–40.
- 831 Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for
832 phylogenetic analysis. *Annu. Rev. Entomol.* 53, 449–472.
- 833 Wróbel, B., 2008. Statistical measures of uncertainty for branches in phylogenetic
834 trees inferred from molecular sequences by using model-based methods. *J.*
835 *Appl. Genet.* 49, 49–67.

Please cite this article in press as: Chintauan-Marquier, I.C., et al. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: The Melanoplinae (Orthoptera: Acrididae). *Mol. Phylogenet. Evol.* (2010), doi:10.1016/j.ympev.2010.07.003

V.3. Histoire évolutive des Podismini

Contribution scientifique correspondant à l'article (B):

Chintauan-Marquier, I.C., Amédégnato C., Pompanon F. *en préparation*. **New Molecular Evidences for the Evolutionary History of Eurasian Podismines (Orthoptera: Acrididae: Melanoplineae).**

V.3.1. Principaux résultats

Les résultats de cette étude suggèrent que les Podismini Eurasiatiques forment un groupe polyphylétique dépourvu de signification évolutive. A l'intérieur des Podismini, les clades proposés par nos résultats, ne soutiennent pas les sous-tribus ou groupes de genres définis sur la base des critères morphologiques (Rehn et Randell 1963) ou cytologiques (Fontana et Vickery 1976). La principale interprétation biogéographique de la phylogénie est que leur distribution actuelle serait le résultat des événements géologiques du Cénozoïque, comme l'orogénèse de principales chaînes montagneuses (e.g., Himalayas, Altay, Alpes etc.) et des grands déserts (e.g., Gobi, Karakoum) dans l'Asie Centrale, et de l'ouverture des mers marginales (e.g., Bering, du Japon) dans l'extrême Est de l'Asie.

V.3.2. Discussion

Nos résultats présentent les Podismini comme un groupe polyphylétique ayant des origines multiples d'âges différents. Cette hypothèse est, d'une part en contradiction avec des recherches moléculaires antérieures (e.g. Chapco et al. 2001, Litzenberger et Chapco 2003, Chintauan-Marquier et al. 2010), qui revendiquent une origine monophylétique pour les Podismini Eurasiatique, et d'autre part, en accord avec l'hypothèse de « faunes multi stratifiées », formulée par les morphologistes pour les Melanoplineae (e.g. Fries et al. 2007) et

les Acrididae, en général (e.g. Amédégato 1977). La contradiction concernant la monophylie des Podismini est induite par la différence d'échantillonnage entre notre article précédent (Chintauan-Marquier et al. 2010) et celui-ci. Ainsi, l'augmentation de l'échantillon antérieurement analysé (voir Chintauan-Marquier et al. 2010) a divisé la branche longue qui liait les ancêtres de la faune de l'Extrême Orient, i.e. *Sinopodisma*, et de l'Europe, i.e. *Italopodisma* et *Podisma*.

Nos résultats proposent, pour la Paléarctique de l'Ouest (Europe), trois origines: (i) Néarctique au Paléocène, (ii) *in situ* datant de l'Eocène et (iii) Est Paléarctique au Miocène. Pour la Paléarctique de l'Est (Asie), les trois origines proposées sont : (i) Néarctique au Paléocène, (ii) *in situ* pendant l' Eocène, (iii) Ouest Paléarctique en Oligocène et (iv) Extrême Est Asiatique (Japon) pour le Miocène. En conséquence, la faune Eurasiatique actuelle des Podismini est le résultat de l'évolution d'une faune ancestrale tempérée et de plusieurs événements de dispersion intercontinentale et de diversification vicariante *in situ* (duplication), dans les conditions climatiques et géomorphologiques du Tertiaire. Ainsi, la distribution insulaire de certains genres endémiques pour des massifs de la chaîne Alpine, comme *Italopodisma* (Apennins), *Odontopodisma* (Alpes, Balkans, Carpates) ou *Podisma*, est un témoin de l'impact des glaciations du Pléistocène sur l'évolution des Podismini. De même, l'aire de répartition disjointe de certains genres, comme *Cophopodisma* (distribution Ouest – Est Paléarctique), pourrait être le résultat des changements induits dans une aire ancestrale continue par des processus tectoniques, d'orogénèse etc.

En conclusion, nos résultats montrent que les clades phylogénétiques à l'intérieur des Podismini ne correspondent ni aux classifications morphologique ou cytologique, ni à la distribution géographique des sous-groupes. Elles sont plutôt en accord avec l'évolution climatique, biotique et orogénique de l'Holarctique au Cénozoïque.

V.3.3. Article - Chintauan-Marquier, I.C., Amédégnato C., Pompanon F. *en préparation*. **New Molecular Evidences for the Evolutionary History of Eurasian Podismines (Orthoptera: Acrididae: Melanoplinae).**

New Molecular Evidences for the Evolutionary History of Eurasian Podismines

(Orthoptera: Acrididae: Melanoplinae)

Ioana C. Chintauan-Marquier¹, Christiane Amédégnato² and François Pompanon¹

¹ *Laboratoire d'Ecologie Alpine, Université Joseph Fourier UMR CNRS 5553, 38041 Grenoble Cedex 09, France*

² *Museum National d'Histoire Naturelle, Département Systématique et Evolution, UMR CNRS 7205 OSEB, Entomologie, Case 50, 45, Rue Buffon, 75005 Paris, France*

Running head: evolutionary history of the Podismini.

Corresponding author:

Ioana C. Chintauan-Marquier

E-mail: icchintauan.marquier@gmail.com

Abstract. — The podismines are Melanoplinae grasshoppers having a Holarctic distribution with a major representation in Eurasia. To investigate their controversial taxonomy and evolutionary history, we studied 25 species representatives of the main Eurasian genera. More than 1650 base pairs of one region of nuclear rDNA (ITS1) and three of mitochondrial DNA (CO2, 12S and 16S) were combined and used to construct parsimony, maximum likelihood and Bayesian phylogenies. The times of species divergence were estimated via a Bayesian molecular clock, and the biogeographic framework was assessed using an event-based parsimony method (DIVA). Our results suggest that the selected Eurasian Podismini are polyphyletic inside the Melanoplinae, and that this group has no evolutionary significance. Within

this group, the clades supported by the phylogenies do not fit the existing classifications. The results are in accordance with a distribution of the Eurasian Melanoplineae fauna resulting from Cenozoic geological events, such as the genesis of high mountain chains (e.g., Himalayas, Altay, Alpes) and large deserts (e.g., Gobi, Karakoum, Taklamakan) in the Central Asia, as well as by the open of marginal seas in the Eastern Asia (e.g., Bering, Japanese and Yellow Seas).

Keywords: grasshopper, Podismini, Melanoplineae, combined phylogeny, mitochondrial and nuclear DNA, molecular dating, biogeography, taxonomy.

The podismines are members of the Melanoplineae acridians, with more than 45 genera and 236 species, distributed throughout Eurasia (191 species), Indo-Malaysia (6 species) and North America (39 species) (e.g., Vickery 1987, Otte 1995). The podismines are one of the main components of the mountain grasshopper fauna in Eurasia, the altitudinal and latitudinal situation of the mountain range being a determining factor of their presence and evolution (e.g. Dreux 1962, Uvarov 1977, Voisin 1979). Although they have a reduced agricultural concern, some of them have served as model organism. For example the alpine grasshopper *Podisma pedestris* has been used for investigating pattern of chromosomal and rDNA pseudogenes evolution (Hewitt 1979, Nichols and Hewitt 1986, Keller et al. 2006).

According to different authors, the melanopline grasshoppers have been divided in two to seven tribes (e.g. Rehn and Randell 1963, Otte 1995, Vickery 1987, 1997), based on either morphological (e.g. Rehn and Randell 1963) or cytological arguments (e.g. Fontana and Vickery 1976). These classifications mainly differ in the hierarchical organization of the groups, with subtribes in one scheme (e.g. Rehn and Randell 1963) being treated as tribes in another (e.g. Otte 1995). In those different classifications, the grouping of the Eurasian melanoploids and their Nearctic relatives, as Podismini or Podisminae, is a common feature.

The most comprehensive morphological definition of this group corresponds to the tribe Podismini defined by Rehn and Randell (1963), and we will rely on this classification in the present paper. Most of phylogenetic studies show the monophyly of the Eurasian podismines with respect to North American melanoplines, but disagree on the scenarios explaining their evolutionary history. Three such scenarios have been proposed: (1) the group originating from an Asian (Angaran) center (Rehn and Rehn 1939, Vickery 1987) would have dispersed in North America via either the Bering Land Bridge or a North Atlantic connection (e.g. Tiffney 1985, Kavanaugh 1988, Askevold 1991, Sanmartin et al. 2001), (2) the splitting of the Laurasian super-continent (Vickery 1986) would have determined the division of melanoplines into vicariant groups including the European Podismini, and (3) the podismines would have evolved in the South-West of North America before a westward dispersion towards Eurasia via the Cenozoic Bering Land Bridges (e.g. Litzenberger and Chapco 2001, 2003, Chintauan-Marquier et al. 2010).

The present study aimed to clarify the evolutionary history of the Eurasian podismines, and to evaluate the concordance between the evolutionary relationships inferred from molecular data and the affinities suggested by current morpho-taxonomies (Rehn and Randell 1963, Vickery 1987) and compilations (Otte 1995, Vickery 1997). More precisely, (i) we assessed the most probable scenarios for the origin and evolution of the European and Asian fauna, and (ii) we dated the events inferred from the phylogenetical analyses. We based our study on combined nuclear and mitochondrial DNA sequences for 25 species representative of the taxonomic and ecological diversity of the Eurasian podismines.

2. Materials and Methods

2.1. Taxon sampling

DNA sequences were obtained from samples of the Caelifera (Insecta, Orthoptera) collection of the Museum National d'Histoire Naturelle (MNHN) in Paris and from samples collected for this study (Table 1). To organize our sampling, we adopted the classification recognising three main tribes of Holarctic Melanoplinae (Rehn and Randell, 1963). Most of the selected genera were represented by their “type species” (see Table 1) as designated in their original morphological description by different authors (e.g. Bei-Bienko and Mishchenko 1951, Harz 1975, Galvani 1987, Otte 1995). The 25 species covered the present Eurasian distribution and ecological diversity of the podismine grasshoppers. Two Palaeartic Catantopinae and four Nearctic Melanoplinae were included as outgroup species (Table 1).

Table 1. Origin of study material and references of sequence data. LECA: Laboratoire d'Ecologie Alpine, UMR CNRS-UJF 5553 Grenoble, France; MNHN: MNHN Caelifera collection ; TS : type species.

Species	Authority	Collecting data	Collection nos.	GenBank nos.			
				ITS1	CO2	16S	12S
Acridoidea							
Acrididae							
Catantopinae							
<i>Oxycatantops spissus</i>	Walker, 1870	Guinea, Nimba, 1984 France, Dordogne,	MNHN Gu8401	GQ354701	—	—	GQ354753
<i>Pezotettix giornae</i> TS	(Rossi, 1794)	2000	MNHN Fr0001	GQ354702	—	xxx	GQ354755
Melanoplinae							
Dactyloptini							
<i>Dactyloptum bicolor bicolor</i> TS	Charpentier, 1843	Mexico, Hidalgo, 1986	MNHN Me8604	GQ354688	—	—	GQ354740
<i>Hesperotettix viridis viridis</i>	(Thomas, 1872)	Mexico, San Luis Potosi, 1986	MNHN Me8605	GQ354692	—	—	GQ354744
Melanoplini							
<i>Melanoplus frigidus</i> TS	(Boheman, 1846)	France, Lautaret, 1999	LECA MFFL0199	GQ354697	xxx	xxx	GQ354749
<i>Phaulotettix compressus</i> TS	Scudder, 1897	Mexico, Hidalgo, 1986	MNHN Me8602	GQ354703	—	—	GQ354756
Podismini							
Miramellae							
<i>Capraiusscola (Miramella)</i> <i>ebneri ebneri</i> TS	Galvani, 1953	Romania, Sinaia, 1966	LECA MEE22021	xxx	xxx	xxx	—
<i>Chortopodisma cobellii</i> TS	Ramme, 1951	Italy , Abruzzo, 2001 France, Pyrenées,	LECA CC8011	xxx	xxx	xxx	xxx
<i>Cophopodisma pyrenaica</i> TS	(Fischer, 1853)	1997	LECA CP9021	xxx	xxx	xxx	xxx
<i>Epipodisma pedemontana</i> <i>pedemontana</i> TS	Brunner, 1882	Russia, 1997	LECA EP14011	xxx	xxx	xxx	xxx
<i>Nadigella (Miramella)</i> <i>formosanta</i> TS	(Fruhstorfer, 1921)	Italy, 1999	LECAMF22082	xxx	—	—	xxx
<i>Parapodisma (Callopodisma)</i> <i>dairisama</i>	Kano, 1996	Japan, 1998	LECA CD28011	—	xxx	xxx	—
<i>Parapodisma mikado</i> TS	(Bolivar, 1890)	Japan, Sapporo, 1997	LECA PARMIK2	xxx	xxx	xxx	xxx
<i>Parapodisma subastris</i>	Huang, 1983	Japan, Sakurai, 1998 Italy, Monte Baldo,	LECA PS28071	xxx	xxx	xxx	xxx
<i>Pseudoprunna baldensis</i> TS	Krauss, 1883	1999	LECA PB37012	xxx	xxx	—	xxx
<i>Sinopodisma punctata</i>	Mistshenko, 1954	Japan, Tatsugou, 1998	LECA SPJT0198	GQ354708	xxx	xxx	GQ354760
Podismae							
<i>Italopodisma costai</i> TS	(Targioni-Tozzetti,	Italy , Abruzzo, 2000	LECA IC0100	xxx	xxx	—	xxx

	1881)						
<i>Italopodisma fiscellana</i>	(La Greca, 1954)	Italy , Abruzzo, 2000	LECA IF19061	xxx	xxx	xxx	xxx
<i>Micropodisma salamandra</i>	(Fischer, 1853)	Austria, 1995	LECA MS21032	xxx	xxx	xxx	xxx
<i>Odontopodisma carpathica</i>	Kis, 1962	Romania, Ghinda, 2000	LECA OC24031	xxx	—	xxx	—
<i>Odontopodisma rubripes</i>	(Ramme, 1931)	Romania, Calimani, 2001	LECA OR24101	xxx	xxx	xxx	xxx
<i>Ognevia longipennis</i>	(Shiraki, 1910)	Japan, Sapporo, 1998	LECA OL25011	xxx	xxx	xxx	xxx
<i>Podisma kanoi</i>	Storozhenko, 1994	Japan, 1998	LECA PK31201	xxx	—	xxx	xxx
<i>Podisma pedestris</i> TS	(Linnaeus, 1758)	France, Belledonne, 2000	LECA PPF0100	GQ354704	xxx	xxx	GQ354757
<i>Pseudopodisma transilvanica</i>	Galvani-Fontana, 1993	Romania, 2001	LECA PT36031	xxx	—	xxx	—
<i>Zubovskya koeppeni parvula</i> TS	(Ikonnikov, 1911)	Japan, M. Ashibetsu, 1997	LECA ZKP43051	xxx	xxx	xxx	xxx
Unassigned subtribe							
<i>Anapodisma miramae</i> TS	Dovnar-Zapolskii, 1932	Russia, 2000	LECA ANMI0100	xxx	xxx	xxx	xxx
<i>Fruhstorferiola okinawaensis</i>	(Shiraki, 1930)	Japan, Okinawa, 1998	LECA FO15041	xxx	xxx	xxx	xxx
<i>Peripodisma tymphii</i> TS	Willemse, 1972	Greece, Epiros, 1999	LECA PT30011	xxx	xxx	xxx	xxx
<i>Tonkinacris</i> sp.	Carl, 1916	Japan, Ouno, 1997	LECA T4101	xxx	xxx	xxx	xxx
Primnoae							
<i>Prumna (Primnoa) primnoa</i> TS	(Motschulsky, 1859)	Mongolia, Selonge, 1997	LECA PP34191	xxx	xxx	xxx	xxx

2.2. DNA extraction, amplification and sequencing

Total DNA was extracted from femoral muscles using the Qiagen DNeasy tissue kit, according to the manufacturer's instructions. The remainder of the specimen was vouchered and maintained in the "Laboratoire d'Ecologie Alpine" from Grenoble.

We sequenced a total of about 455 base pairs (bp) of one region of nuclear rDNA (ITS1) and about 1206 bp of three regions of mitochondrial DNA (CO2, 12S and 16S). We used the same primers for both DNA amplification and sequencing (Table 2). The PCR were performed in a final volume of 25 µl containing 2-5 µl of DNA, and 20-23 µl buffer solution with 10 µM of each primer, 10 mM of each dNTP, 25 mM MgCl₂, and 5u/µl of AmpliTaq Gold polymerase (Applied Biosystems, Foster City, California). PCR were performed according to the following protocol: initial polymerase activation at 95°C for 10 min; then for 35-45 cycles: denaturation at 95°C for 30 sec, annealing (temperatures in Table 2) for 60 sec, elongation at 72°C for 2 min, and final elongation step at 72°C for 10 min. PCR products corresponding to a single fragment were purified using the Qiagen QIAquick PCR

purification kit. Otherwise, we used a combination of gel separation and subsequent purification with the Qiagen QIAEX II gel extraction kit. To increase the quantity of target DNA for most degraded samples, PCR products were ligated into a pGEM-T Easy vector (Promega) overnight. Then 2 μ l of the ligation reaction were used to transform JM109 High Efficiency Competent Cells (Promega) according to the manufacturer's recommendations. Positive clones were then used as a template for a second PCR amplification.

Table 2. Descriptive statistics for the unambiguous aligned sequence datasets for the ingroup and outgroup combined. Statistics were estimated after removal of primer sequences and the best model of evolution selected with the AIC criterion in Modeltest 3.7 (Posada and Crandall 1998). df = degree of freedom.

Loci	No. taxa	No. characters	no/% variable sites	no/% parsimony informative sites	Model of evolution	AIC	(-) lnL	Average frequencies				Constancy of base frequencies (χ^2 and P values)	
								A	C	G	T	all sites	variable sites
ITS1	30	455	121/27	53/12	HKY+G	3136.69	1563.35	0.1792	0.3370	0.3214	0.1625	13;P=1;df=90	26;P=1;df=90
CO2	22	379	175/46	92/24	TIM+G	4624.38	2305.19	0.3693	0.1631	0.1472	0.3202	19;P=1;df=90	43;P=1;df=90
16S	24	472	128/27	63/13	TVM+I+G	3959.43	1970.71	0.3247	0.1235	0.1806	0.3711	6;P=1;df=90	25;P=1;df=90
12S	27	355	110/31	59/17	TVM+I+G	3099.12	1540.56	0.4045	0.1657	0.1030	0.3268	16;P=1;df=90	51;P=1;df=90
Combined	31	1661	534/32	267/16	GTR+I+G	15957.36	8215.37	0.3096	0.2020	0.1932	0.2952	220,8;P=0;df=90	108,5;P=0,1;df=90

Purified PCR products were used as templates in 20 μ l BigDye Terminator Cycle Sequencing kit version 2 (Applied Biosystems) reactions, with and 25 cycles and the following steps 96°C for 30s, the annealing temperature for 30s, and 60°C for 4 min. Excess dye terminators were removed by sephacryl/sephadex column purification. Sequencing reactions were electrophoresed on an ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems). Sequencing reactions were carried out on both DNA strands and SeqScape 2.5 (Applied Biosystems) was used to reconcile chromatograms of complementary strands. Autapomorphies were verified on the original chromatograms. Ambiguous results were checked by sequencing either a different DNA extract from the same individual, or an extract from another conspecific individual. However, the quality of museum preserved specimen

varied considerably according to the samples, and the degradation of DNA did not allow the amplification of target sequences for all species (see Table 1).

2.3. Sequences analysis

2.3.1. Alignment and gap treatment

We generated 134 sequences that were edited in Sequence Navigator 1.01 (PE Applied Biosystems, Inc.). Multiple alignments were generated for each locus using ClustalX 1.81 (Thompson et al. 1997) with default options, and then reviewed and corrected by eye. Both the primary sequence and the secondary structure of rDNA were taken into account when aligning the sequences. Positional homology of rDNA sequence alignments was constrained manually using the structural method described in Kjer (1995), and information about the secondary structure of rDNA on insects (e.g. Kjer 2004, Page 2000). Ambiguous aligned regions were identified by following the method proposed by Lutzoni et al. (2000). The alignments are available as Supplementary files.

2.3.2. Characteristics of molecular data

Base composition, substitution rates, and rate heterogeneity across sites were estimated under the most likely model of evolution suggested by the Akaike Information Criterion (AIC; Akaike 1974) in Modeltest 3.06 (Posada and Crandall 1998) for each dataset (Table 2). The constancy of base frequencies across taxa was evaluated with a chi-square (χ^2) “goodness-of-fit” test using PAUP*. Tests were performed with all sites and variable sites only (Table 2) in order to assess the potential confounding effects of invariant sites (e.g. Buckley et al. 2001).

2.4. *Phylogenetic analysis*

We conducted maximum parsimony (MP), maximum likelihood (ML) and Bayesian analyses on separate and combined data sets to address questions regarding the levels of homoplasy, incongruence, and information content within and among loci. Independent analyses were conducted on the ingroup alone to determine the effect of distant outgroups (e.g. Lyons-Weiler et al. 1998, Collin 2003) not only on the rooting of the ingroup but also on the recovered topology within the ingroup. The trees were edited using TREVIEW 1.3 (Page 1996).

2.4.1. *Outgroup selection*

In order to avoid the possible effects of a distant outgroup, preliminary analyses were performed to choose outgroups among the Acrididae close relatives of podismines as suggested by previous studies (e.g. Amédégnato 1977, Rehn and Randell 1963, Chintauan-Marquier et al. 2010), following the procedure detailed in Chintauan-Marquier et al. (2010).

2.4.2. *Maximum parsimony*

MP analyses were carried out on the ingroups alone and on the ingroups + outgroups. Heuristic searches, on separate and combined data, were implemented in PAUP* under the MP criteria with accelerated transformation (ACCTRAN) option to optimize the state of unordered (Fitch) characters, 100 random sequence addition replicates, the TBR branch-swapping and gaps treated as missing. Strict consensus trees were then constructed. When the equally parsimonious trees had very different topologies and their strict consensus tree had a low resolution, we opted for the equally parsimonious tree with the highest negative log likelihood (-lnL) using the Shimodaira-Hasegawa test (WSH; Shimodaira and Hasegawa 1999) implemented in CONSEL 0.1i (Shimodaira and Hasegawa 2001) with ten sets of

100,000 bootstrap replicates. To explore tree islands in a multidimensional scaling, based on tree-to-tree distance metrics, we used the Tree Set Visualization 2.1 module (Amenta and Klingner 2002) in the Mesquite package (Maddison and Maddison 2006) as suggested by Hillis et al. (2005).

2.4.3. Maximum likelihood

ML searches were conducted with PAUP*. We first evaluated the models of DNA substitution that best fit each data partition (Table 2) following the AIC implemented in Modeltest 3.06 (Posada and Crandall 1998) with the default settings. In all ML searches, we estimated appropriate ML model parameters using the data partition of interest and a neighbour-joining (NJ) topology. These estimated values were fixed, and a full heuristic ML search was conducted with 10 random addition sequence replicates, retaining all minimal trees, and TBR branch swapping. Then we re-estimated model parameters on the ML tree and used these new values for a new search. We repeated this process until the new search found the same topology as the previous search and all parameter estimates were identical.

2.4.4. Bayesian analyses

Bayesian Markov chain Monte Carlo (MCMC) analyses were performed with MrBayes 3.0b4 (Ronquist and Huelsenbeck 2003). We performed two simultaneous independent Metropolis-coupled MCMC runs starting from different random points. We set each run to have one cold and five heated chains to allow better mixing of the MCMC chain and minimize the chance of being trapped in local optima (Pereira et al. 2007). We performed runs assuming the same topology to be shared among all partitions, and each partition allowed to have its own model of DNA evolution according to the AIC. MCMC samples were taken in

every 500th cycle. We plotted the log-likelihood of sampled topologies to determine the burn-in period in which the MCMC had reached a stationary status (Huelsenbeck et al. 2001). Post-burn-in samples from the independent runs were used to construct a 50% majority rule consensus trees. The proportion of trees in which nodes were recovered after the burn-in period was interpreted as the posterior probability (PP) of that node (Ronquist and Huelsenbeck 2003).

2.4.5. Node support

Support for internal branches was assessed in PAUP* by nonparametric bootstrapping (nPB) (Felsenstein 1985, Efron et al. 1996) with 1,000 - 2,000 pseudoreplicates for MP and 100 – 500 pseudoreplicates for ML, using full heuristic searches with ten - twenty random addition sequence replicates, TBR branch swapping and one tree held at each step during stepwise addition (e.g. DeBry and Olmstead 2000). Posterior probabilities (PP) for individual nodes were estimated in a Bayesian MCMC framework as presented above. For ease of visual comparison to nPB values, we present these probabilities as numbers ranging from 0 to 100.

2.5. Biogeographic framework

The history of the distribution of Podismini was reconstructed applying an event-based parsimony method using DIVA 1.1 (Ronquist 1996). An exact search according to the dispersal-vicariance optimization proposed by Ronquist (1997) was applied to the data with outgroups included in order to restrict the ancestral distribution at the basal node (Sanmartin 2003, Donato 2006). DIVA 1.1 assumes that speciation is caused by vicariance and only imposes a cost for a dispersal or extinction event. The areas considered by us in DIVA for the Holarctical region correspond to the historically persistent landmass according to

palaeogeographic reconstructions (Cox 1974). They are: (i) the Western Palaearctic (WP or E) as Europe, North Africa and Asia west of the former Turgai Sea, (ii) the Eastern Palaearctic (EP or A) as non-tropical Asia east of the Turgai Sea, and (iii) the Nearctic as North-America, including the temperate Mexican distributions (N) (Fig. 3). The present geographical distributions of the selected genera of Podismini (Table 1) are available as Supplementary file (Table S3).

2.6. Temporal diversification

A Likelihood Ratio Test (LRT) showed that the evolution of the combined data partition did not conform to the molecular clock hypothesis ($P > 0.05$ for 29 degrees of freedom). Therefore, we estimated the age of nodes and the 95% credibility intervals from our combined DNA sequences data using the Bayesian molecular clock method implemented in MultiDivtime (Thorne and Kishino 2002). The assumption of constant rates over time is not required for this method, and multiple constraints on node ages are allowed.

We considered as time constraints (Table 3, Fig. 2): the age of the Catantopinae from Europe as suggested by Zeuner (1941) based on a fossil from the Upper Miocene (node 27: 11.6-5.3 MYA), and the split between the Nearctic and Palaearctic faunas during the Early Palaeocene-Late Oligocene (nodes 26 and/or 20: 64.21-27.30 MYA) (Chapco et al. 2001, Amédégnato et al. 2003, Chintauan-Marquier et al. 2010). The ages of the geological periods and the chronostratigraphic references were taken from the 1999 Geological Time Scale of the Geological Society of America (www.geosociety.org/science/timescale/timescl.pdf). All analyses were duplicated to confirm successful convergence for the estimation of divergence ages.

3. Results

3.1. Sequence analyses

A high content of G/C (66%) and A/T (69-73%), for the nuclear (455 sites) and mitochondrial (1206 sites) DNA respectively, was noted for each taxon (Table 2), in agreement with values for other melanoplinae (Chapco et al. 1999, Litzberger and Chapco 2001) and insects in general (e.g. Simon et al. 1994, 2006). Base frequencies were constant for all partitions taking all sites except combined data, and for all partitions taking variable sites only. Across all fragments, 534 sites (32.15%) were variable with 267 sites (16.07%) parsimony-informative; the latter were distributed principally (80.15%) between the mitochondrial loci. Models of evolution and associated parameters for the ML analysis are given in Table 2.

3.2. Phylogenetic analyses

For all phylogenetic methods, combining the nuclear and mitochondrial data improved tree resolution and nodal support. Adding the Acrididae outgroups (see Table 1) did not modify the basic ingroup topologies for the Eurasian podisminae. Moreover, varying factors such as the presence of autapomorphies or of invariant sites had no significant effect on tree topologies. The MP search yielded 32 equally parsimonious trees; length, consistency and retention indexes were 1061, 0.6249 and 0.5279, respectively. Similar topologies were obtained with the ML and Bayesian analyses. A weighted Shimodaira-Hasegawa test, using the combined data set and the corresponding model selected by the AIC, showed that the best *a posteriori* topology was the ML one (Fig. 1). The result of this test is available as Supplementary Material (Table S2).

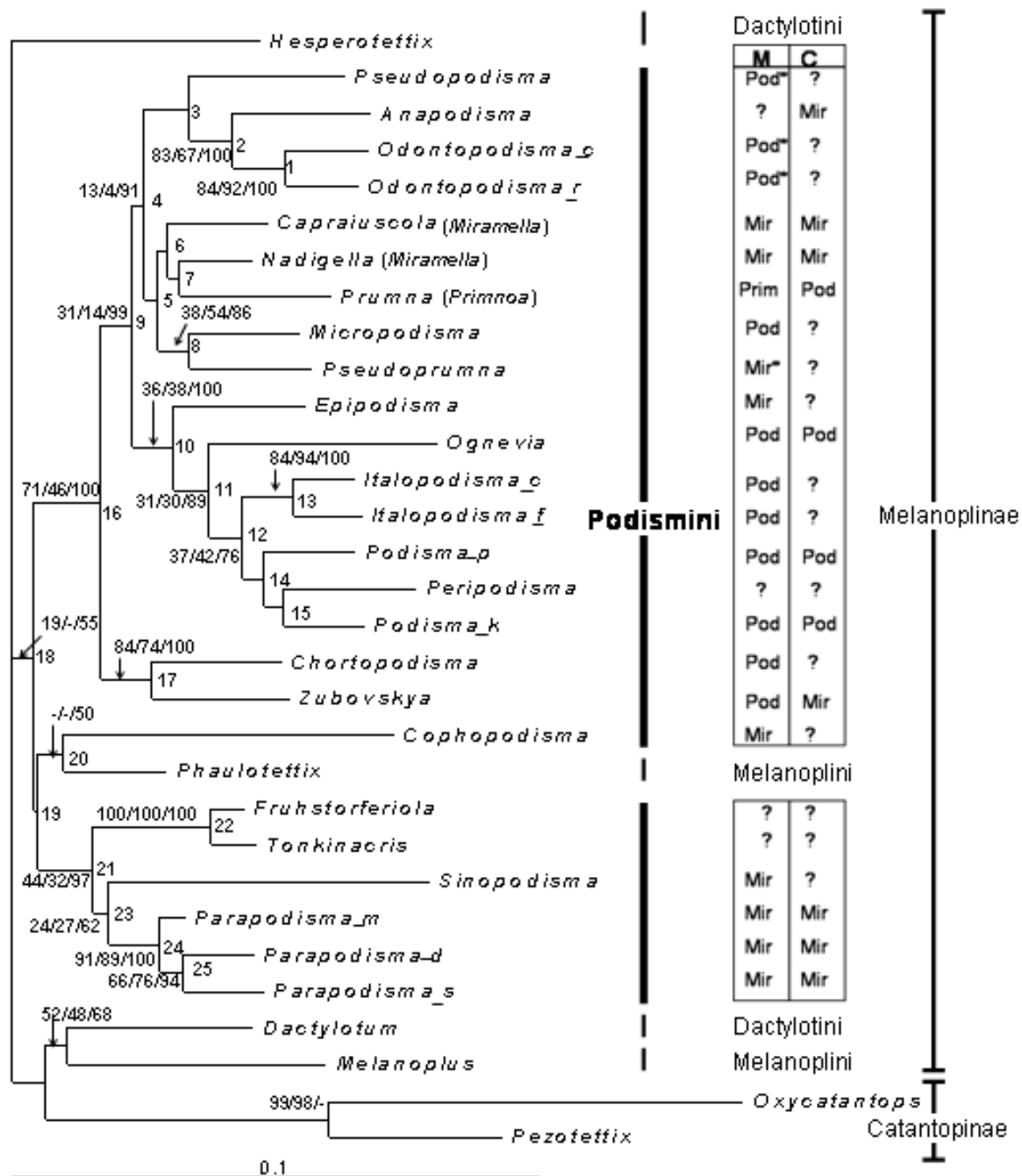


Figure 1. Maximum likelihood (ML) tree for combined data. Species names are abbreviated to a single letter; full names are included in Table 1. Summary of subtribal placements in Miramellae (Mir: 2n=21) and Podismae (Pod; 2n=23) as based on morphological (Rehn and

Randell 1963) and cytological (Fontana and Vickery 1976) data. Numbers indicate nodes.

Signes: * doubtful subtribal affiliation.

The Bayesian posterior probabilities for the nodes were generally higher than the bootstrap values. About 74% of the nodes connecting the podismines were supported by posterior probabilities higher than 50% (Fig. 1). The Bayesian analysis gives strong supports for middle level (e.g., nodes 4, 9, 12, 21) nodes that are not supported in the ML analysis. We tested one possible source of these discrepancies, i.e. model misspecification (c.f., Buckley et al. 2001, Wróbel 2008), by using different models of evolution in the ML analyses (data not shown). This did not reconcile the differences between the two methods in supporting the same topology. Thus, we considered the posterior probabilities and bootstrap values as potential upper and lower bounds of node support, as suggested by Douady et al. (2003).

Our results (Fig. 1) do not support the monophyly of the Podismini (node 18), as defined by Rehn and Randell (1963), due to the European genus *Cophopodisma* which seems to be closer to the American genus of Melanoplini, *Phaulottetix* (node 20). They suggest the existence of a basal Asian clade of Podismini (node 21), divided into an Eastern (nodes 23-25) and an Eastern - South-Eastern (node 22) branches respectively. The European Podismini *Cophopodisma* and the American Melanoplini *Phaulotettix* appear as sister taxa (node 19) of the Asian clade (node 21). However, the support for this branching of the tree is low hence unequivocal conclusions cannot be drawn.

Our results (Fig. 1) also propose the existence of a second basal clade inside the Podismini (node 16), dominated by endemic genera from the mountains of Southern and Middle-Eastern Europe (i.e., Alps, Carpathians). In its turn, this clade splits in two sub-clades,

one (node 9) clustered around the genus *Podisma*, and the other (node 17) around the genus *Zubovskya*.

Furthermore, the Asian node 21 joins only members of the *Miramellae*, while the Eurasian nodes 16, 17 and 9 mix members of both *Miramellae* and *Podismae* groups (Rehn and Randell 1963). Thus, our results reject the monophyly of these two subtribes, defined on either morphological (Rehn and Randell 1963) or cytological (Fontana and Vickery 1976) bases.

3.3 *Biogeographic analysis*

The unambiguous DIVA optimal area reconstructions, with the maximal number of ancestral areas set to four, are summarized in Fig. 3. DIVA suggested four vicariance events (allopatric speciations, nodes 2, 7, 11 and 20) and twelve duplication events that equally affect middle and terminal branches (sympatric speciations, nodes 1, 4, 5, 9, 12, 13, 16, 21-25). At least seven nodes (3, 6, 8, 10, 14, 15, and 17) correspond to possible dispersals, and eight extinction events would have occurred (nodes 2, 7, 11, 20).

3.4 *Temporal diversification*

The results of the molecular dating, using the age of a fossil Catantopinae and the splitting between the Eurasian and North American melanoplines as calibration points on the Bayesian topology (Fig. 2), with all outgroups included, are presented in the Table 3. The present fauna is likely to have evolved from old lineages, including North American Melanoplineae (node 26, 18, 19, 20), since the Early Palaeocene to the Late Pleistocene (60.64 – 0.59 MYA). Melanoploid lineages within the zoogeographical regions of the Palaearctic have been diverging from at least 16 MYA (e.g., nodes 13 and 22).

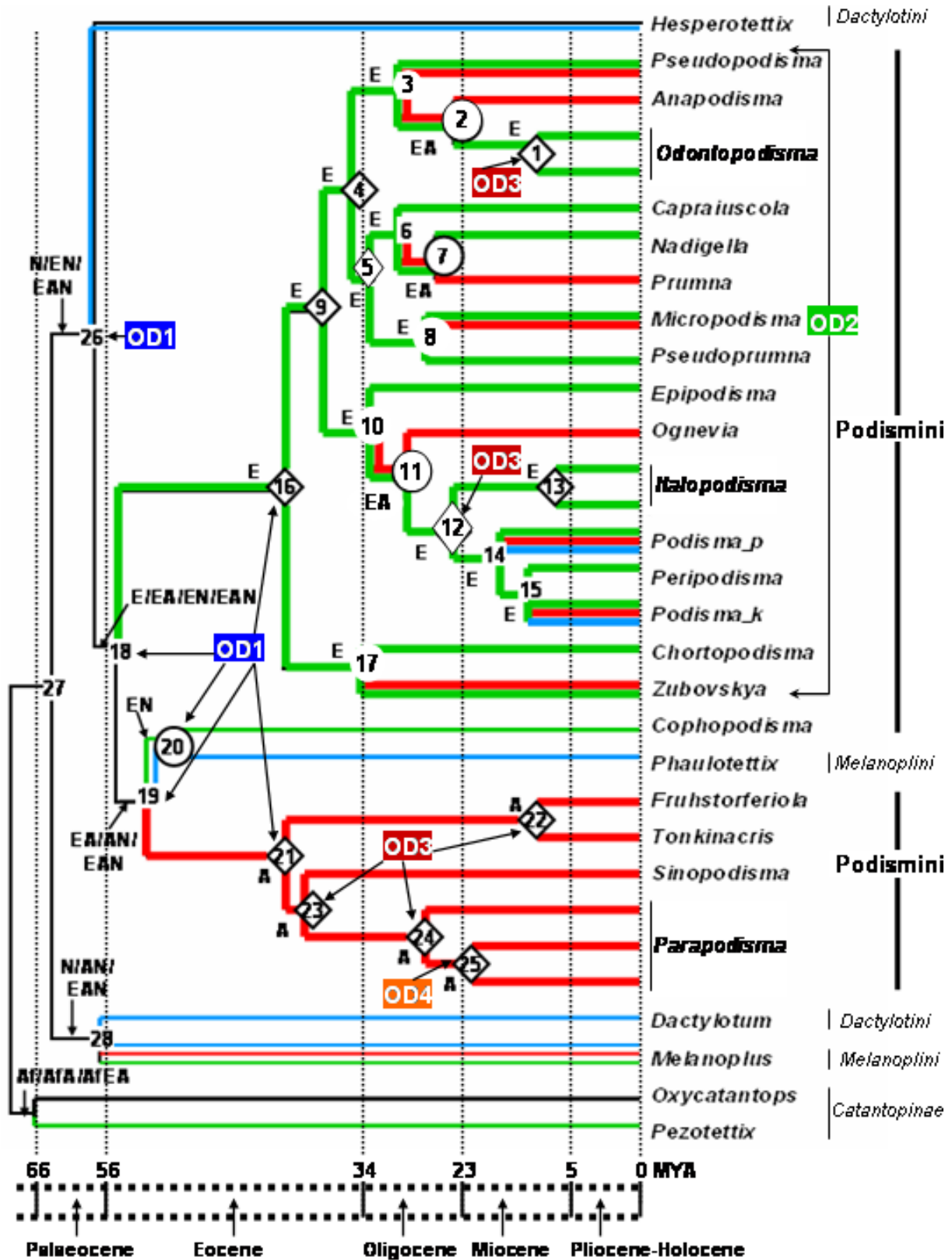


Figure 2. Chronogram of Podismini and reconstruction of ancestral areas as suggested by DIVA. Africa (Af), Europe (E), Asia (A) and North America (N). Branch colours indicate the biogeographical distribution of the taxa and their hypothetical ancestors: green – Europe, red

– Asia, blue – North America. Numbers indicate nodes. Symbols: *circle*: vicariance; *rhomb*: duplication (sympatric speciation).

4. Discussion

Three scenarios have previously been proposed to explaining the origin and diversification of the podismines. They are: (i) an ancient Laurasian origin about 200 million years before present, or at least before the separation of the continents of North and South America (Vickery 1987), (ii) an ancient Angarian origin followed by eastward dispersions in North America, in the Pleistocene or even earlier (Rehn 1958), and (iii) a Laurentian origin followed by several westward migrations to the Palaeartic (Rehn and Randell 1963, Chintauan-Marquier et al. 2010) *via* the Bering Land Bridges (e.g. Carbonell 1977, Litzenberger and Chapco 2001, 2003) about less than 65 MYA.

Our results clearly reject the first scenario and suggest that the selected Eurasian Podismini constitute a polyphyletic group inside the Melanoplineae, with no evolutionary significance. Accordingly, these Palaeartic podismines would have multiple origins of different ages for, as already advanced for the Melanoplineae (e.g. Fries et al. 2007) and Acrididae, in general (e.g. Amédégnato 1977). Thus, for the Western Palaeartic (WP) fauna of podismines, our results suggest (Fig. 3): (i) a Palaeocene (60.64 – 3.27 MYA) Nearctic origin (OD1), (ii) a Mid Eocene (44.80 – 0.65 MYA) *in situ* origin (OD2) and (iii) an Early Miocene (23.46 – 0.62 MYA) East Palaeartic origin (OD3). For the Eastern Palaeartic (EP) podismines, our results propose: (i) a Palaeocene (60.64 – 3.27 MYA) Nearctic origin (OD1), (ii) a Mid Eocene (41.65 – 0.59 MYA) *in situ* origin (OD3), (iii) an Early - Late Oligocene (28.73 – 0.8 MYA) West Palaeartic origin (OD2), and (iv) an Early Miocene (21.33 – 0.86 MYA) Japanese origin (OD4).

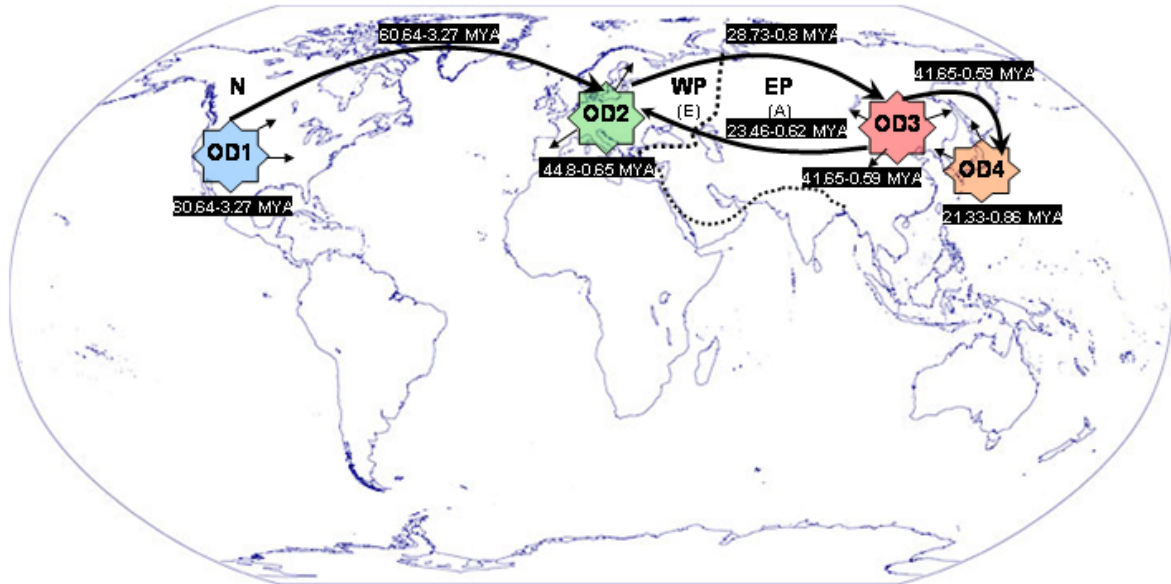


Figure 3. Biogeographic scenario for the evolution of the Palearctic Podismini. The diagram represents the main migration and diversification events inferred from the combined rDNA phylogenies and an event-based parsimony method using DIVA 1.1. Datations result from the analysis of combined DNA sequences using the Bayesian molecular clock method implemented in MultiDivtime. Nearctic (N), Western Palaeartic (WP) or Europe (E) and Eastern Palaeartic (EP) or Asia (A).

Thus more precisely, the ancestors of podismines would correspond to a cold-temperate melanoplinea fauna (node 26: 60.64-4.77 MYA Fig. 2), which would evolved, in the Early Tertiary, in the Western North-America (OD1 Fig. 3) (Rehn 1958, Vickery 1987, Litzenger and Chapco 2001, 2003). These ancestors would have dispersed either to Europe (OD2), *via* the Trans-Atlantic Land Bridges (e.g. Tiffney 1985, Sanmartin et al. 2001) (nodes 26→18→16/19→20 levels) or to Asia (OD3), *via* the Trans-Beringian Land Bridges (e.g. Rehn and Randell 1963, Norlander et al. 1996, Sanmartin et al. 2001) (nodes 19 → 21 levels). Our results (Fig. 2) also suggest biotic exchanges between Europe and Asia (nodes 2, 7, 8, 11, 14, 15) and the reciprocal way (nodes 1, 12), after the closing of the Turgai Strait less than 30

MYA. This is consistent with Litzemberger and Chapco's (2001), but disagree with LaGreca (1977), who proposed a podismines' colonization of Europe from Siberia during the Pleistocene (<2 Myr).

The mixed-mesophytic forests and the steppe vegetation, which would become dominant in the cold and dry climate of the Eocene – Miocene (Sanmartin et al. 2001), and would have lasted until late Pleistocene in the Palaearctic, would have privileged not only the aforementioned intercontinental dispersal events (Figs. 2-3), but also the diversification (duplication) *in situ* of the cold-temperate ancestors in Europe (OD2) (nodes 2-11, 13-15, 17) and Asia (OD3) (nodes 22-24). Allopatric taxa such as *Cophopodisma* (west-east Palaearctic), *Micropodisma* (Alps-Caucasian) and *Parapodisma* (continental Asia – Japanese islands) would result from such vicariant evolution. In the Pleistocene, the distribution areas of the corresponding podismines expanded and coalesced, and got narrower during the interglacial periods. In these periods, the Podismini fauna would have been restricted to a series of refuges located mainly at high altitudes and/or Northern latitudes, what would have resulted in the present insular distribution of taxa endemics for the Alpidic mountain ranges, such as *Italopodisma* (Apennines), *Odontopodisma* (Alps, Balkans, Carpathians), *Peripodisma* (Pindus Mountains) or *Podisma* (Alps, Carpathians, Pyrénées) (e.g. La Greca 1977, Kis 1978, Otte 1995, Sergeev 1998). Actually, our results (Fig. 2) support the existence of two lines of evolution inside the Eurasian Podismini (Fig. 2), “*Podisma*” (node 9) and “*Zubovskya*” (node 17), as suggested by Rehn (1958) based on morphology. These would result from a Mid Eocene European origin (node 16: 44.8-3.14 MYA) followed by vicariant duplication events (e.g. nodes 9, 4, 5, 1, 12, 13) in the aforementioned mountain ranges of the Alpidic belt. The dispersal events would have lasted until the late Pleistocene in Asia and Europe (OD2 – OD3 Fig. 3). Together with the vicariant diversification events *in situ*, these processes would have been influenced by the Eocene – Mid Miocene genesis of high mountain chains (e.g. Himalayas, Altay) and large deserts (e.g. Gobi, Karakoum, Taklamakan) in the Central Asia

(e.g. nodes 1, 7, 11 Fig. 2), as well as by the opening of marginal seas (e.g. Japanese Sea) in the Eastern Asia. Thus, the disjunct geographical distribution of some genera, like *Cophopodisma* (disjunction West – East Palaearctic), would be the result of changes imposed on an originally Palaearctic continuous distribution (e.g. by changes in the distribution of land, by continental or tectonic shifts, mountain-building, etc.).

The present study shows that the clades supported by the phylogenies are not in accordance neither with the existing morphological or cytological classifications of the podismine grasshoppers, nor with their present geographic distribution. However, the scenarios about the origin and evolution of these insects are in agreement with the historical biogeographic and geologic events of the Cenozoic in Holarctic (e.g. Tiffney 1985, Enghoff 1995, Sanmartin et al. 2001).

References

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.
- Amédégnato, C. 1977. Etude des Acridoidea centre et sud americains (Catantopinae sensu lato). Anatomie des génitalia, classification, répartition, phylogénie. Unpublished Thesis, University of Paris VI, 385 pp.
- Amédégnato, C., W. Chapco, and G. Litzemberger. 2003. Out of South America? Additional evidence for a southern origin of melanopline grasshoppers. *Mol. Phylogenet. Evol.* 29: 115-119.
- Amenta, N., and J. Klingner. 2002. Case study: Visualising sets of evolutionary trees. 8th IEEE Symposium on Information Visualization 2002: 71-74.

- Askevold, I.S. 1991. Classification, reconstructed phylogeny, and geographic history of the New World members of *Plateumaris* Thomson, 1859 (Coleoptera, Chrysomelidae: Donaciinae). *Memoirs of the Entomol. Soc. of Canada* 157: 5-175.
- Bei-Bienko & Mishchenko. 1951. *Locusts and Grasshoppers of the U.S.S.R. and Adjacent Countries* 1.
- Buckley, T. R., C. Simon, and G. K. Chambers. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Syst. Biol.* 50: 67-86.
- Carbonell, C. S. 1977. Origin, evolution, and distribution of the Neotropical acridomorph fauna (Orthoptera): A preliminary hypothesis. *Rev. Soc. Entomol. Argentina* 36: 153-175.
- Chapco, W., W. R. Kuperus and G. Litzenberger. 1999. Molecular phylogeny of Melanopline grasshoppers (Orthoptera: Acrididae): the genus *Melanoplus*. *Ann. Entomol. Soc. Am.* 92: 617-623.
- Chapco, W., G. Litzenberger, and W. R. Kuperus. 2001. A molecular biogeographic analysis of the relationships between North American melanopline grasshoppers and their Eurasian and South American relatives. *Mol. Phylogenet. Evol.* 18: 460-466.
- Chintauan-Marquier, I.C., Jordan, S., Berthier, B., Amédégnato, C., Pompanon, F. 2010. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: the Melanoplineae (Orthoptera: Acrididae). *Molecular Phylogenetics and Evolution: in press.*

- Collin, R. 2003. Phylogenetic relationships among Calyptraeid gastropods and their implications for the biogeography of marine speciation. *Syst. Biol.* 52: 618-640.
- Cox, C.B. 1974. Vertebrate palaeodistributional patterns and continental drift. *Journal of Biogeography* 1: 75-94.
- DeBry, R. W. and R. G. Olmstead. 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Syst. Biol.* 49: 171-179.
- Donato, M. 2006. Historical biogeography of the family Tristiridae (Orthoptera: Acridomorpha) applying dispersal-vicariance analysis. *Journal of arid environments* 66: 421-434.
- Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. P. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20: 248-254.
- Dreux P. 1962. Recherches écologiques et biogéographiques sur les Orthoptères des Alpes françaises. *Annales des Sciences Naturelles et de Zoologie* 3: 323-766.
- Efron, B., E. Halorran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93: 13429-13434.
- Enghoff, H. 1995. Historical biogeography of the Holarctic : area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics* 11: 223-263.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

- Fontana, P.G., V.R. Vickery. 1976. Cytotaxonomic studies on the genus *Boonacris*. I. The “eastern” taxa and a comparison with the related genera *Dendrotettix* and *Appalachia* (Orthoptera: Catantopidae: Podismini). *Can. J. Genet. Cytol.* 18: 625-652.
- Fries, M., W. Chapco, and D. Contreras. 2007. A molecular phylogenetic analysis of the Oedipodinae and their intercontinental relationships. *Journal of Orthoptera Research* 16: 115-125.
- Galvagni. 1987. In Baccetti [Ed.]. *Evolutionary Biology of Orthopteroid Insects*. Ellis Horwood series in entomology and acarology: 212, 217.
- Harz, K. 1975. Die Orthopteren Europas II. *Series Entomologica* 11:1-939.
- Hewitt, G. M. 1979. *Animal cytogenetics* 3, Insecta 1, Orthoptera. Gebrüder Borntraeger, Berlin.
- Hillis, D. M., T. A. Heath, and K. ST. John. 2005. Analysis and visualization of tree space. *Syst. Biol.* 54: 471-482.
- Huelsenbeck, J. P., F. Ronquist and B. Hall. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755.
- Kavanaugh, D.H. 1988. The insect fauna of the Pacific Northwest coast of North America: present patterns and affinities and their origins. *Memoirs of the Entomological Societu of Canada* 144: 125-149.
- Keller, I., I. C. Chintauan-Marquier, P. Veltsos and R. A. Nichols. 2006. Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics* 174: 863-874.

- Kis, B. 1978. Cheie pentru determinarea ortopterelor din România. Partea II. Subordinul Caelifera. Studii si comunicari. Muzeum Brukenthal. St. Nat. 22: 233-276.
- Kjer, K.M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogenet. and Evol.4: 314-330
- Kjer, K.M. 2004. Aligned 18S and Insect phylogeny. Systematic Biology 53:506-514.
- LaGreca, M. 1977. L'evoluzione Plio-Pleistocenica degli insetti di alta montagna. Boll. Zool. 44: 261-285.
- Litzenberger, G. and W. Chapco. 2001. Molecular phylogeny of selected Eurasian podismine grasshoppers (Orthoptera: Acrididae). Ann. Entomol. Soc. Am. 94: 505 -511.
- Litzenberger, G., and W. Chapco. 2003. The North American Melanoplineae (Orthoptera: Acrididae): a molecular phylogenetic study of their origins and taxonomic relationships. Ann. Entomol. Soc. Am. 96: 491-497.
- Lyons-Weiler, J., G. A. Hoelzer, and R. J. Tausch. 1998. Optimal outgroup analysis. Biol. J. Linn. Soc. 64: 493-511.
- Maddison, W. P. and D. R. Maddison. 2006. Mesquite: a modular system for evolutionary analysis. Version 1.11, <http://mesquiteproject.org>.
- Nichols, R.A., Hewitt, G.M. 1986. Population structure and the shape of a chromosomal cline between two races of *Podisma pedestris* (Orthoptera: Acrididae). Biol. J. Linn. Soc. 29: 301-316.
- Otte, D. 1995. Orthoptera species file 4. The Orthopterists' Society and the Academy of Natural Sciences of Philadelphia, Philadelphia, PA.

- Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.
- Page, R. D. M. 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. *Nucleic Acids Research* 28: 3839-3845.
- Pereira, S. L., K. P. Johnson, D. H. Clayton, and A. J. Baker. 2007. Mitochondrial and nuclear DNA sequences support a Cretaceous origin of Columbiformes and a dispersal-driven radiation in the Paleogene. *Syst. Biol.* 56: 656-672.
- Posada D and Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Rehn, J.A.G., Rehn, J.W.H. 1939. Studies of certain cyrtacanthacridoid genera (Orthoptera: Acrididae). Part 1. The Podisma complex. *Trans. Amer. Ent. Soc.* 65: 61-69.
- Rehn, J. A. G. 1958. The origin and affinities of the Dermaptera and Orthoptera of western North America. In Hubbs, C. L. *Zoogeography* 12. Horn-Shafer Co., Baltimore.
- Rehn, J. A. G., and R. L. Randell. 1963. A preliminary analysis of the lines of the supertribe Melanoplina Orthoptera; Acrididae; Cyrtacanthacridinae). *Proc. Acad. Natl. Sci. (Phila.)* 115: 1-32.
- Ronquist, F. 1996. DIVA version 1.1. Computer program and manual available by anonymous FTP from Uppsala University (<ftp.uu.se> or <ftp.systbot.uu.se>).
- Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46: 195-203.

- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-4.
- Sanmartin, I., H. Enghoff, and F. Ronquist. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biol.J.Linn.Soc.* 73: 345-390.
- Sanmartin, I. 2003 Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palaeoartic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography* 30: 1883-1897.
- Sergeev, M. 1998. Conservation of orthopteran biological diversity relative to landscape change in temperate Eurasia. *Journal of Insect Conservation* 2: 247-252.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16: 1114-1116.
- Shimodaira, H., and M. Hasegawa. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246-1247.
- Simon, C., Frati, F., Beckenback, A., Crespi, B., Liu, H., and Flook, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Annals Entomol. Soc. Am.* 87 : 651-701.
- Simon, C., Buckley, T.R., Frati, F., Stewart, J. and Beckenbach, A. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 37: 545-579.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24: 4876-4882.

- Thorne, J. L. and H. Kishino. 2002. Divergence time estimation and rate evolution with multilocus data sets. *Syst. Biol.* 51: 689-702.
- Tiffney, B. H. 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* 66: 243-273.
- Uvarov, B. 1977. Grasshoppers and locusts. A handbook of general acridology. Vol. 2. Centre for the Overseas Pest Research, London.
- Vickery, V. R. 1987. The northern Nearctic Orthoptera: their origins and survival. In: Bacetti, B. C. (Ed.), *Evolutionary Biology of Orthopteroid Insects*. Ellis Horwood, West Sussex, England: 581-591.
- Vickery, V. R. 1997. Classification of Orthoptera (sensu stricto) or Caelifera, pp. 5-40. In S. K. Gangwere, M. C. Muralirangan, and M. Muralirangan [eds.], *The bionomics of grasshoppers, katydids and their kin*. CAB International, New York.
- Voisin, J.F. 1979. *Autoécologie et biogéographie des Orthoptères du Massif Central*. Thèse de Doctorat d'état, Université Pierre et Marie Curie (Paris 6).
- Wróbel, B. 2008. Statistical measures of uncertainty for branches in phylogenetic trees inferred from molecular sequences by using model-based methods. *J. Appl. Genet.* 49: 49-67.
- Zeuner. 1941. The classification of the Dectinae hitherto included in *Platycleis* Fieb. or *Metrioptera* Wesm. (Orthoptera, Saltatoria). *Transactions of the Entomological Society of London* 91: 1-50.

Chapitre VI. – ADN ribosomal chez le criquet alpin *Podisma pedestris* (Linnaeus, 1758): pseudogènes et évolution (article C).



Chapitre VI. – ADN ribosomal chez le criquet alpin *Podisma pedestris*

(Linnaeus, 1758): pseudogènes et évolution.

VI.1. Problématique et démarche scientifique

L'évolution moléculaire de l'ADN ribosomal (rDNA) est essentiellement différente de celle des gènes présents en copie unique dans la cellule. Les copies de rDNA n'évoluent pas indépendamment l'une de l'autre : elles suivent un processus d'évolution concertée, qui résulte en une homogénéisation structurale et en un niveau bas de divergence génétique, au sein d'une même espèce. Par contre, les différences génétiques s'accumulent entre les espèces. Pour les applications phylogénétiques, l'élimination de la variabilité intra individuelle du rDNA par évolution concertée est une hypothèse préalable souvent invoquée. Pourtant, un nombre croissant d'études indique que, pour de nombreux taxons, le taux d'homogénéisation du rDNA pourrait être trop bas pour éviter un niveau élevé de polymorphisme intraspécifique.

Si l'évolution concertée peut se faire grâce à des processus moléculaires comme le crossing-over inégal et la conversion génique (e.g. Elder et Turner 1995 ; Graur et Li 2000), ces mécanismes pourraient aussi générer des pseudogènes. Pour cette raison, certains auteurs s'interrogent sur l'utilité du rDNA dans les analyses phylogénétiques (Alvarez et Wendel 2003) bien que leurs séquences soient encore utilisées intensément pour une large palette de taxons (plus de 3000 publications depuis 2000).

Un aspect important de ces mécanismes d'uniformisation et translocation des séquences est leur dynamique, qui pourrait avoir un sens évolutif en elle-même. Chez les orthoptères, les loci de rDNA diffèrent par leur nombre et leur localisation dans le génome entre les populations d'une même espèce (e.g. Flanagan et al. 1999, Cabrero et al. 2003). Dans ce contexte, le criquet alpin *Podisma pedestris* est intéressant pour étudier la dynamique évolutive du génome car il possède un génome d'une très grande taille (18000 Mb ;

Westerman et al. 1987), qui pourrait être la conséquence directe du bas taux relatif de perte d'ADN observé chez cette espèce.

Dans ce contexte, nous nous sommes attachés (i) à identifier et dater les pseudogènes d'ADN ribosomal chez *P. pedestris* et (ii) à proposer des scénarios concernant le rôle des pseudogènes dans l'évolution du génome.

VI.2. Contribution scientifique correspondant à l'article (C):

Keller, I., Chintauan-Marquier, I.C., Veltsos, P., Nichols, R.A. 2006. **Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution.** Genetics 174: 863-871.

VI.2.1. Principaux résultats

Nos résultats montrent qu'un seul individu peut contenir plusieurs groupes d'ADN ribosomal, très divergents entre eux. Les analyses phylogénétiques nous ont permis d'identifier huit groupes de séquences d'ADN ribosomal: deux groupes supposés fonctionnels et six identifiés comme des pseudogènes, d'après les mutations présentes dans la région 3' du fragment 18S. La datation moléculaire de la phylogénie suggère une co-existence des ces groupes de plus de 10 millions d'années, avec la divergence la plus récente il y a 12,6 à 6,6 million d'années. L'analyse des séquences amplifiées dans des conditions différentes de PCR montre que la fréquence d'amplification des deux groupes des gènes supposés fonctionnels est variable.

VI.2.2. Discussion

Nous pensons que les mutations au niveau du locus 18S, qui est d'habitude bien conservés entre des espèces très divergentes, sont une preuve que les six groupes mentionnés

antérieurement sont bien des pseudogènes. La datation suggérant une évolution indépendante de plus de 11 millions d'année est en contradiction avec l'hypothèse qu'un tel segment d'ADN garderait une unité structurale par le processus d'évolution concertée. Nous pensons que, dans le cas de *Podisma pedestris*, le mécanisme d'évolution concertée est défaillant car le génome de cette espèce est particulièrement grand (e.g. Westerman et al. 1987). De plus, il est possible que la grande taille du génome de *Podisma* soit générée par accumulation des pseudogènes. Ce processus est plausible d'autant plus qu'il a été prouvé que, le taux relatif de perte d'ADN ribosomal chez *Podisma pedestris* est inférieur celui de *Drosophila melanogaster* ou *Caenorhabditis elegans* (Petrov et al. 1996, Bensasson et al. 2001), ce qui favorise la pérennité des pseudogènes dans les cellules.

VI.2.3. Article - Keller, I., **Chintauan-Marquier, I.C.**, Veltsos, P., Nichols, R.A. 2006.

Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution.

Genetics 174: 863-871.

Ribosomal DNA in the Grasshopper *Podisma pedestris*: Escape From Concerted Evolution

Irene Keller,^{*†} Ioana C. Chintauan-Marquier,^{*†} Paris Veltsos^{*} and Richard A. Nichols^{*}

^{*}School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, United Kingdom and

[†]Laboratoire d'Ecologie Alpine, Université Joseph Fourier, 38041 Grenoble Cedex 09, France

Manuscript received May 26, 2006

Accepted for publication August 7, 2006

ABSTRACT

Eukaryote nuclear ribosomal DNA (rDNA) typically exhibits strong concerted evolution: a pattern in which several hundred rDNA sequences within any one species show little or no genetic diversity, whereas the sequences of different species diverge. We report a markedly different pattern in the genome of the grasshopper *Podisma pedestris*. Single individuals contain several highly divergent ribosomal DNA groups. Analysis of the magnitude of divergence indicates that these groups have coexisted in the *Podisma* lineage for at least 11 million years. There are two putatively functional groups, each estimated to be at least 4 million years old, and several pseudogene groups, many of which are transcribed. Southern hybridization and real-time PCR experiments show that only one of the putatively functional types occurs at high copy number. However, this group is scarcely amplified under standard PCR conditions, which means that phylogenetic inference on the basis of standard PCR would be severely distorted. The analysis suggests that concerted evolution has been remarkably ineffective in *P. pedestris*. We propose that this outcome may be related to the species' exceptionally large genome and the associated low rate of deletion per base pair, which may allow pseudogenes to persist.

THE molecular evolution of ribosomal DNA (rDNA) differs fundamentally from that of single-copy genes. Individual rDNA copies do not evolve independently from each other but are homogenized horizontally, so that little differentiation is typically observed between rDNA sequences within a tandem array of any one species. Conversely, differences do accumulate between species. This pattern is referred to as concerted evolution and may involve a number of molecular processes such as unequal crossing over and gene conversion (Elder and Turner 1995; Gaur and Li 2000). For phylogenetic applications, in particular, it has often been assumed that concerted evolution has essentially eliminated within-individual rDNA variation. However, an increasing number of studies indicate that, in many taxa, the rate of homogenization may be too low to prevent significant levels of intraspecific rDNA polymorphisms. Often this variation is restricted to the noncoding rDNA regions such as the internal transcribed spacers (ITS1 and ITS2) or the intergenic spacer (IGS), which separates the transcribed units (Cr ease 1995; Copenhaver and Pikaard 1996; O'Donnell and Cigel nik 1997; Fenton et al. 1998; Gernandt and List on 1999; Hugal l et al. 1999; Denduangboripant and Cronk 2000; Harris and Crandal l 2000; Whang et al. 2002; Parkin and Butlin 2004). However, there are also numerous reports of rDNA pseudogenes where

the coding regions are freed from functional constraints (Buckler et al. 1997; Hartmann et al. 2001; Mayol and Rossello 2001; Muir et al. 2001; Márquez et al. 2003; Razafimandimbison et al. 2004; Ruggiero and Procaccini 2004).

Substantial differentiation within individuals, corresponding to several million years of divergence, has been found in plants. These large differences, however, may be attributed to divergence that has accumulated in different species and has then been brought together by hybridization (Lim et al. 2000; Hartmann et al. 2001; Muir et al. 2001). The rDNA of hybrids can diverge rapidly from either parent within two or three generations (Kovarík et al. 2004). For this reason, and because of the divergence between different rDNA arrays and pseudogenes, some authorities consider rDNA to be a poor choice for phylogenetic analysis (Álvarez and Wendel 2003). Nevertheless, rDNA sequences are still extensively used in a wide range of taxa: a citation search detects 3000 papers since 2000.

More important than these concerns are the dynamics of sequence homogenization and translocation, which are of evolutionary interest in themselves. There is some evidence that the chromosomal location of rDNA arrays is particularly labile in the Orthoptera, with loci differing in number and genomic location in different populations of the same species (Flanagan et al. 1999; Cabrero et al. 2003). In the grasshopper *Podisma pedestris*, there is additional interest in the rDNA because evidence from restriction digests shows strong linkage

^{*}Corresponding author: Schweizerische Vogelwarte Sempach, 6204 Sempach, Switzerland. E-mail: irene.keller@vogelwarte.ch

disequilibrium between a particular rDNA variant and a chromosomal fusion (Dal I as et al. 1988), which cannot be explained by close physical proximity to the main rDNA locus (Bel I a et al. 1991). This disequilibrium hints at the possibility of selection maintaining the association. Further, *P. pedestris* has an exceptionally large genome of 18,000 Mb (West er man et al. 1987), which may be a direct consequence of the low relative rate of DNA loss observed in this species. Bensasson et al. (2001) have shown that, in *P. pedestris*, pseudogenes are likely to be slowly eroded by substitutions, whereas in species with high rates of DNA loss (e.g., *Drosophila*) they will not persist long enough for much genetic divergence to accumulate.

In this study, we surveyed cloned DNA to characterize variation in the ITS1 of the rDNA genic region within and between individuals of *P. pedestris*. To characterize the diversity within the genome, we have compared the sequences obtained from genomic DNA by standard PCR and high-stringency PCR and from ribosomal RNA by reverse transcription PCR (RT-PCR). Pseudogenes have been distinguished from putatively functional sequences by DNA sequence analysis and timings estimated for the date of separation of different lineages. We assess different possible explanations for, and implications of, the extraordinarily deep divergence that we detect among the different lineages.

MATERIALS AND METHODS

Study species and sample collection: The grasshopper species *P. pedestris* (Orthoptera, Acrididae, Melanoplinae) has a boreal and subalpine distribution. The species includes two races, which are called "fused" (or neo-XY) and "unfused" (or XO), respectively, due to differences in their X chromosomes (John and Hewitt 1970). The two races meet and interbreed along a narrow zone that runs from east to west following the ridges of the Alpes Maritimes in southern France (Bart on and Hewitt 1981). This hybrid zone is likely to have existed at its present location for ~ 8000 years (Nichols and Hewitt 1986).

Samples of *P. pedestris* were collected in 2003 and 2004 in the Alpes Maritimes from six pure populations on each side of the hybrid zone (Table 1). The insects were killed and the hind legs stored in 100% ethanol. For RNA analysis, testicular tissue was stored in RNAlater buffer (Ambion, Austin, TX) at 4°. As outgroups, we included sequences from eight other species of the subfamily Melanoplinae that had been obtained by I. C. Chintauan-Marquier as part of her Ph.D. thesis (Table 1).

Our study was divided into two main parts. First, the general diversity of rDNA was characterized by cloning and sequencing of PCR products. And second, on the basis of these results, we developed assays to investigate the frequency of specific rDNA variants.

The rDNA diversity in *P. pedestris*

DNA extraction and PCR: DNA was extracted from muscle tissue of one or both hind legs using either a high salt extraction method (Aljanabi and Martinez 1997) or a QIAGEN (Chatsworth, CA) DNeasy tissue kit according to the manufacturer's instructions. ITS1 was PCR amplified using the primers 18S(f) (59CCTTTGTACACCGCCCGT) and ITS6(r) (59GT

TCATGTGTCCTGCAGTTCAC), which target the coding regions flanking ITS1 (Sharpe et al. 2000). The resulting PCR product was ~ 650 bp in length and included 119 bp from the 39 end of the 18S coding region and all of ITS1. PCR amplifications were carried out in volumes of 25 µl using a PTC-220 DNA Engine Dyad Peltier thermal cycler (MJ Research, Watertown, MA). Each reaction contained ~ 5–10 ng of genomic DNA, 3 mM MgCl₂, 0.12 mM of each dNTP, 12.5 pmol of each primer, 0.25 units of BIOTAQ DNA polymerase (Bioline) and the associated NH₄ buffer at 13 concentration. An initial denaturation step of 5 min at 95° was followed by 35 cycles of amplification (30 sec at 94°, 30 sec at 62°, 1 min at 72°) and a final elongation step of 10 min at 72°.

The diversity of ITS1 was studied in particular detail in two individuals, one from the fused (individual BL1) and one from the unfused race (individual CM1). Several steps were taken to avoid problems that may arise in PCR surveys of multigene families such as rDNA. PCR can give misleading results due to the preferential amplification of certain target sequences and/ or the formation of PCR artifacts. The impact of stochastic events can be reduced by stopping the PCR at an early stage and pooling several independent amplifications before further analysis (Wagner et al. 1994; Kanagawa 2003). For the cloning and sequencing of rDNA in individuals BL1 and CM1, we combined the PCR products of four independent reactions, each of which had included only 29 cycles of amplification. There may also be a systematic bias in the efficiency with which different target sequences are amplified. For example, it has been shown that nonfunctional rDNA copies (i.e., pseudogenes) may be favored under standard PCR conditions (Buckler et al. 1997). For this reason, PCR was also performed under highly denaturing reaction conditions by adding 5% dimethyl sulfoxide (DMSO). This approach has been found to enhance the amplification of functional rDNA copies, possibly because they have high secondary structure stability (Buckler et al. 1997).

Similarly, the primers and stringent conditions used to obtain the outgroup sequences had been specifically developed to amplify only putatively functional sequences in Orthoptera. The strategy was successful in *Podisma* since it recovered group 1 sequence (see below; Figure 2). It is notable that, as in *Podisma*, the genic region of the outgroup sequences is identical to that of mouse (with the exception of *Melanoplus* sp., which was therefore treated separately in the analysis). This sequence conservation suggests strong selective constraints consistent with functionality.

RNA extraction and RT-PCR: For the individuals BL1 and CM1 we also studied ITS1 variation in RNA to investigate which copies are actually transcribed. The testicular tissue stored in RNAlater (Ambion) was homogenized by applying it to a QIAshredder column (QIAGEN) in a solution of 100 µl lysis buffer and 0.7 ml β-mercaptoethanol. RNA was then extracted using an Absolutely RNA RT-PCR Miniprep kit (Stratagene, La Jolla, CA) following the manufacturer's instructions with the suggested modifications for extremely small samples. The extraction protocol included an on-column DNase treatment.

cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen, San Diego) in a volume of 20 µl. Each reaction contained 400 units Superscript II, the associated first-strand buffer at 13 concentration, 10 mM DTT, 10 mM of each dNTP (dNTP set, Amersham Biosciences), 40 units RNasin Plus RNase inhibitor (Promega, Madison, WI), 50 pmol each of primers 18S(f) and ITS6(r) and 1.75 µl RNA template. The reaction was incubated at 37° for 1 hr. The cDNA was then used as template in a standard PCR as described above.

Cloning and sequencing: Before cloning, 20 µl of PCR product was run on a 1.5% agarose gel, and the band was excised and purified using a QIAquick gel extraction kit

TABLE 1
18S partial sequences and ITS1 sequences obtained in this study

Species	Authority	Origin	Individual	Sequences		Accession nos.
				18S	ITS1	
Melanoplinae	Scudder, 1897					
Podismini	Yakobson and Bianchi, 1905					
<i>P. pedestris</i>	Linnaeus, 1758	France				
Unfused		Belledonne	PPD1	1	1	AM295160
		Crête des Mélèzes	CM1	27	26	AM183598/ 600/ 603/ 604/ 612/ 621-623/ 636/ 648-665
		Crête des Mélèzes	CM2	3	3	AM183609/ 614/ 615
		Crête des Mélèzes	CM3	1	1	AM183611
		Crête des Mélèzes	CM4	1	1	AM183630
		Crête des Mélèzes	CM5	3	3	AM183606/ 625/ 626
		Col Bas	CB1	1	3	AM183620/ AM238436/ 437
		Lac Noir	LN1	2	3	AM183607/ 632/ AM238438
Fused		Blayeul	BL1	29	26	AM183588/ 593/ 594/ 597/ 601/ 602/ 608/ 616-619/ 624/ 628/ 629/ 631/ 634/ 635/ 637-647/ 666
		Blayeul	BL2	3	3	AM183589/ 590/ 595
		Blayeul	BL3	1	1	AM183596
		Blayeul	BL4	1	1	AM183599
		Blayeul	BL5	2	2	AM183627/ 633
		Chabanon	CH1	2	2	AM183587/ 591
		La Tartonne	TA1	3	3	AM183592/ 605/ 610
<i>Cophopodisma pyrenaica</i>	Fischer, 1853	Aude	CPD1	1	1	AM295161
<i>Italopodisma acuminata</i>	La Greca, 1969	Italy: Abruzzo	IPAA1	1	1	AM295162
<i>Snopodisma punctata</i>	Mistshenko, 1954	Japan: Tatsugou	SPP1	1	1	AM295163
Melanoplini	Scudder, 1897					
<i>Phaulotettix compressus</i>	Scudder, 1897	Mexico: Hidalgo	PLC1	1	1	AM295164
<i>Melanoplus frigidus</i>	Boheman, 1846	France: Lautaret	MLF1	1	1	AM295165
Dactyloptini	Rehn and Randell, 1963					
<i>Dactyloptum bicolor</i>	Charpentier, 1843	Mexico	DTBB1	1	1	AM295166
Jvarini	Ronderos, 1983					
<i>Jvarus</i> sp.	Giglio-Tos, 1898	South America	JSP1	1	1	AM295167
Dichroplini	Liebermann, 1949					
<i>Dichroplus</i> sp.	Stål, 1973	Ecuador: Tungurahua	DCHSP1	1	1	AM295168

(QIAGEN). The PCR fragments were then ligated into a pGEM-T Easy vector (Promega) overnight. Two microliters of the ligation reaction were used to transform JM109 High Efficiency Competent Cells (Promega) according to the manufacturer's recommendations. Individual positive clones were then used as a template for PCR amplification of ITS1 with the primers and conditions described above. The PCR products were cleaned and sequenced on an ABI 3700 sequencer. Table 1 lists the number of sequences obtained per individual and population. Note that differences between the number of 18S and ITS1 sequences arise in cases where either the beginning or the end of a sequence was of poor quality and omitted from further analyses. The EMBL accession numbers of the sequences are listed in Table 1.

Phylogenetic analyses: All sequences were split into two parts representing the 39end of the 18S and the ITS1, respectively. Partial 18S sequences were aligned in BioEdit 7.0.1 (Hall 1999) with published sequences from the grasshopper *Chorthippus parallelus* (EMBL accession no. AY585651), the fruit

fly *Drosophila melanogaster* (M21017), and the mouse *Mus musculus* (J00623).

The ITS1 sequences were edited manually in Sequence Navigator 1.01 (PE Applied Biosystems) and multiple alignments were generated using ClustalX 1.81 (Thompson et al. 1997). Different gap opening/ extension costs (10/ 0.2, 10/ 2, 10/ 5, 15/ 0.2, 15/ 5, 18/ 0.2, and 18/ 5) were used to identify the hypervariable and "more conserved" regions in different alignments of the same data. The final, unambiguous alignment contained 87 sequences of 437 bp of the ITS1 from different melanoplines (supplemental Table 1 at <http://www.genetics.org/supplemental/>).

The phylogenetic analyses were conducted for the ingroup alone as well as for ingroup and outgroups combined. As outgroups, we selected eight species representing six of the seven tribes of the Melanoplinae subfamily (Table 1) that had been suggested as close relatives of *P. pedestris* by previous molecular studies (Litzenberger and Chapco 2001; Amédegnato et al. 2003; I. C. Chintauan-Marquier, unpublished data).

Characteristics of the molecular data: Prior to phylogenetic reconstruction, the homogeneity of base frequencies across taxa was evaluated with a chi-square goodness-of-fit test using PAUP* 4.0b10 (Swofford 2002) because compositional bias among sequences can interfere with tree topology (Pereira et al. 2002). We included all sites and variable sites only to assess the potentially confounding effects of invariant sites (Waddell et al. 1999). The parameters allowed to vary in model fitting (eg, base composition, substitution rates, and rate heterogeneity across sites) were estimated under the most likely model of evolution suggested by the Akaike information criterion (Akaike 1974) in Modeltest 3.0 (Posada and Crandall 1998).

Maximum parsimony: Heuristic searches were implemented in PAUP* 4.0b10 (Swofford 2002) under the maximum-parsimony (MP) criteria with the accelerated transformation (ACCTRAN) option to optimize the state of unordered (Fitch) characters, 100 random sequence addition replicates, the tree bisection-reconnection (TBR) branch swapping, and gaps treated as a fifth character. Unweighted MP methods do not always take full advantage of the information contained in DNA sequences due to the presence of homoplasious characters. To deal with this problem, a weighted parsimony analysis was also conducted using the rescaled consistency index on an initial tree in successive approximations of character weighting (Farris 1969).

Maximum likelihood: The maximum-likelihood (ML) search was conducted using PAUP* 4.0b10 as follows: We first evaluated the model of DNA substitution that best fit our data partition, using the Akaike information criterion in Modeltest 3.0 (Posada and Crandall 1998). Then we calculated the ML model parameters using the data partition of interest and a neighbor-joining topology. These estimated values were fixed, and a full heuristic ML search was conducted with 10 random addition sequence replicates, retaining all minimal trees, and TBR branch swapping. After this first ML search was completed, we reestimated model parameters on the ML tree and used these new values to search again. We repeated this process until the new search found the same topology as the previous search and all parameter estimates were identical.

Bayesian analyses: Bayesian Markov chain Monte Carlo (MCMC) analyses were performed with MrBayes 2.01 (Huelsenbeck and Ronquist 2001) with the GTR 1 + I + G model. To assess the coverage of tree space, we performed three independent Metropolis-coupled MCMC runs started from different random points in parameter space (Huelsenbeck and Bollback 2001; Huelsenbeck et al. 2002). We applied different temperatures (0.2°, 0.5°, 1°) to influence the rate of switching between chains. Each run had 10 million iterations, with one cold and three incrementally heated chains per run and with trees sampled every 500 generations. The number of trees to be discarded as burn-in (Huelsenbeck and Ronquist 2001) was determined graphically. The average log-likelihood values (6 SD) at stationarity for the independent runs were calculated with Microsoft Excel and compared for convergence (Huelsenbeck and Bollback 2001). The set of model substitution parameters were analyzed to determine the mean, variance, and 95% credible interval. To estimate posterior probabilities (PPs) of recovered branches (Larget and Simon 1999; Huelsenbeck and Ronquist 2001), 50% majority rule consensus trees were calculated from the remaining trees in MrBayes 2.01.

Node support: Support for internal branches was assessed in PAUP* 4.0b10 (Swofford 2002) by nonparametric bootstrapping (nBT) (Felsenstein 1985) with 2000 pseudoreplicates for MP, using full heuristic searches with 30 random addition sequence replicates, TBR branch swapping, and one tree held at each step during stepwise addition (DeBry and

Olmstead 2000). PPs for individual nodes were estimated in a Bayesian MCMC framework with the runs executed as above. Nodes with nBT values $\geq 65\%$ and PPs $\geq 75\%$ were shown as resolved (Hillis and Bull 1993; Jordan et al. 2003).

Molecular dating: A likelihood-ratio test (LRT; Felsenstein 1981; Goldman 1993a,b; Yang and Roberts 1995) showed that the data did not conform to the model of a constant molecular clock (LRT χ^2 283, P χ^2 0, 85 d.f.). We therefore implemented Yoder and Yang's local clock method (Yoder and Yang 2000). We made use of the general time reversible plus G-distributed rate variation (GTR 1 + G) option with local clocks implemented in PAML v3.13 (Yang 1997). Each branch was assigned to a rate category after inspection of the branch lengths in Figure 2 and separate rates (R1–R3) were estimated for each. The branches in each of the three categories are identified in Figure 2: R1 applies to the branches in blue, R2 to those in red, and R3 to the remainder.

We used three alternative calibrations of the molecular clock to cross-check the timings estimated from the genetic divergence. Each calibration was based on a date that had previously been proposed for a particular node on the phylogeny: (1) the separation of the South American melanoplinae thought to correspond to an incursion into North America 34.43–74.05 MYA (Chapco et al. 2001; Amedegnato et al. 2003; I. C. Chintauan-Marquier, unpublished data); (2) more recently, the North American melanoplinae are thought to have split from the Eurasian in the early Tertiary/late Cretaceous 20.53–66 MYA (Chapco et al. 2001; Amedegnato et al. 2003; I. C. Chintauan-Marquier, unpublished data); and (3) the segregation of the European from the Asian melanoplinae fauna has been estimated at 16.13–39.81 MYA (Litzenberger and Chapco 2001; I. C. Chintauan-Marquier, unpublished data). In each case, the date required by PAML's maximum-likelihood dating procedure was taken at the midpoint of the above ranges. The ages of the geological periods and chronostratigraphic references have been taken from the 1999 Geological Time Scale of the Geological Society of America (<http://www.geosociety.org/science/timescale/timescl.pdf>).

Detection of potential recombinants: We used two different statistical approaches to detect recombinant sequences. The RDP method (Martin et al. 2005) and the MaxChi method (Maynard Smith 1992; Posada and Crandall 2001) were applied to the 77 complete sequences.

Frequency of specific variants

Restriction digests: On the basis of the results of the sequencing it was possible to identify restriction enzymes that were expected to cut only one particular group of sequences. *Avall* (G¹GWCC) is expected to cut group 1 sequences (see below) into two fragments of ≈ 150 and 500 bp. *Avall* restriction sites are also present in three other sequences. However, digestion of these sequences is expected to produce fragments of a different length. The restriction enzyme *AlwNI* (CAGNNN¹CTG) is expected to cut only sequences belonging to group 2 and to produce two fragments of ≈ 400 and 250 bp. There are no *AlwNI* recognition sites in any non-group 2 sequences.

To investigate the within-individual frequency of the different sequence variants in more detail, DNA from all *P. pedestris* individuals used for sequencing (except PPD1; Table 1) was analyzed by restriction digestion. PCR was carried out as described above for the individuals BL1 and CM1 using both standard and highly denaturing reaction conditions. A total of 25 μ l of PCR reaction was incubated with 10 units of *Avall* or *AlwNI* (New England Biolabs, Beverly, MA) at 37° for 18 hr. The restriction fragments were then separated on a 1.5% agarose gel.

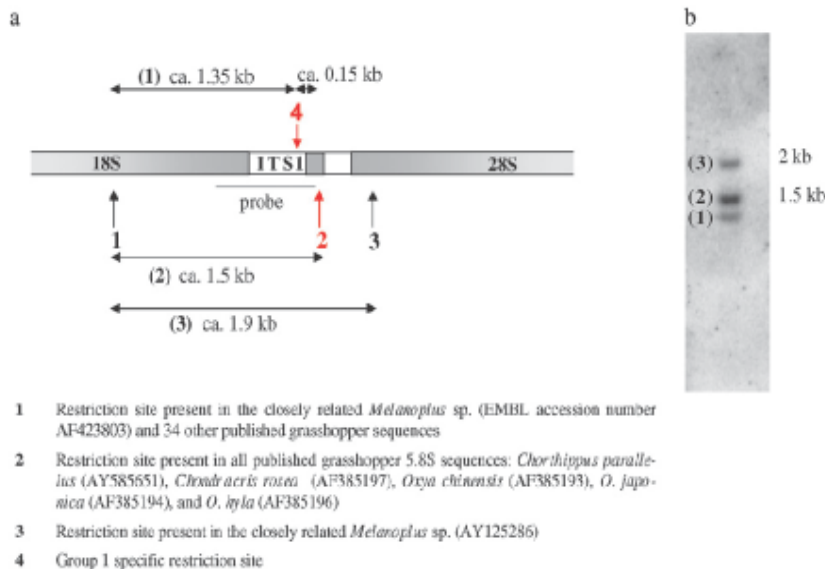


Figure 1.—(a) Location of *Ava*I restriction sites in the rDNA genomic region predicted on the basis of our own and published grasshopper sequences as referenced in the legend. The restriction sites indicated in red (2 and 4) can potentially be blocked by methylation. Site 4 in the ITS1 is characteristic of group 1 sequences. (b) *Ava*I digest of genomic DNA from *P. pedestris*.

- 1 Restriction site present in the closely related *Melanoplus* sp. (EMBL accession number AF423803) and 34 other published grasshopper sequences
- 2 Restriction site present in all published grasshopper 5.8S sequences: *Chorthippus parallelus* (AY585551), *Chorthippus rosea* (AF385197), *Oxya chinensis* (AF385193), *O. japonica* (AF385194), and *O. lyda* (AF385196)
- 3 Restriction site present in the closely related *Melanoplus* sp. (AY125286)
- 4 Group 1 specific restriction site

Southern hybridization: Southern blot analyses were performed to estimate the relative frequency of group 1 sequences in the genome of *P. pedestris*. A total of 3–5 mg of genomic DNA from seven different individuals was digested to completion with *Ava*I. The digested DNA was ethanol precipitated and resuspended in 25 ml TE buffer before electrophoresis through a 1.2% agarose minigel for 8 hr. The DNA fragments were then blotted onto a nylon membrane (Hybond-N1, Amersham Biosciences) following the protocol of Sambrook et al. (1989). As a probe, ITS1 was amplified as described above and 100 ng of purified PCR product was labeled with [³²P]CTP using Ready-To-Go DNA labeling beads (Amersham Biosciences) according to the manufacturer's instructions. Prehybridization and hybridization of the membranes were carried out at 42° in solutions containing 50% formamide, 63 SSC, 53 Denhardt's reagent (prehybridization solution only), 0.5% SDS, and 100 mg/ml denatured, fragmented salmon sperm DNA. After overnight hybridization, membranes were washed in large volumes of 23 SSC and 0.5% SDS twice for 5 min, in 23 SSC and 0.1% SDS for 15 min, and in 0.13 SSC and 0.5% SDS for 30 min at room temperature and again for 30 min at 65°. They were then exposed to a storage phosphor screen for 1–2 hr, and the results were analyzed on a Typhoon 9200 scanner (Amersham Biosciences).

On the basis of published grasshopper sequences it was possible to predict the lengths of *Ava*I restriction fragments (Figure 1a). Unmethylated group 1 sequences are expected to produce two bands at 1.35 and 0.15 kb, while all other rDNA variants should give one band at 1.5 kb. In methylated DNA, restriction sites 2 and 4 could be blocked, resulting in just one fragment of 1.9 kb (Figure 1a).

Quantitative PCR with allele-specific primers: Quantitative PCR (qPCR) was used to estimate the relative frequency of the main groups of rDNA sequence. Primers specific for group 1 and group 2, respectively, were designed on the basis of sequencing results. We selected positions that were diagnostic of one particular group, i.e., where all sequences in that group shared a nucleotide that was not present in any of the other sequences. Both the forward and reverse primers were designed to end on such a diagnostic base and show a perfect match only to the target sequence and a 39 mismatch to all others. The group 1-specific forward and reverse primers were 59GCTCCTCTCAAACCTAACGCAC and 59CAGAATA YATACAGGCGGGACC, and the group 2-specific forward and

reverse primers were 59CGTGTGCGGCATTGACAG and 59CGAAGCTCGCAGAATACATCT. To monitor the specificity of the primers, clones with known inserts (either group 1 or group 2) were used as PCR templates. Further, we sequenced PCR products to confirm allele-specific amplification from genomic DNA.

Since total rDNA content is thought to vary considerably among individuals (as indicated by fluorescent in situ hybridization on cytogenetic preparations; not shown), we standardized the qPCR results for the main groups by comparison with those for a 180-bp fragment from the 39end of the 18S coding region. This fragment was amplified using the primer 18S(f) described above and the new primer 18S(r) (59GATCCT TCCGCAGGTTTAC). These primers are expected to show a perfect match to all functional sequences and a good match to relatively young pseudogenes, which have not yet accumulated mutations in the highly conserved region.

The qPCR analysis was performed in a PTC-200 Peltier thermal cycler (MJ Research) with a Chromo4 fluorescence detector (GRI). Reactions were carried out in a 20-ml volume using a DyNAmo HS SYBR Green qPCR kit (New England Biolabs) according to the manufacturer's recommendations. Four replicates of each sample were analyzed with each set of primers. The cycling protocol consisted of an initial denaturation step of 15 min at 95°, 40 cycles of 10 sec at 94°, 30 sec at a primer-specific annealing temperature (T_A), 30 sec at 72° followed by a data acquisition step, and a final elongation of 5 min at 72°. The annealing temperatures for the different primer pairs were as follows: 18S, T_A ¼ 57°; group 1 specific, T_A ¼ 60°; group 2 specific, T_A ¼ 56°.

For each sample, the qPCR software determined the number of cycles, $C(t)$, after which the fluorescence intensity crossed a threshold that was set at five standard deviations above the baseline fluorescence. From these $C(t)$ values, the initial template concentration was estimated by comparing the samples to a serial dilution of a standard included in each run. As this standard, we used 16 different dilutions (in a twofold dilution series) of PCR products obtained from clones containing either a group 1 or a group 2 insert. These 16 dilutions spanned the template concentrations typically observed for genomic samples. Both series were included in runs with the 18S primers. On the basis of data from 12 individuals from both races, we estimated the average copy number of group 1 and group 2 sequences relative to the number of conserved 18S sequences.

Direct sequencing: To further investigate the genomic frequency of the different rDNA variants, DNA extractions from two individuals were sequenced directly without a preceding PCR step. 18S(f) was used as a sequencing primer. The electrophoretogram was analyzed for evidence of multiple peaks at the sites of SNPs that distinguish the variants.

RESULTS

High levels of intraspecific variability: The ITS1 sequences obtained from *P. pedestris* revealed a remarkable level of intraspecific variation with an average pairwise difference of 7.7% (SD $\frac{1}{4}$ 3.8%). Only two cDNA sequences from individual BL1 were identified as potential recombinants by MaxChi, while RDP did not find any evidence of recombination (data not shown). Figure 2 shows a maximum-likelihood tree based on ITS1, including sequences from eight related species.

The *P. pedestris* sequences form eight principal clusters that are well supported in all trees (Figure 2, Table 3). Clusters 1 and 2 have no more than one difference in the last 58 bp of 18S genic DNA from the conserved sequence found in *Drosophila*, *Chorthippus*, and *Mus*. They were therefore identified as putatively functional. The other six groups (groups 3–8) are identified as pseudogenes because they contain multiple differences in this highly conserved region. Most of these substitutions are shared by several sequences within the group. The three different methods for obtaining sequences recovered different proportions from each category (Table 2). Standard PCR and cDNA analysis recovered a high frequency of pseudogenes as well as members of the putatively functional groups 1 and 2. The frequencies obtained by these two methods were not significantly different (χ^2 -test, $P \frac{1}{4}$ 0.14). In contrast, PCR with DMSO recovered almost exclusively putatively functional group 1 sequences in a significantly higher proportion than the other methods (χ^2 -test, $P = 0.001$).

Different rDNA variants have coexisted in the *P. pedestris* genome for ≈ 10 million years: The age of the different nodes was estimated using the three distinct calibration points, all of which gave comparable results (Figure 2, Table 3). We found that the most recent common ancestor between some of the most divergent rDNA groups is likely to have existed between 10 and 20 million years ago in the Miocene (nodes 94 and 95, Table 3). This includes the split between the two putatively functional groups 1 and 2 (see previous paragraph). All eight well-supported groups (Figure 2) are likely to have been evolving independently since at least the late Miocene, with the most recent split 6.6–12.6 million years ago

between groups 2 and 3 (node 105). Lineages within both groups 1 and 2 have been diverging for at least 3.7 million years (nodes 96 and 107).

One putatively functional group occurs in much higher copy number than the other: The results of the Southern blots indicate that unmethylated group 1 sequences are found at a relatively high copy number but may not be the most common rDNA type. Digests with *Ava*I always resulted in three bands of \approx 1.3, 1.5, and 1.9 kb (Figure 1b). The middle band at 1.5 kb was always the brightest, followed by the band at 1.9 kb and finally the band at 1.3 kb. The intensity of the two weaker bands was generally very similar and amounted to 34–57% of the intensity of the brightest band. In two of the seven individuals, there were three additional, very weak bands at \approx 1, 0.75, and 0.4 kb attributable to other restriction sites. The size of the three bright bands corresponds to that predicted from published sequences (Figure 1a). Band 3 at 1.9 kb would be produced from methylated DNA and band 2 at 1.5 kb from sequences lacking the ITS1 site, while band 1 at 1.3 kb would result from digests of group 1 sequences (Figure 1b). The expected 0.17-kb fragment from group 1 sequences was not detected, most probably because of its short overlap with the probe. Our interpretation is, therefore, that nonmethylated group 1 sequences (band 1) are common, but make up less than half the total rDNA. This provides a lower boundary for the genomic frequency of group 1. The true frequency might be underestimated considerably if bands 2 and 3 contain a high proportion of methylated group 1 sequences.

The qPCR experiments show that group 1 does make up the majority (88%; SE 7%) of rDNA units with relatively well-conserved 18S (i.e., functional rDNA and young pseudogenes), whereas group 2 sequences are very rare, accounting for \approx 1%. For all the standard calibration series, we observed an almost perfect linear relationship between $C(t)$ values and the log of the initial copy number ($R^2 \approx 0.99$ in all cases). This result indicates that the initial copy numbers of the three different targets could indeed be estimated reliably for the genomic samples by qPCR.

The results of the direct sequencing of genomic DNA further supported the high frequency of group 1 sequences among the sequences with a conserved 18S. Without the initial PCR step, which has been shown to selectively amplify some variants (see below), the electrophoretogram traces were interpretable for only short runs (\approx 200 bp). These were, however, sufficient to show that all SNP loci were unambiguously of the group

Figure 2.—Maximum-likelihood tree of ITS1 sequences from *P. pedestris* and eight related species. Numbers in red indicate nodes that are well supported in both Bayesian and weighted maximum-parsimony analysis (boldface type), in Bayesian analysis only (italic), or in weighted maximum-parsimony analysis only (underlined). Sequences from individuals of the unfused race are represented by triangles; sequences from individuals of the fused race by squares. C1–3 indicate the nodes used as calibrations (see Table 3). Different symbols are used for sequences from different individuals. To estimate divergence times, the branches in blue were assumed to evolve under rate R1 and the branches in red under rate R2.

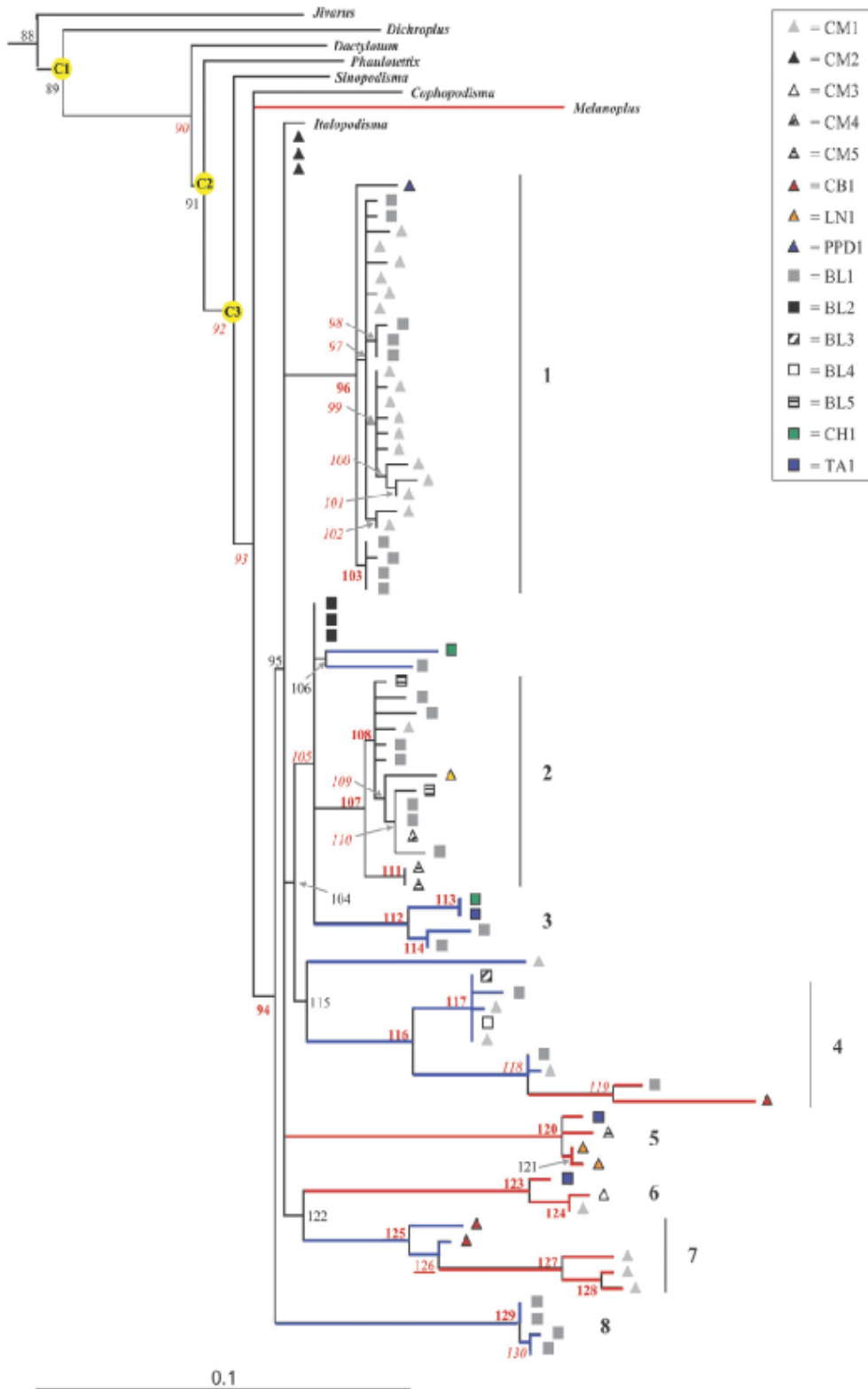


TABLE 2

Putatively functional sequences and pseudogenes obtained by cloning the PCR product from three different protocols: standard PCR, high-stringency PCR (with DMSO), and (RT-PCR)

	Putatively functional		Pseudogenes Groups 3–8
	Group 1	Group 2	
Standard PCR	3	9	23
PCR with DMSO	18	1	0
RT-PCR	5	4	7

The eight groups represent well-supported clades in a maximum-likelihood tree, of which two are putatively functional (groups 1 and 2) and the rest are pseudogenes (Figure 2).

1 type, with no detectable evidence of the alternative nucleotides found in the other rDNA groups.

PCR conditions strongly influence the amplification efficiency of different rDNA variants. The investigation of the sequences amplified under different PCR conditions confirmed that group 1 sequences are the most common product under the denaturing PCR conditions produced by DMSO. In each of the 14 individuals, *A*vaI digestion produced a very strong double band corresponding to the expected group 1 products, with only a very faint undigested band indicating other groups. There was no visible digestion by the group 2-specific enzyme, *Alw*NI—in agreement with the qPCR results suggesting that group 2 is relatively rare.

PCR under standard conditions resulted in a significantly different and very variable composition, corresponding to the diversity of sequences that had previously been cloned. Group 2-specific bands were detectable in all 14 individuals, whereas the proportion of group 1 sequences was generally lower and varied widely between individuals. In fact, group 1 sequences were completely undetectable in individuals from population BL, even though they were clearly present in the sequences amplified by PCR with DMSO.

DISCUSSION

Intragenomic rDNA diversity. We have detected eight distinct groups of ITS1 sequences in *P. pedestris* individuals. Phylogenetic analysis indicates that they have been evolving independently for ~11 million years. Mutations at 18S sites that are otherwise perfectly conserved between disparate species provide convincing evidence that six of the *Podisma* groups consist of pseudogenes, while two might be functional (groups 1 and 2). Southern blots indicate that the genomic frequency of pseudogenes is possibly higher than that of functional units. One of the putatively functional variants (group 1) also occurs at high copy number and is clearly the most frequent among sequences with a relatively well-

conserved 18S coding region. Interestingly, this common functional variant is scarcely amplified at all under standard PCR conditions. Group 2, on the other hand, occurs at very low copy numbers and therefore seems unlikely to have much functional significance.

Even within group 1, some of the lineages have been diverging for at least 3.7 million years. This result is in contrast to the expectation that functional arrays would be particularly prone to homogenization through concerted evolution (Kovarik et al. 2004). What could explain a low rate of homogenization in *Podisma*? Several studies strongly suggest that concerted evolution is much more potent for sequences on the same chromosome than for loci on different chromosomes (Schlötterer and Tautz 1994; Copenhaver and Pikaard 1996; Polanco et al. 1998; Parkin and Butlin 2004). The results of fluorescent in situ hybridization indeed show that multiple rDNA loci are present in *P. pedestris*. From two to seven loci have been detected, fused populations generally having two to three, whereas unfused populations vary from one population to another, usually having more (*P. veltsos*, unpublished data). However, the sequencing results detected many more haplotypes than there are loci, implying that individual arrays must have more than one haplotype and indicating low rates of concerted evolution even at the intrachromosomal level. Individual CM1, for example, has six different rDNA loci but a minimum of 14 different haplotypes from group 1 alone (Figure 2). Furthermore, the different sequence variants persist in the *P. pedestris* genome for millions of years, while the number and chromosomal location of the arrays changes extraordinarily rapidly. Populations BL and CM, for example, have no rDNA loci in common but are remarkably similar at the sequence level (Figure 2).

This inefficiency of concerted evolution could be associated with the exceptionally large genome of *P. pedestris* (Westerman et al. 1987), a genome that is 100 times larger than that of *Drosophila melanogaster* (Hartl 2000). There are extensive heterochromatic blocks (Bella et al. 1991) and restriction digests (Figure 1b) show evidence of methylation, unlike *Drosophila*. Methylated sequences and DNA condensation, possibly associated with heterochromatinization, have been implicated in markedly reducing rates of concerted evolution (Lim et al. 2000).

It is possible that there is a causal link between genome size and the abundance of rDNA pseudogenes in this species. It has been demonstrated that the relative rate of DNA loss (per kilobases of sequence) is significantly lower in *P. pedestris* than in model organisms such as *D. melanogaster* or *Caenorhabditis elegans* with their compact genomes (Petrov et al. 1996; Robertson 2000; Bensasson et al. 2001). Pseudogenes in these small genomes are being removed rapidly relative to mutation (Petrov 2001), so that little genetic divergence accumulates over their lifetime before they are

TABLE 3

Maximum-likelihood estimates of divergence times (in million years) under a local clock using three calibrations (C1, C2, and C3) and support of the nodes based on Bayesian (PP) and weighted maximum-parsimony analysis (BT)

Node	Divergence times			Mean	95% C.I.		Support of nodes	
	C1	C2	C3		Lower limit	Upper limit	PP	BT
88	60.593	107.581	95.119	87.764	60.219	115.309	NS	NS
89	54	95.876	84.77	78.215	53.667	102.764	NS	NS
90	26.563	47.162	41.699	38.475	26.399	50.550	100	NS
91	24.219	43	38.019	35.079	24.070	46.089	50	NS
92	17.836	31.668	28	25.835	17.726	33.943	81	NS
93	13.483	23.939	21.166	19.529	13.400	25.659	75	NS
94	11.25	19.974	17.66	16.295	11.181	21.409	87	94
95	9.914	17.603	15.564	14.360	9.853	18.868	NS	NS
96	3.739	6.638	5.869	5.415	3.716	7.115	100	87
97	2.5	4.438	3.924	3.621	2.485	4.757	86	54
98	0.63	1.118	0.989	0.912	0.626	1.198	100	60
99	1.912	3.395	3.002	2.770	1.900	3.639	87	NS
100	1.509	2.679	2.368	2.185	1.500	2.871	95	64
101	0.91	1.617	1.429	1.319	0.904	1.733	91	60
102	1.399	2.484	2.196	2.026	1.390	2.662	94	63
103	0.53	0.942	0.833	0.768	0.527	1.010	90	86
104	9.425	16.734	14.796	13.652	9.367	17.936	54	NS
105	6.645	11.798	10.431	9.625	6.604	12.645	92	NS
106	4.814	8.547	7.557	6.973	4.784	9.161	NS	NS
107	4.335	7.698	6.806	6.280	4.308	8.251	100	94
108	3.77	6.693	5.917	5.460	3.747	7.173	86	66
109	3.201	5.684	5.025	4.637	3.181	6.092	76	63
110	1.767	3.138	2.774	2.560	1.756	3.363	97	63
111	0	0	0	0.000	0.000	0.000	100	98
112	2.096	3.721	3.29	3.036	2.083	3.988	100	99
113	0	0	0	0.000	0.000	0.000	100	100
114	1.253	2.224	1.966	1.814	1.245	2.383	95	71
115	9.12	16.192	14.316	13.209	9.064	17.355	NS	NS
116	5.338	9.478	8	7.605	5.231	9.979	100	93
117	0.435	0.772	0.683	0.630	0.432	0.828	100	100
118	2.275	4.039	3.571	3.295	2.261	4.329	100	NS
119	1.326	2.354	2.081	1.920	1.318	2.523	100	NS
120	0.618	1.098	0.971	0.896	0.614	1.177	100	100
121	0.171	0.303	0.268	0.247	0.170	0.325	60	53
122	7.866	13.965	12.348	11.393	7.818	14.968	NS	NS
123	1.023	1.817	1.606	1.482	1.017	1.947	100	100
124	0.238	0.423	0.374	0.345	0.237	0.453	100	98
125	3.93	6.977	6.169	5.692	3.906	7.478	100	96
126	3.049	5.414	4.787	4.417	3.030	5.803	55	78
127	1.118	1.986	1.756	1.620	1.111	2.129	100	100
128	0.341	0.605	0.535	0.494	0.339	0.648	100	98
129	0.289	0.513	0.453	0.418	0.287	0.550	100	100
130	0.124	0.22	0.194	0.179	0.123	0.236	96	64

Node, number of a particular node as indicated in Figure 2; 95% C.I., 95% confidence interval; PP, posterior probability; BT, bootstrap support; NS, node not supported.

eroded away by deletion. The *Drosophila* genome, in particular, is characterized by a remarkable paucity of pseudogenes (Harrison et al. 2003). In *Podisma*, by contrast, the rate of removal is five hundred times lower than in *Drosophila*, and only 5% of the point substitution rate (Bensasson et al. 2001), which would help explain why the pseudogenes have accumulated such remarkable divergence.

It is, at first sight, surprising that the cDNA sequences obtained from transcribed rRNA include many potential pseudogenes. However, pseudogenes can still be transcribed, provided that their transcriptional control elements remain functional. Expressed pseudogenes have indeed been documented (Mighell et al. 2000; Hirotsune et al. 2003; Yamada et al. 2003). It follows that transcription is not an indicator of functionality

(Bailey et al. 2003). Pseudogene transcripts can actually be more abundant than RNA from functional paralogues (Fujii et al. 1999). It is possible that the pool of standing rRNA in the cell has a disproportionately high frequency of pseudogenes. ITS1 is included in the primary transcript and is excised during the processing of the precursor rRNA (van Nues et al. 1995). If this process is less efficient for pseudogene transcripts, they could persist in the cell and hence be the more frequent template for our RT-PCR. One particular attribute of the cDNA sequences, which could interfere with efficient splicing, is that three quarters of them contain deletions in the ITS1. Our observation of very deeply diverged pseudogenes in cDNA indicates that at least some of these copies are accessible to the nucleus's transcriptional machinery, although not necessarily within active nucleoli.

Methodological issues associated with the study of multigene families: Our study illustrates PCR drift and a particularly severe case of bias (Table 2), which can affect surveys of multigene families (Wagner et al. 1994; Kanagawa 2003). The results obtained under standard reaction conditions appear to provide a particularly inaccurate estimate of the actual genomic frequencies of different sequence variants. Group 2 sequences, for example, seem to be preferentially amplified and are consistently present in standard PCR products even though their genomic frequency is very low (as demonstrated by three approaches: qPCR, direct sequencing, and digestion of PCR products). The most highly repeated functional group, group 1, on the other hand, is often not amplified at all under standard PCR conditions. Fenton et al. (1998) and Hartmann et al. (2001) also found that particular rDNA types were readily amplified by PCR even though Southern blots showed that their genomic frequency was very low.

The addition of DMSO to the reaction dramatically alters the frequency of the different sequence variants in the pool of PCR products. Group 1 sequences consistently predominate while group 2 sequences become undetectable in the restriction digests. Buckler et al. (1997) have suggested that DMSO may facilitate the amplification of sequences with a more stable secondary structure. At first glance, this would seem consistent with our results, since the ITS1 of group 1 units tends to have higher predicted secondary-structure stability than the other variants (data not shown). However, a follow-up experiment indicated that PCR efficiency was little affected by these sequence differences. Using standard conditions, we amplified mixtures containing varying proportions of group 1 and group 2 clones. In contrast to the results from the genomic DNA extracts, group 1 sequences were clearly detectable in restriction digests of the PCR product whenever they were present in the template mixture. Further, more group 1 sequences were recovered from the genomic DNA extracts when the annealing temperature was increased to 68°, con-

sistent with the proposal that the effect of the DMSO had been to increase stringency.

Using the hybrid zone to date the dynamics of the rDNA loci: The biogeography of the Alps indicates that the hybrid zone must be $\approx 10,000$ years old and that the initial separation of the chromosomal races occurred at least 100,000 years ago and possibly several times longer ago than that (Barton and Hewitt 1981). The existence of the different rDNA haplotypes on both sides of the zone is therefore consistent with our much older estimates for their age of origin. Even though Dal Mas et al. (1988) found very strong linkage disequilibrium between the presence of a particular IGS restriction site and the chromosomal fusion in one locality, they did not find the association elsewhere. One possible mechanism that could both make and break such associations is the movement of the rDNA variants around the genome. This process does indeed seem to be occurring, as reflected in the differing number and location of rDNA loci on either side of the zone, and indeed between populations on the same side of the zone. It follows, then, that the ancient rDNA lineages within *Podisma* are persisting despite this substantial intra-chromosomal mobility having occurred since they arose.

We thank Lee Crainey, Thomas Giger, and Raj Joseph for substantial help with molecular work and Hussna Jahan and Joe Nichols for assistance in the field. We are very grateful to Andrew Leitch, Carlo Largiadèr, Douda Bensasson, and Steve Rossiter for helpful discussions and to anonymous reviewers for valuable comments on earlier versions of this manuscript. This work was financially supported by fellowships from the Swiss National Science Foundation (grant PBEEA-104447) and the Roche Research Foundation to I.K. and a Natural Environment Research Council grant to R.A.N. (NER-B-S2003-00859).

LITERATURE CITED

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19: 716–723.
- Aljanabi, S., and I. Martinez, 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* 25: 4692–4693.
- Álvarez, I., and J. F. Wendel, 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29: 417–434.
- Amedegnato, C., W. Chapco and G. Litzenberger, 2003. Out of South America? Additional evidence for a southern origin of melanopline grasshoppers. *Mol. Phylogenet. Evol.* 29: 115–119.
- Bailey, C. D., T. G. Carr, S. A. Harris and C. E. Hughes, 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol. Phylogenet. Evol.* 29: 435–455.
- Barton, N. H., and G. M. Hewitt, 1981. A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution* 35: 1008–1018.
- Bella, J. L., M. Westerman, C. Lopezferandez, J. Delatorre, J. M. Rubio et al., 1991. Sex-chromosome and autosome divergence in *Podisma* (Orthoptera) in Western-Europe. *Genet. Sel. Evol.* 23: 5–13.
- Bensasson, D., D. A. Petrov, D.-X. Zhang, D. L. Hartl and G. M. Hewitt, 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol. Biol. Evol.* 18: 246–253.
- Buckler, E. S., A. Ippolito and T. P. Holtsford, 1997. The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics* 145: 821–832.
- Cabrero, J., F. Perfectti, R. Gomez, J. P. M. Camacho and M. D. Lopez-Leon, 2003. Population variation in the A chromosome distribution of satellite DNA and ribosomal DNA in the grasshopper *Eyprepocnemis plorans*. *Chromosome Res.* 11: 375–381.

- Chapco, W., G. Litzenberger and W. R. Kuperus, 2001 A molecular biogeographic analysis of the relationship between North American melanoploid grasshoppers and their Eurasian and South American relatives. *Mol. Phylogenet. Evol.* 18: 460–466.
- Copenhaver, G. P., and C. S. Pikaard, 1996 Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* 9: 273–282.
- Crease, T. J., 1995 Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* 141: 1327–1337.
- Dallas, J. F., N. H. Barton and G. A. Dover, 1988 Interracial rDNA variation in the grasshopper, *Podisma pedestris*. *Mol. Biol. Evol.* 5: 660–674.
- DeBry, R. W., and R. G. Olmstead, 2000 A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Syst. Biol.* 49: 171–179.
- Denduangboripant, J., and Q. C. B. Cronk, 2000 High intraindividual variation in internal transcribed spacer sequences in *Aeschynanthus* (Gesneriaceae): implications for phylogenetics. *Proc. Biol. Soc.* 267: 1407–1415.
- El der, J. F., and B. J. Turner, 1995 Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70: 297–320.
- Farris, J. S., 1969 A successive approximations approach to character weighting. *Syst. Zool.* 18: 374–385.
- Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17: 368–376.
- Felsenstein, J., 1985 Confidence limits on phylogenies—an approach using the bootstrap. *Evolution* 39: 783–791.
- Fenton, B., G. Malloch and F. Germa, 1998 A study of variation in rDNA ITS regions shows that two haplotypes coexist within a single aphid genome. *Genome* 41: 337–345.
- Flanagan, N. S., P. L. Mason, J. Gosálvez and G. M. Hewitt, 1999 Chromosomal differentiation through an alpine hybrid zone in the grasshopper *Chorthippus parallelus*. *J. Evol. Biol.* 12: 577–585.
- Fujii, G. H., A. M. Morimoto, A. E. Berson and J. B. Bolen, 1999 Transcriptional analysis of the PTEN/MMAC1 pseudogene, (phi)PTEN. *Oncogene* 18: 1765–1769.
- Gernandt, D. S., and A. Liston, 1999 Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *Am. J. Bot.* 86: 711–723.
- Golman, N., 1993a Simple diagnostic statistical tests of models for DNA substitutions. *J. Mol. Evol.* 37: 650–661.
- Golman, N., 1993b Statistical tests of models of DNA substitutions. *J. Mol. Evol.* 36: 182–198.
- Graur, D., and W. H. Li, 2000 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95.
- Harris, D. J., and K. A. Crandall, 2000 Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. *Mol. Biol. Evol.* 17: 284–291.
- Harrison, P. M., D. Milburn, Z. Zhang, P. Bertone and M. Gerstein, 2003 Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res.* 31: 1033–1037.
- Hartl, D. L., 2000 Molecular melodies in high and low C. *Nat. Rev. Genet.* 1: 145–149.
- Hartmann, S., J. D. Nason and D. Bhatnagar, 2001 Extensive ribosomal DNA gene variation in the columnar cactus *Lophocereus*. *J. Mol. Evol.* 53: 124–134.
- Hillis, D. M., and J. J. Bull, 1993 An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42: 182–192.
- Hiratsune, S., N. Yoshida, A. Chen, L. Garrett, F. Sugiyama et al., 2003 An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423: 91–96.
- Huel senbeck, J. P., and J. P. Bollback, 2001 Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* 50: 351–366.
- Huel senbeck, J. P., and F. Ronquist, 2001 MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.
- Huel senbeck, J. P., B. Larget, R. E. Miller and F. Ronquist, 2002 Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51: 673–688.
- Hugal, A., J. Stanton and C. Moritz, 1999 Reticulate evolution and the origins of ribosomal internal transcribed spacer diversity in apomictic Meloidogyne. *Mol. Biol. Evol.* 16: 157–164.
- John, B., and G. M. Hewitt, 1970 Inter-population sex chromosome polymorphism in the grasshopper *Podisma pedestris*—I. Fundamental facts. *Chromosoma* 31: 291–308.
- Jordan, S., C. Simon and D. Polhemus, 2003 Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Caenagrionidae). *Syst. Biol.* 52: 89–109.
- Kanagawa, T., 2003 Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Biosci. Bioeng.* 96: 317–323.
- Kovarik, A., R. Matyasek, K. Y. Lim, K. Skalicka, B. Koukalova et al., 2004 Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana allotetraploids*. *Biol. J. Linn. Soc.* 82: 615–625.
- Larget, B., and D. L. Simon, 1999 Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16: 750–759.
- Lim, K. Y., A. Kovarik, R. Matyasek, M. Bezdek, C. P. Lichtenstein et al., 2000 Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. *Chromosoma* 109: 161–172.
- Litzenberger, G., and W. Chapco, 2001 Molecular phylogeny of selected Eurasian Podismine grasshoppers (Orthoptera: Acrididae). *Ann. Entomol. Soc. Am.* 94: 505–511.
- Márquez, L. M., D. J. Miller, J. B. MacKenzie and M. J. H. van Oppen, 2003 Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral *Acropora*. *Mol. Biol. Evol.* 20: 1077–1086.
- Martin, D. P., C. Williamson and D. Posada, 2005 RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260–262.
- Maynard Smith, J., 1992 Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34: 126–129.
- Mayol, M., and J. A. Rossello, 2001 Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol. Phylogenet. Evol.* 19: 167–176.
- Mighell, A. J., N. R. Smith, P. A. Robinson and A. F. Markham, 2000 Vertebrate pseudogenes. *FEBS Lett.* 468: 109–114.
- Muir, G., C. C. Fleming and C. Schlötterer, 2001 Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol. Biol. Evol.* 18: 112–119.
- Nichols, R. A., and G. M. Hewitt, 1986 Population structure and the shape of a chromosomal cline between two races of *Podisma pedestris* (Orthoptera: Acrididae). *Biol. J. Linn. Soc.* 29: 301–316.
- O'Donnell, K., and E. Cigelnik, 1997 Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* 7: 103–116.
- Parkin, E. J., and R. K. Butlin, 2004 Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Mol. Biol. Evol.* 21: 1595–1601.
- Perreira, S. L., A. L. Baker and A. Wajntal, 2002 Combined nuclear and mitochondrial DNA sequences resolve generic relationships within the Cracidae (Galliformes, Aves). *Syst. Biol.* 51: 946–958.
- Petrov, D. A., 2001 Evolution of genome size: new approaches to an old problem. *Trends Genet.* 17: 23–28.
- Petrov, D. A., E. R. Lozovskaya and D. L. Hartl, 1996 High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384: 346–349.
- Polanco, C., A. I. Gonzalez, A. de la Fuente and G. A. Dover, 1998 Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. *Genetics* 149: 243–256.
- Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
- Posada, D., and K. A. Crandall, 2001 Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* 98: 13757–13762.

- Razafimandimbison, S. G., E. A. Kellogg and B. Bremer, 2004 Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleaceae (Rubiaceae). *Syst. Biol.* 53: 177–192.
- Robertson, H. M., 2000 The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* 10: 192–203.
- Ruggiero, M. V., and G. Procaccini, 2004 The rDNA ITS region in the Lessepsian marine angiosperm *Halophila stipulacea* (Forsk.) Aschers. (Hydrocharitaceae): intragenomic variability and putative pseudogenic sequences. *Mol. Biol. Evol.* 58: 115–121.
- Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schlötterer, C., and D. Tautz, 1994 Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* 4: 777–783.
- Sharpe, R. G., R. E. Harbach and R. Butlin, 2000 Molecular variation and phylogeny of members of the *Minimus* group of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). *Syst. Entomol.* 25: 263–272.
- Swofford, D. L., 2002 PAUP* Version 4.0b10 for Macintosh. Sinauer Associates, Sunderland, MA.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876–4882.
- van Nues, R. W., J. Venema, J. M. J. Rientjes, A. Dirks-Mulder and H. A. Raué, 1995 Processing of eukaryotic pre-rRNA: the role of the transcribed spacers. *Biochem. Cell Biol.* 73: 789–801.
- Waddell, P. J., Y. Cao, J. Hauf and M. Hasegawa, 1999 Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid invariant sites LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *Syst. Biol.* 48: 31–53.
- Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof et al., 1994 Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst. Biol.* 43: 250–261.
- Westerman, M., N. H. Barton and G. M. Hewitt, 1987 Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity* 58: 221–228.
- Whang, I.-J., J. Jung, J.-K. Park, G.-S. Min and W. Kim, 2002 Intragenomic length variation of the ribosomal DNA intergenic spacer in a malaria vector, *Anopheles sinensis*. *Mol. Cells* 14: 158–162.
- Yamada, K., J. Lim, J. M. Dale, H. Chen, P. Shinn et al., 2003 Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* 302: 842–846.
- Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555–556.
- Yang, Z., and D. Roberts, 1995 On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* 12: 451–458.
- Yoder, A. D., and Z. H. Yang, 2000 Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17: 1081–1090.

Communicating editor: G. C. Gibson

Chapitre VII. Discussion générale – Perspectives



Dans les recherches de phylogénie moléculaire chez les Melanoplinae (Insecta: Orthoptera: Acrididae), nous avons rencontré quelques difficultés dans (i) l'amplification des fragments d'ADN ciblés, et (ii) la résolution des analyses phylogénétiques. Les principaux facteurs responsables de ces problèmes seraient la taille du génome et l'ancienneté du groupe choisi comme modèle d'étude.

VII.1. La taille du génome et les pseudogènes chez les criquets (Insecta: Orthoptera: Caelifera)

Chez les acridiens, le génome a une taille importante et sa structure est particulièrement complexe. Le génome du criquet alpin *Podisma pedestris*, par exemple, est un des plus grand génome du monde animal avec 18000 Mb, ce qui représente cent fois la taille du génome de *Drosophila* et quatre fois celui de *Homo sapiens* (Bensasson 2001). Cette grande taille est, en grande partie, due à l'accumulation de séquences particulières, des NUMTs (« NUclear MiTochondrial-like Sequences »), *i.e.* des séquences d'ADN nucléaire ressemblant à des gènes mitochondriaux et qui mutent plus rapidement que les séquences copiées (e.g. Bensasson et al. 2001). Il semble qu'il existe une corrélation inverse entre la taille du génome des criquets et le taux de perte de l'ADN (Bensasson et al. 2001). Le faible taux relatif de perte d'ADN chez les criquets pourrait être expliqué par un potentiel d'accumulation des mutations ponctuelles qui dépasse celui d'accumulation des indels (Bensasson et al. 2001; Keller et al. 2006). Le principal facteur responsable des différences entre les taux de perte d'ADN, entre les criquets, les grillons et *Drosophila* par exemple, semble être la taille de la délétion: chez les Melanoplinae, comme *Podisma* et *Italopodisma*, les délétions de plus d'une base semblent très rares (Bensasson et al. 2001).

Pour éviter l'amplification et le séquençage des NUMTs dans nos travaux de phylogénie, nous avons appliqué les techniques suivantes:

a) avant le séquençage:

nous avons défini des amorces spécifiques pour chaque fragment d'ADN et pour chaque clade identifié par des analyses phylogénétiques préliminaires, et nous avons purifié sur gel la bande ayant la taille (longueur en base nucléotidique) attendue dans le cas d'un produit PCR multi bandes;

b) après le séquençage:

nous avons évalué et comparé le taux de mutations de séquences amplifiées, en sachant que celui des NUMTs est intermédiaire celui des gènes nucléaires et mitochondriales (Bensasson et al. 2001). Nous avons aussi comparé nos séquences avec les pseudogènes déposés dans des "banque de pseudogènes" comme celle initiée par Bensasson (2001) (<http://www.pseudogene.net/>). Pour un même extrait d'ADN, nous avons aussi répété les PCR et le séquençage, pour comparer en suite la constance de la position phylogénétique des séquences provenant d'un même individu, dans les topologies issues d'une même technique d'analyse.

Malheureusement, jusqu'à présent aucune méthode de séparation des gènes mitochondriaux de leurs copies nucléaires ne s'est avérée totalement fiable, fait qui met en doute la fiabilité des méthodes d'identification des espèces sur une séquence unique comme le "DNA barcoding" (e.g. Moulton et al. 2010).

VII.2. L'utilité de la phylogénie moléculaire dans la classification des Melanoplinae

Les recherches phylogénétiques, de parcimonie, maximum de vraisemblance et Bayésienne, que nous avons menées dans le cadre de notre thèse, nous ont permis d'élaborer et argumenter des hypothèses sur l'origine et l'évolution des acridiens analysés, à tous les

niveaux taxinomiques choisis : la sous-famille pour les Melanoplinae, la tribu pour les Podismini et l'espèce pour *Podisma pedestris* (Linnaeus, 1758).

Quelques remarques méthodologiques émergent de notre travail de thèse:

- pour tous les niveaux taxinomiques analysés, l'analyse des données combinées, nucléaires et mitochondriales, améliore la résolution des topologies et le support des noeuds phylogénétiques par rapport aux analyses des fragments séparés ;
- la méthode de maximum de vraisemblance a donné les topologies avec la meilleure résolution ;
- la pondération de caractères moléculaires et le codage des « indels » (Lutzoni et al. 2001) n'améliorent pas nos résultats phylogénétiques.

Le tableau qui s'impose pour l'histoire évolutive moléculaire des unités taxonomiques analysés, i.e. Podismini et Melanoplinae, est celui des groupes polyphylétiques, dépourvus d'un sens évolutif, et donc en désaccord avec les classifications morphologiques en vigueur. Des origines multiples, des dispersions intercontinentales et des diversifications *in situ* sont proposées pour expliquer la genèse et l'évolution des faunes actuelles de Podismini Eurasiatique et des Melanoplinae, en corrélation avec les conditions climatiques, géotectoniques et biotiques du Tertiaire.

VII.3. Quelques perspectives immédiates

Une partie de notre travail de thèse représentée par des résultats phylogénétiques à des niveaux taxonomiques supérieurs à la sous-famille, comme : (1) la famille, i.e. les **Acrididae**, et (2) la super-famille, i.e. les **Acridomorpha** (sensu Dirsh, 1956), demeure inexploitée pour l'instant. Ces résultats constituent la base de deux autres articles en préparation.

Nos résultats préliminaires, partiellement présentés dans le présent manuscrit, s'inscrivent dans la lignée des débats ouverts par le réquisitoire : « plus de taxons ou plus des caractères pour améliorer la résolution des arbres phylogénétiques ? » (voir Zwickl et Hillis 2002, Hillis et al. 2003, Wiens 2005). Sans avoir la prétention de répondre à cette question, nos résultats font pencher la balance en faveur de l'augmentation du nombre de taxons analysés.

De même, nos résultats confortent la constatation que l'ajout des taxons améliore la résolution phylogénétique plus vite que l'ajout de caractères (voir Graybeal 1998, Caterino et al. 2000). Les résultats contradictoires sur la monophylie des Podismini seraient justement un témoin de l'impact que l'échantillonnage pourrait avoir sur la phylogénie.

Les hypothèses phylogénétiques issues de notre travail de thèse orienteraient les recherches moléculaires futures vers un choix des taxons conscient, ciblé, mieux adapté à la problématique soulevée.

De plus, le nombre des taxons à analyser étant un facteur essentiel et donc limitant dans les recherches évolutives, on peut entrevoir aisément l'importance des collections de musées comme source d'information pour les taxons nouveaux, avec un statut taxonomique incertain, rares, difficilement accessibles ou disparus.

De ce point de vue, les séquences d'ADN que nous avons extrait des taxons, souvent « type », des collections du Museum National d'Histoire Naturelle de Paris, trouveront facilement leur utilité future dans les recherches phylogénétiques, à côté des séquences provenant des individus nouvellement récoltés sur le terrain...

Bibliographie sélective

(hors articles)



- Alvarez, I., Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phyl. Evol.* 29 : 417-434.
- Amédégnato, C., 1974. Les genres d'Acridiens neotropicaux, leur classification par familles, sous familles et tribus. *Acrida* 3 : 193-204.
- Amédégnato, C., 1977. Etude des Acridoidea centre et sud americains (Catantopinae sensu lato). Anatomie des génitalia, classification, répartition, phylogénie. Unpublished Thesis, University of Paris VI, 385 pp.
- Amédégnato, C., 1993. African – American relationships in the acridians (Insecta, Orthoptera). In: George, W., Lavocat, R. (Eds.), *The Africa – South America connection*. Clarendon Press, Oxford, pp 59-75.
- Amédégnato, C., Chapco, W., Litzemberger, G., 2003. Out of South America? Additional evidence for a southern origin of melanopline grasshoppers. *Mol. Phylogenet. Evol.* 29: 115-119.
- Bensasson D, Zhang DX, Hewitt GM. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol Biol Evol.* 17: 406-415.
- Bensasson, D., D. A. Petrov, D.-X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol. Biol. Evol.* 18: 246-253.
- Blanchet, E., C. Pages, L. Blondin, Billot C., Rivallan R., Vassal J.M., Lecoq M., Risterucci A.M.. 2010. Isolation of microsatellite markers in the *Calliptamus* genus (Orthoptera, Acrididae). *Journal of Insect Science* 133: 1-6.
- Brunner von Wattenwyl, C. 1893. Révision du système des Orthoptères et description des espèces rapportées par M. Leonardo Fea de Birmanie. *Ann. Mus. Stor. Nat. Genova*, 13: 1-230.

- Cabrero, J., Perfectti, F., Gomez, R., Gamacho, J.P.M., Lopez-Leon, M.D. 2003. Population variation in the A chromosome distribution of satellite DNA and ribosomal DNA in the grasshopper *Eyprepocnemis plorans*. *Chromosome Res.* 11: 375-381.
- Carbonell, C. S., 1977. Origin, evolution, and distribution of the Neotropical acridomorph fauna (Orthoptera): A preliminary hypothesis. *Rev. Soc. Entomol. Argentina* 36, 153-175.
- Carpenter FM. 1992. Superclass Hexapoda. In: Kaesler RL, ed. *Treatise on invertebrate paleontology*. Boulder: The Geological Society of America and the University of Kansas.
- Castresana, J., Feldmaier-Fuchs, G., Yokobori, S., Satoh, N., Pääbo, S. 1998. The Mitochondrial Genome of the Hemichordate *Balanoglossus carnosus* and the Evolution of Deuterostome Mitochondria. *Genetics* 150: 1115-1123.
- Caterino, M.S., Soowon Cho, Sperling, F.A.H. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annu. Rev. Entomol.* 45: 1-54..
- Chapco, W., Litzenberger, G., Kuperus, W.R., 2001. A molecular biogeographic analysis of the relationships between North American melanopline grasshoppers and their Eurasian and South American relatives. *Mol. Phylogenet. Evol.* 18, 460-466.
- Chapco, W. 2006. A note on the phylogenetic position of *Duartettix montanus* within the subfamily Melanoplineae. *J. Orth. Research.* 15: 59-63.
- Chintauan-Marquier, I.C., Jordan, S., Berthier, B., Amédégnato, C., Pompanon, F. 2010. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: the Melanoplineae (Orthoptera: Acrididae). *Molecular Phylogenetics and Evolution*: in press.
- Darlu, P., Tassy, P. 1993. *La reconstruction phylogénétique*. Masson, Paris.
- Darwin, C. 1861. Letter to Henry Fawcett, cité par Steven J. Gould dans "Dinosaurs in the haystack," *Natural History* 3(92):2-13.

- DeBry, R.W., Abele, R.W. 1995. The relationship between the parsimony and maximum-likelihood analyses: tree scores and confidence estimates for three real data sets. *Mol. Biol. Evol.* 12: 291-297.
- Delsuc, F., Douzery, E.J.P. 2004. Les méthodes probabilistes en phylogénie moléculaire: (1) Les modèles d'évolution des séquences et le maximum de vraisemblance. *Biosystema* 22: 59-74.
- Delsuc, F., Douzery, E.J.P. 2004. Les méthodes probabilistes en phylogénie moléculaire: (2) L'approche bayésienne. *Biosystema* 22: 75-86.
- De Ricqlès, A. 1996. Biologie historique et évolutionnisme. Leçon inaugurale, Collège de France.
- Dirsh, V.M. 1956. The phallic complex in Acridoidea (Orthoptera) in relation to taxonomy. *Trans. R. Ent. Soc. Lond.* 108.
- Dirsh, V.M., 1975. Classification of the Acridomorphoid Insects. Clasey Ltd., Faringdon, Oxon.
- Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20, 248-254.
- Douzery, E.J.P., 2002. Les datations en phylogénie moléculaire: de l'ADN et des protéines jusqu'aux fossiles et réciproquement. Dossier d'Habilitation à Diriger les Recherches (H.D.R.). Université Montpellier II.
- Douzery, E.J.P., Blanquart, S., Criscuolo, A., Delsuc, F., Douady, C., Lartillot, N., Philippe, H., Ranwez, V. 2010. "Phylogénie moléculaire". Pp. 183-243 dans «Biologie évolutive»: collection LMD [F. Thomas, T. Lefèvre et M. Raymond édés]. De Boeck, Bruxelles.
- Dowton M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. *Mol. Phylogenet. Evol* 11: 283-287.

- Edwards, A.W.F., Cavalli-Sforza, L.L. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21: 550-570.
- Elder, J.F., Turner, B.J. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70: 297-320.
- Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* 22: 240-249.
- Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer Associates, Inc., Publishers Sunderland, Massachusetts.
- Fitch, W.M., Margoliash, E. 1967. A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. *Biochem. Genet.* 1: 65-71.
- Flanagan, N.S., Mason, P.L., Gosalvez, J., Hewitt, G.M. 1999. Chromosomal differentiation through an alpine hybrid zone in the grasshopper *Chorthippus parallelus*. *J. Evol. Biol.* 12: 577-585.
- Flook, P.K., Rowell, C.H.F., 1997. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Mol. Phyl. Evol.* 8: 177-192.
- Fries, M., Chapco, W., Contreras, D., 2007. A molecular phylogenetic analysis of the Oedipodinae and their intercontinental relationships. *Journal of Orthoptera Research* 16: 115-125.
- Fontana, P.G., V.R. Vickery. 1976. Cytotaxonomic studies on the genus *Boonacris*. I. The "eastern" taxa and a comparison with the related genera *Dendrotettix* and *Appalachia* (Orthoptera: Catantopidae: Podismini). *Can. J. Genet. Cytol.* 18: 625-652.
- Gillespie, J.H. 1984. The molecular clock may be an episodic clock. *Proc. Natl. Acad. Sci. USA* 81: 8009-8013.
- Gorokhov, A.V. 1995. Contribution to the system and evolution of the order Orthoptera. *Zoologicheskii Zhurnal*, Nauka.

- Graur, D., Li, W.H. 2000. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Graybeal, A. 1998. Is better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47: 9-17.
- Guindon, S., Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696-704.
- Harz, K. 1975. *Die orthopteren Europas II*. Dr. W. Junk B.V. Publishers, Hague.
- Hennig, W. 1966. *Phylogenetic Systematics*. University of Illinois Press (Urbana).
- Hickson, R.E., Simon, C., Perrey S.W. 2000. The Performance of Several Multiple-Sequence Alignment Programs in Relation to Secondary-Structure Features for an rRNA Sequence *Mol. Biol. Evol.* 17:530–539.
- Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J. 2003. Is sparse taxon sampling a problem for phylogenetic inference? *Syst. Biol.* 52: 124-126.
- Holder, M.T., Lewis, P.O., Swofford, D. 2010. The Akaike information criterion will not choose the no common mechanism model. *Syst. Biol.* 59: 477-485.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755.
- Jago, N.D., 1979. The breakup of Pangea and prediction of numbers of subfamily elements, living and extinct in Acridoidea. *Proceedings of the 2nd triennial meeting of the Pan American Acridological Society, Montana, USA: 139-162.*
- Keller, I., Chintauan-Marquier, I.C., Veltsos, P., Nichols, R.A. 2006. Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics* 174: 863-871.

- Kevan, SP. 1982. Synopsis and classification of living organisms. Parker - Mcgrow-Hill Book Co.
- Kirby, W.F. 1902. Report on a collection of African Locustidae formed by Mr. W.L. Distant chiefly from the Tansvaal. Trans. Ent. Soc. Lond.: 57-114, 231-241.
- Knowles, L.L., Otte, D. 2000. Phylogenetic analysis of montane grasshoppers from western North America (Genus *Melanoplus*, Acrididae: Melanoplineae). Ann. Entomol. Soc. Am. 93: 421-431.
- Kukalová-Peck J. 1991. Fossil history and the evolution of hexapod structures. In: Naumann ID, ed. The insects of Australia, a textbook for students and researchers, 2nd edn. Melbourne: Melbourne University Press, 141–179.
- Lemmon, A.R., Milinkovitch, M.C. 2002. The metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. Proc. Nat. Acad. Sci. 99: 10516-10521.
- Levenshtein V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10: 707–710.
- Liò, P., Goldman, N. 1999. Using protein structural information in evolutionary inference: transmembrane proteins. Mol. Biol. Evol. 16: 1696-1710.
- Linnaeus, C. 1758. Systema Naturae. 10th ed. 1: 424-433.
- Litzenberger, G. and W. Chapco. 2001. Molecular phylogeny of selected Eurasian podismine grasshoppers (Orthoptera: Acrididae). Ann. Entomol. Soc. Am. 94: 505 -511.
- Litzenberger, G., Chapco, W., 2003. The North American Melanoplineae (Orthoptera: Acrididae): a molecular phylogenetic study of their origins and taxonomic relationships. Ann. Entomol. Soc. Am. 96: 491-497.
- Lutzoni, F., Wagner, P., Reeb, V., Zoller, S., 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49: 628-651.

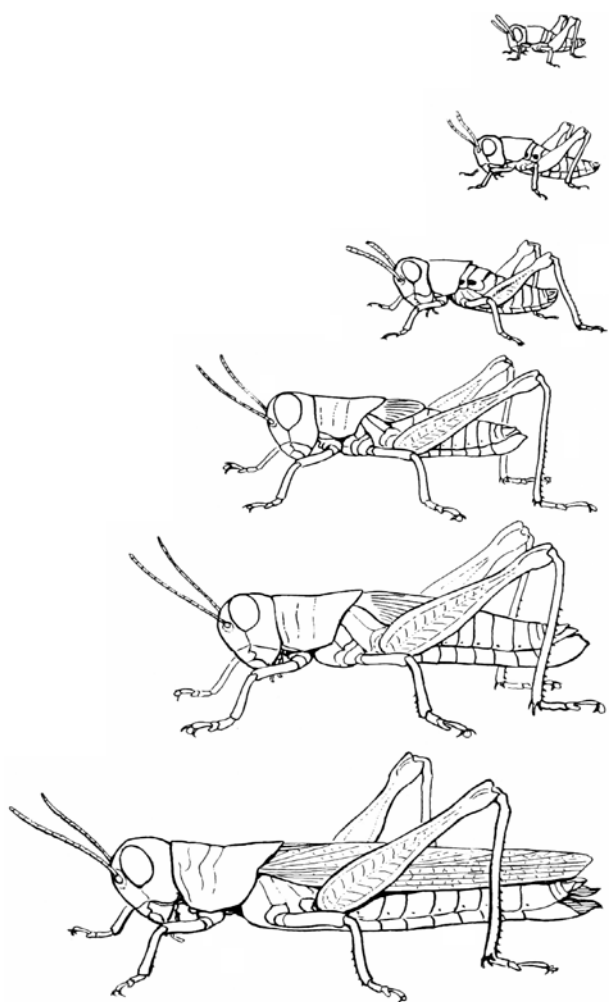
- Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis. Version 1.11, <http://mesquiteproject.org>.
- Moulton MJ, Song H, Whiting MF. 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources* 10: 615–627.
- Mullis, K.B., Faloona, F.A. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Meth. Enzymol.* 155: 335-350.
- Naylor, G.J.P., Brown, W.M. 1998. Amphioxus mitochondrial DNA, Chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47: 61-76.
- Nei, M., Kumar, S. 2000. *Molecular phylogenetics and evolution*. Oxford University Press.
- O’Hara, R.J. 1997. Population thinking and tree thinking in systematics. *Zool. Scr.* 26: 323-329.
- Otte, D., 1995. Orthoptera species file 4. The Orthopterists’ Society and the Academy of Natural Sciences of Philadelphia, Philadelphia.
- Page, R.D.M, Holmes, E.C. 2001. *Molecular evolution: a phylogenetic approach*. Blackwell Science Ed.
- Peel AD, Telford MJ, Akam M. 2006. The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. Proc Biol Sci. 273:1733-1742.
- Petrov, D. A., E. R. Lozovskaya, and D. L. Hartl. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384:346-349.
- Philippe H., Douzery E. 1994. The pitfalls of molecular phylogeny based on four species as illustrated by the Cetacea/Artiodactyla relationships. *Journal of Mammalian Evolution* 2: 133-152.
- Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

- Rehn, J.A.G., Randell, R.L., 1963. A preliminary analysis of the lines of the supertribe Melanoplina Orthoptera; Acrididae; Cyrtacanthacridinae). Proc. Acad. Natl. Sci. (Phila.) 115, 1-32.
- Roberts, H. R. 1941. A comparative study of the subfamilies of the Acrididae (Orthoptera) primarily on the basis of their phallic structures. Proceedings Academy Natural Sciences Philadelphia 93:201–246.
- Rokas, A., J. Kathirithamby and P. W. H. Holland (1999). Intron insertion as a phylogenetic character: the engrailed homeobox of Strepsiptera does not indicate affinity with Diptera. Insect Mol. Biol. 8: 527-530.
- Rokas, A. and P. W. H. Holland (2000). Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15: 454-459.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-4.
- Ronquist, F., Deans A.R.. 2010. Bayesian phylogenetics and its influence on insect systematics. Annu. Rev. Entomol. 55: 189-206.
- Rowell, C.H.F., Carbonell, C.S., 1977. *Baeacris talamancensis* gen. and sp. Nov. (Acrididae, Melanoplinae), a Neotropical montane grasshopper; its implications for the origin of the Dichroplini and the Costa Rican paramo. *Acrida* 6, 55-74.
- Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, F., Crozier, R.H. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55: 421-438.
- Sharov. 1968. Trudy Paleontologicheskogo Instituta 118:19, 152.
- Sharov AG. 1971. Phylogeny of the Orthopteroidea. Jerusalem: Israel Program for Scientific Translations.

- Spaethe J, Briscoe AD. 2004. Early duplication and functional diversification of the opsin gene family in insects. Mol Biol Evol. 21:1583-1594.
- Storozhenko, S.Y. 1997. Classification of order Grylloblattida (Insecta), with description of new taxa. J. Rus. Entomol. Soc. 42: 1-20.
- Swofford, D.L., 1995. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic Inference. In Hillis DM, Moritz D, and Mable BK, editors, Molecular Systematics, pp. 407-514. Sinauer Associates, Sunderland, Massachusetts.
- Teruel M, Cabrero J, Perfectti F, Camacho JP. 2010. B chromosome ancestry revealed by histone genes in the migratory locust. Chromosoma 119:217-225.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876-4882.
- Thorne, J.L., Kishino, H., 2002. Divergence time estimation and rate evolution with multilocus data sets. Syst. Biol. 51: 689-702.
- Torrusio, S., Cigliano M.M., De Wysiecki M.L. 2002. Grasshopper (Orthoptera: Acridoidea) and plant community relationships in the Argentine pampas. Journal of Biogeography 29: 221-229.
- Vickery, V.R., 1987. The northern Nearctic Orthoptera: their origins and survival. In Bacetti, B.C. (Ed.), Evolutionary Biology of Orthopteroid Insects. Ellis Horwood, West Sussex, England, pp 581-591.
- Vickery, V.R., 1997. Classification of Orthoptera (sensu stricto) or Caelifera. In Gangwere, S.K., Muralirangan, M.C., Muralirangan, M. (Eds.), The bionomics of grasshoppers, katydids and their kin. CAB International, New York, pp 5-40.

- Walter, D.E., Winterton, S. 2007. Keys and the crisis in taxonomy: extinction or reinvention? *Annu. Rev. Entomol.* 52:193-208.
- Westerman, M., N. H. Barton, and G. M. Hewitt. 1987. Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity* 58:221-228.
- Wiens, J.J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst. Biol.* 54: 731-742.
- Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annu. Rev. Entomol.* 53: 449-472.
- Zeuner, F.E. 1939. Fossil Orthoptera Ensifera. British Museum of Natural History.
- Zwickl, D.J., D.M. Hillis. 2002. Increased taxon sampling greatly reduced phylogenetic error. *Syst. Biol.* 51: 588-598.
- Zuckerandl, E., Pauling, L. 1965. Evolutionary divergence and convergence in proteins. In V. Bryson and H.J. Vogel (eds.). *Evolving genes and proteins*. Academic Press, New York: 97-166.

Annexes



Annexe 1. Suppléments de l'article II.2.C. - Chintauan-Marquier, I.C., Jordan, S., Berthier, B., Amédégato, C., Pompanon, F. 2010. Evolutionary history and taxonomy of a short-horned grasshopper subfamily: the Melanoplineae (Orthoptera: Acrididae). *Molecular Phylogenetics and Evolution*: in press.

TABLE S1. Oligonucleotide primers used. ^a The primers' positions make reference to the *Drosophila melanogaster* rDNA complete sequence (GenBank accession M21017) for the nuclear sequences, and to the *Drosophila yakuba* entire genome sequence (GenBank accession XO3240) for the mitochondrial genes.

Primer name	Sequence (5' – 3')	Position ^a	Source	Annealing temperature
18S fragment 18S				
18SP2	TTCTTGATTCGGTGGGTG	1337	Flook and Rowel, 1998	54°C – 56°C
18SAF1	TTCGGTGGGTGGTGGTGCAT	1344	This study	45°C – 55°C
srDNA-6P	TAATGATCCTTCCGCAGGTTCA	1995	Vawter, 1991	54°C – 56°C
18SAR1	GATCCTCCGCAGGTTACCTA	1991	This study	45°C – 55°C
ITS1 fragment				
ITS1A	GTCGTAACAAGGTTTCCGTAGG	1952	Kuperus and Chapco, 1994	43°C – 54°C
ITS1FB	GTAACAAGGTTTCCGTAGGTGA	1955	This study	44°C – 62°C
ITS1R	TGCTGCGTTCATCGACCCA	2771	This study	47°C – 62°C
ITS1RB	TTGCTGCGTTCATCGACCC	2772	This study	47°C – 62°C
ITS1B	TTTGTGCGTTCATCGACC	2773	Kuperus and Chapco, 1994	43°C – 54°C
28S fragment				
28S-01	GACTACCCCTGAATTTAAGCAT	3313	Kim et al., 2000	45°C – 53°C
28SF1	CCCCTCTGAACCACGGAGTCTAACAT	4048	This study	48°C – 58°C
28SR-01	GACTCCTTGTCGGTGTTC AAG	4075	Kim et al., 2000	48°C – 57°C
28SR1	ATCCTGAGGGAAACTTCGGAAGGAACCAGC	4428	This study	48°C – 58°C
12S fragment				
SR-J-14233 (12Sbi)	AAGAGCGACGGGCGATGTGT	14233	Simon et al., 1994	45°C – 50°C
12SF1	AGCGACGGGCGATGTGTACAC	14236	This study	45°C – 59°C
SR-N-14588 (12Sai)	AAACTAGGATTAGATACCCTATTAT	14588	Simon et al., 1994	45°C – 50°C
12SR1	CTAGGATTAGATACCCTATTATT	14591	This study	45°C – 59°C

TABLE S2. Genera geographical distribution matrix used in DIVA: 1- presence, 0 – absence.

Genus	Ethiopian region ER	Western Palaeartic WP	Eastern Palaeartic EP	Western Nearctic WN	Eastern Nearctic EN	Northern South-America NSA	Southern South-America SSA
<i>Prionotropis</i>	0	1	0	0	0	0	0
<i>Maculiparia</i>	0	0	0	0	0	1	0
<i>Marellia</i>	0	0	0	0	0	1	1
<i>Oxycatanotops</i>	1	0	0	0	0	0	0
<i>Pezotettix</i>	0	1	0	0	0	0	0
<i>Copiocera</i>	0	0	0	0	0	1	1
<i>Coscineuta</i>	0	0	0	0	0	1	0
<i>Italopodisma</i>	0	1	0	0	0	0	0
<i>Podisma</i>	0	1	0	0	0	0	0
<i>Sinopodisma</i>	0	0	1	0	0	0	0
<i>Aidemona</i>	0	0	0	1	0	1	0
<i>Melanoplus</i>	0	1	1	1	1	0	0
<i>Netrosoma</i>	0	0	0	1	0	1	0
<i>Phaulotettix</i>	0	0	0	1	0	1	0
<i>Dactylotum</i>	0	0	0	1	1	1	0
<i>Hesperotettix</i>	0	0	0	1	1	0	0
<i>MNHN Pe01-18</i>	0	0	0	0	0	1	0
<i>Chibchacris</i>	0	0	0	0	0	1	0
<i>Dichroplus</i>	0	0	0	0	0	0	1
<i>Hazelacris</i>	0	0	0	0	0	1	0
<i>Neopedies</i>	0	0	0	0	0	0	1
<i>Scotussa</i>	0	0	0	0	0	0	1
<i>Meridacris</i>	0	0	0	0	0	1	0
<i>ME86-06</i>	0	0	0	0	0	1	0
<i>Pediella</i>	0	0	0	0	0	1	0
<i>Propedies</i>	0	0	0	0	0	1	0

Annexe 2. Suppléments de l'article II.3.C. - Chintauan-Marquier, I.C., Amédégato C., Pompanon F. *en préparation*. New Molecular Evidences for the Evolutionary History of Eurasian Podismines (Orthoptera: Acrididae: Melanoplinae).

Table S1. Oligonucleotide primers used in this study. a The positions of the nuclear primers make reference to the *Drosophila melanogaster* rDNA complet sequence from GenBank (accession M21017; Tautz et al., 1988). The mitochondrial primers have been named with reference to the *Drosophila yakuba* entire genome sequence from GenBank (accession XO3240; Clary and Wolstenholme, 1985) as suggested by Simon et al. (1994).

Primer name	Sequence (5' - 3')	Position ^a	Source	Annealing temperature
Fragment ITS1 (500 bp)				
ITS1A	GTCGTAACAAGGTTTCCGTAGG	1952	Kuperus and Chapco, 1994	43°C - 54°C
ITS1FB	GTAACAAGGTTTCCGTAGGTGA	1955	This study	44°C - 62°C
ITS1R	TGCTGCGTTCTTCATCGACCCA	2771	This study	47°C - 62°C
ITS1RB	TTGCTGCGTTCTTCATCGACCC	2772	This study	47°C - 62°C
ITS1B	TTTGCTGCGTTCTTCATCGACC	2773	Kuperus and Chapco, 1994	43°C - 54°C
Fragment COII (400 bp)				
C2-J-3279 (A-171)	GGTCAAACAATTGAGTCTATTTGAAC	3254	Simon et al., 1994	43°C - 54°C
C2-N-3661 (Barbara)	CCACAAATTTCTGAACATTGACCA	3684	Simon et al., 1994	45°C - 59°C
Fragment 16S (500 bp)				
LR-J-12887 (16Sbr, 16Sb)	CCGGTCTGAACTCAGATCACGT	12866	Simon et al., 1994	45°C - 50°C
LR-N-13398 (16Sar)	CGCCTGTTTAAACAAAACAT	13417	Simon et al., 1994	45°C - 59°C
Fragment 12S (350 bp)				
SR-J-14233 (12Sbi)	AAGAGCGACGGGCGATGTGT	14233	Simon et al., 1994	45°C - 50°C
12SF1	AGCGACGGGCGATGTGTACAC	14236	This study	45°C - 59°C
SR-N-14588 (12Sai)	AAACTAGGATTAGATACCCTATTAT	14588	Simon et al., 1994	45°C - 50°C
12SR1	CTAGGATTAGATACCCTATTATT	14591	This study	45°C - 59°C

Table S2. Results of the Shimodaira-Hasegawa test with RELL bootstraps applied to the maximum parsimony (MP), maximum likelihood (ML) and Bayesian (B) topologies, using the combined data set and the model selected by the AIC.

Tree	-lnL	Diff	-ln L	P
MP	8222.64	6.99	0.38	
ML	8215.65	best		
B	8233.45	17.80	0.11	

Table S3. The present geographical distributions of the selected genera of Podismini (Table 1) used in DIVA.

Taxon	Geographical distribution			
	Africa	Europe	Asia	North America
<i>Anapodisma</i>	0	0	1	0
<i>Capraiuscola</i>	0	1	0	0
<i>Chortopodisma</i>	0	1	0	0
<i>Cophopodisma</i>	0	1	0	0
<i>Epipodisma</i>	0	1	0	0
<i>Fruhstorferiola</i>	0	0	1	0
<i>Italopodisma_c</i>	0	1	0	0
<i>Italopodisma_f</i>	0	1	0	0
<i>Micropodisma</i>	0	1	1	0
<i>Nadigella</i>	0	1	0	0
<i>Odontopodisma_c</i>	0	1	0	0
<i>Odontopodisma_r</i>	0	1	0	0
<i>Ognevia</i>	0	0	1	0
<i>Parapodisma_d</i>	0	0	1	0
<i>Parapodisma_m</i>	0	0	1	0
<i>Parapodisma_s</i>	0	0	1	0
<i>Peripodisma</i>	0	1	0	0
<i>Podisma_k</i>	0	1	1	1
<i>Podisma_p</i>	0	1	1	1
<i>Prumna</i>	0	0	1	0
<i>Pseudopodisma</i>	0	1	1	0
<i>Pseudoprumna</i>	0	1	0	0
<i>Sinopodisma</i>	0	0	1	0
<i>Tonkinacris</i>	0	0	1	0
<i>Zubovskya</i>	0	1	1	0
<i>Dactyloptum</i>	0	0	0	1
<i>Melanoplus</i>	0	1	1	1
<i>Oxycatantops</i>	1	0	0	0
<i>Pezotettix</i>	1	1	1	0
<i>Phaulotettix</i>	0	0	0	1
<i>Hesperotettix</i>	0	0	0	1

