

**QUAND LA PHYLOGÉOGRAPHIE
RENCONTRE L'HISTOIRE DES GÈNES :
L'ANALYSE CLADISTIQUE EMBOÎTÉE (*NESTED CLADE ANALYSIS*)
ET L'INFÉRENCE DE PROCESSUS POPULATIONNELS
RÉCURRENTS ET ÉPISODIQUES.**

Jesús GÓMEZ-ZURITA

Molecular Systematics Laboratory, Department of Entomology
The Natural History Museum
Cromwell Road, London SW7 5BD, UK
e-mail : j.gomez-zurita@nhm.ac.uk

Abstract. — Phylogeography deals with the study of the spatial distribution of genetic diversity on a phylogenetic framework, i.e. using the information in the space and time axes. Two major groups of processes are considered to shape the geographic distribution of genetic variation for a given set of populations or closely related species. Firstly, recurrent or contemporary processes such as restricted gene flow are a major force determining population structure, and the patterns observed are believed to arise from the interplay of migration and genetic drift. Secondly, episodic or historical processes, including fragmentation of populations, past demographic events such as contiguous or long distance range expansions can also play an important role in the extant geographic distribution of genes in natural populations. The traditional and widely accepted approach for the analysis of population genetic structure is based upon Wright's *F* and derived statistics. However, since these methods do not use genealogical information they are unable to deduce historical processes from the data. This type of historical information has to be obtained from the evolutionary history of the studied genes tracking that of the populations bearing them within certain limits of confidence. When a gene genealogy is available, its overlay on the geography is the traditional tool used in phylogeography for historical inference, though this approach is very limited. Nested geographical cladistic analysis was designed to bridge the gap between these two methodologies, implementing the information used in phylogeography (genealogical structure and geographical distribution) but also incorporating predictions from

coalescent theory based on allele frequencies. The combination of these different sources of information makes it possible to recognise patterns derived from population structure and history, resulting in a promising tool that is becoming increasingly used in phylogeographic analyses. Here I will illustrate the strengths and limitations of this methodology in a case-study using mtDNA data from a sample of 157 specimens of a species complex in the leaf beetle genus *Timarcha*, comparing it with the type of information obtained from the same set of sequences using conventional approaches.

INTRODUCTION

En biologie évolutive moderne, un des champs d'étude les plus féconds est l'analyse de la distribution géographique de la variabilité génétique dans les populations naturelles, et l'explication des causes qui conduisent à une corrélation particulière entre leur structure génétique et leur arrangement spatial. L'étude de la subdivision géographique des populations présente un intérêt indiscutable notamment pour les biologistes de l'évolution et l'étude de la spéciation, pour les biogéographes, et pour les généticiens de la conservation.

Pour mener l'étude de ce type d'associations, une très grande variété de méthodes a été développée dont l'essor a été conditionné par la disponibilité de nouvelles sources de données moléculaires et le

progrès des fondements théoriques. À l'origine, les études basées sur l'utilisation des allozymes, qui ne prenaient pas en compte les relations évolutives entre les allèles observés, nécessitaient l'utilisation des approches traditionnelles de la génétique des populations dans lesquelles les estimations du flux génique sont déduites des données de fréquences d'allèles dans les localités géographiques échantillonnées (FELSENSTEIN, 1982 ; SLATKIN, 1985, 1987). Les techniques de RFLP, et plus récemment les méthodes d'amplification (PCR) et de séquençage d'ADN, associées au progrès rapide des algorithmes de reconstruction phylogénétique, ont permis l'utilisation de l'information généalogique dans les études de distribution spatiale de la variation génétique. L'étude de la distribution géographique d'allèles avec prise en compte de la dimension temporelle, implicite dans les phylogénies d'allèles, a conduit à l'introduction formelle du terme phylogéographie et de la discipline correspondante (AVISE *et al.*, 1987). La phylogéographie a été définie comme la discipline traitant des principes et des processus contrôlant la distribution géographique des lignées généalogiques, et en particulier ceux mis en jeu au sein d'une espèce ou entre des espèces étroitement apparentées (AVISE, 2000). L'approche traditionnelle en phylogéographie consiste à superposer les généalogies de gènes du groupe de populations analysé sur la géographie. Cela permet de mettre en évidence l'association de clades particuliers avec des localisations géographiques spécifiques et de formuler des hypothèses sur la cause de ces associations. Toutefois, en plus de cette approche directe, des méthodologies plus puissantes ont été conçues qui permettent d'obtenir une information plus détaillée à partir à la fois des variables géographiques et des variables génétiques et généalogiques. Parmi ces méthodes, l'une des plus prometteuses est l'analyse cladistique emboîtée (NCA, *nested cladistic analysis*) de TEMPLETON *et al.* (1995) que j'appellerai ici analyse cladistique géographique emboîtée (NGCA, *nested geographic cladistic analysis*) afin de souligner le fait qu'elle s'applique à des données géographiques et pas seulement phylogénétiques.

Dans cet article, je vais argumenter en faveur de l'utilisation de la NCGA en phylogéographie en décrivant ses points forts par comparaison avec les approches traditionnelles. Certaines limites de cette méthodologie seront également discutées. Enfin, l'utilisation et l'application de la méthode sera illustrée par une réanalyse de données d'ADN mitochondrial dans le cadre de l'étude de la phylogéographie d'un complexe d'espèces de chrysomèles (GÓMEZ-ZURITA *et al.*, 2000a).

Méthodes hiérarchiques et généalogiques en phylogéographie : applications et limites

Comme indiqué plus haut, un objectif essentiel de la phylogéographie est de déterminer s'il existe une corrélation entre la dispersion géographique de populations (ou d'espèces apparentées) et leurs différentes structures génétiques. En général, on peut distinguer deux groupes majeurs de facteurs déterminant ce type d'associations : les processus actuels et récurrents, et les processus historiques, épisodiques ou non-récurrents (fig. 1). Les processus récurrents comprennent essentiellement la structure de la population et le rôle joué par le flux génique et la dérive génétique. Par ailleurs, les deux types principaux de processus historiques déterminant les associations géographiques d'haplotypes sont la fragmentation des populations et les mouvements de populations (ces derniers incluant l'expansion de l'aire de distribution et la colonisation à longue distance). Les approches qui étudient les processus populationnels récurrents et épisodiques peuvent être désignées respectivement comme les méthodes hiérarchiques et généalogiques. Toutefois, la distinction n'est pas absolue, et des méthodes généalogiques ont été également mises en œuvre pour l'étude de la structure des populations et du flux génique (par ex. SLATKIN & MADDISON, 1989).

Les méthodes hiérarchiques pour l'analyse de la structure des populations sont fondées sur l'analyse des différences génétiques parmi des sous-ensembles définis *a priori* de l'échantillon global (appelés populations). Les mesures standard utilisées sont les statistiques *F* de WRIGHT (1931) pour

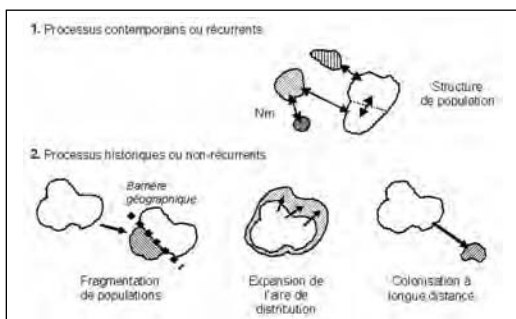


Figure 1. Hypothèses explicatives alternatives pour l'association entre la géographie et la composition génétique des populations naturelles. Les différentes trames montrent les différentes constitutions génétiques des populations et les flèches représentent la dispersion des individus.

données appariées, basées sur le modèle en îles, et toutes les statistiques qui en dérivent ; elles-mêmes basées sur différents modèles, comme le modèle à pas japonais (« stepping stone ») de KIMURA (1953), ou d'autres indices apparentés adaptés à l'utilisation du RFLP ou des données de séquences moléculaires (par ex., NEI, 1982 ; LYNCH & CREASE, 1990 ; HUDSON *et al.*, 1992a).

L'inconvénient majeur de ces méthodes quand on les applique aux études biogéographiques est qu'elles n'intègrent aucune information temporelle, et donc qu'elles ne peuvent séparer les événements passés des contemporains. Ces méthodes de génétique des populations n'utilisent que des patrons spatiaux de distribution allélique et font la supposition que les structures observées sont le résultat de processus tels que la migration et la dérive génétique.

Indépendamment du manque de puissance de ces approches pour rendre compte des changements temporels de flux génique, elles ont des limites très bien connues et doivent donc être utilisées avec précaution dans les études phylogéographiques (BOSSART & PROWELL, 1998). La première critique tient au fait que chacun des modèles permettant d'estimer le nombre de migrants par génération (Nm) comporte des postulats sous-jacents qui ne sont pas nécessairement satisfaits par l'ensemble des popu-

lations étudiées, et généralement le modèle approprié de flux génique n'est pas connu. De plus, ces méthodes moyennent les estimations de flux génique pour l'ensemble de l'échantillon et ne tiennent pas compte de l'hétérogénéité des modèles, ou des paramètres des modèles, applicables dans différentes zones de l'aire de distribution des populations. Un autre problème est que, étant donnée la relation non-linéaire entre Nm et la statistique F_{ST} (ou de son équivalent) dans la gamme inférieure à 0,1 (c'est-à-dire dans le cas où le flux génique est élevé) peuvent représenter des différences dans le nombre de migrants allant de quelques individus à plusieurs millions (TEMPLETON, 1998). Enfin, ces statistiques sont des estimateurs de Nm particulièrement dépourvus de sens biologique quand la différenciation génétique et le flux de gènes ne sont pas reliés, par exemple en cas de colonisation de zones géographiques nouvelles, quand les populations partagent un ancêtre commun récent ou quand elles ont conservé des allèles ancestraux. Dans tous ces cas, les valeurs de F_{ST} estimées sont inférieures à 1, impliquant un échange génétique entre les populations, même en absence complète de flux génique (SLATKIN & MADDISON, 1989 ; TEMPLETON, 1998).

Les méthodes généalogiques utilisent l'information contenue dans les phylogénies d'allèles ou d'haplotypes. Dans le contexte de la présente discussion, elles peuvent être utilisées pour répondre à des questions similaires concernant la structure des populations en relation avec le flux génique (SLATKIN & MADDISON, 1989). Toutefois, en tant qu'outil pour étudier les événements épisodiques en biologie des populations, elles peuvent aussi être appliquées pour inférer des changements démographiques survenus historiquement dans les populations (analyses graphiques de lignées au cours du temps ; NEE *et al.*, 1996), ou pour des analyses phylogéographiques formelles, qui superposent des phylogénies sur la géographie (AVISE, 2000).

Les méthodes et les estimateurs incluant la structure généalogique sont généralement beaucoup plus efficaces que les statistiques pour données appariées

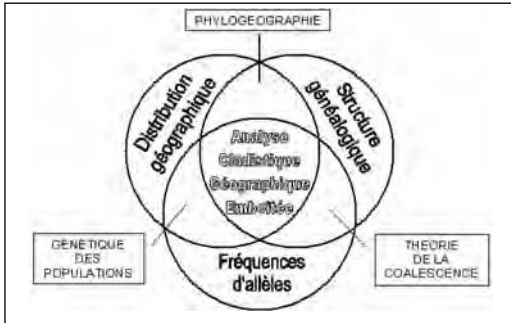


Figure 2. La nature synthétique de l'analyse cladistique emboîtée incorporant des éléments d'analyse génétique des populations traditionnelle, de phylogéographie et de théorie de la coalescence.

utilisées dans les méthodes hiérarchiques (FELSENSTEIN, 1992). Néanmoins, elles posent de nouveaux problèmes quand on les applique à la biogéographie. D'abord, l'information obtenue à partir de la généalogie d'un seul gène peut être trompeuse. L'introgression mitochondriale ou des différences dans les taux d'échantillonnage des lignées (lineage sorting rates) peuvent entraîner une discordance entre l'arbre du gène et l'arbre des espèces ou des populations sous-jacent, de telle sorte que des estimations fiables peuvent nécessiter des données de nombreux locus (BROWER *et al.*, 1996 et références incluses). Un autre problème avec l'approche classique en phylogéographie, qui consiste à superposer la structure généalogique sur la distribution géographique de la variation génétique, est que bien qu'il soit très intuitif de trouver des associations entre les deux, elles ne fournissent pas d'information sur les causes de l'association (TEMPLETON, 1998). Dans tous les cas cette approche devrait être utilisée comme un outil pour formuler des hypothèses, mais pas comme une méthode d'inférence, parce qu'elle conduit généralement à des inférences implicites et subjectives (TEMPLETON, 1998). Enfin, il n'y a pas d'évaluation de l'adéquation des tailles d'échantillons pour discriminer entre les hypothèses alternatives (TEMPLETON, 1998).

L'approche synthétique d'analyse cladistique géographique emboîtée (NGCA)

La NGCA a été développée avec l'idée de surmonter la plupart des limites, déjà mentionnées, des approches traditionnelles pour l'étude de la phylogéographie. Dans ce but, elle combine toutes les sources d'information capables d'expliquer la structuration géographique de la diversité génétique, notamment les données de fréquences alléliques, la distribution géographique et la structure généalogique des données (fig. 2). Pour cela, elle incorpore des éléments de génétique des populations et d'analyse phylogéographique avec les prédictions de la théorie de la coalescence. L'utilisation simultanée de ces différentes variables, dans un cadre où les généalogies de gènes sont considérées comme un ensemble emboîté d'haplotypes, permet de tester différentes hypothèses de structure et d'histoire des populations.

TEMPLETON *et al.* (1995) et TEMPLETON (1998) ont décrit en détails la logique et les procédures pour la NGCA. Je vais simplement résumer ici les principales étapes de l'analyse.

Dans un premier temps, on rassemble des données génétiques qualitatives (RFLP ou séquences) pour inférer l'histoire évolutive du groupe d'individus analysé, qui, dans une étude phylogéographique, comporte typiquement des populations d'une même espèce ou de plusieurs espèces étroitement apparentées. A ce niveau de divergence génétique, quand des haplotypes ancestraux et dérivés coexistent dans l'échantillon et que ces haplotypes sont séparés seulement par quelques mutations, les méthodes phylogénétiques telles que le maximum de parcimonie ou le maximum de vraisemblance, adaptées à la résolution de relations dichotomiques, éprouvent des difficultés à déterminer leurs relations de parenté (HUELSENBECK & HILLIS, 1993 ; CRANDALL, 1994). Les méthodes de reconstruction de réseaux sont alors recommandées, et TEMPLETON *et al.* (1992) ont développé un algorithme statistique de parcimonie permettant d'obtenir les réseaux d'haplotypes ou d'allèles en accord avec les limites de la parcimonie pour un ensemble particulier d'haplotypes. Cet

algorithme est mis en œuvre dans le programme «TCS : estimating gene phylogenies» *version* 1.13 (CLEMENT *et al.*, 2000) (http://bioag.byu.edu/zoology/crandall_lab/tcs.htm).

Une fois que le réseau de parcimonie maximale est construit, sa topologie est utilisée pour définir une série emboîtée d'haplotypes étroitement apparentés en suivant un ensemble de règles spécifiques d'emboîtement (TEMPLETON *et al.*, 1987 ; TEMPLETON & SING, 1993 ; CRANDALL, 1996). Ce système emboîté détermine la structure des tests ultérieurs d'association entre diversité génétique et distribution géographique. Ces tests sont basés sur des analyses de contingence avec permutations (ROFF & BENTZEN, 1989) de deux manières différentes : (i) tester la corrélation entre les haplotypes d'un groupe donné et les localités dans lesquelles se trouvent ces haplotypes, et (ii) tester la déviation entre les distances géographiques réelles entre sites d'échantillonnage et une situation de non-association géographique des haplotypes. Pour le second test, deux types de distances géographiques sont définis dans le contexte du système emboîté (TEMPLETON *et al.*, 1995). Le premier type est la distance de clade (clade distance, D_c), c'est-à-dire la distance géographique moyenne entre les individus portant les haplotypes dans un groupe donné et le centre géographique de ce groupe particulier. Le second est la distance de clade emboîtée (nested clade distance, D_n), c'est-à-dire la distance géographique moyenne entre les individus portant les haplotypes dans un groupe particulier et le centre géographique du groupe de niveau supérieur qui le contient. Ces distances, et leur différence ($I-T$) entre les clades intérieur et apical (« tip »), sont calculées pour chaque groupe et ces différences sont comparées avec une situation de panmixie simulée par permutation au hasard des données dans les matrices de contingence d'occurrence d'haplotypes dans chaque localité. En utilisant cette procédure, il est possible de voir si D_c , D_n , $(I-T) D_c$ et/ou $(I-T) D_n$ sont significativement plus grands ou plus petits qu'attendu sous l'hypothèse d'une non association géographique des haplotypes. Les résultats sont ensuite interprétés à la lumière des prédictions

basées sur différents modèles de structure et d'histoire des populations, ces derniers incluant les fragmentations passées, les expansions d'aire de répartition géographique des populations ou les colonisations à longue distance (TEMPLETON *et al.*, 1995 ; TEMPLETON, 1998). Les tests de contingence par permutation de cette sorte peuvent être faits avec le logiciel «GeoDis» *version* 2.0 (POSADA *et al.*, 2000) (http://bioag.byu.edu/zoology/crandall_lab/geodis.htm).

Les bases de l'analyse cladistique emboîtée ont été posées il y a plus de dix ans (TEMPLETON *et al.*, 1987) et son application aux analyses phylogéographiques a été proposée assez rapidement (TEMPLETON *et al.*, 1995). Toutefois, sa popularité ne s'est développée qu'avec la sortie du logiciel permettant d'obtenir les réseaux d'haplotypes et de faire les tests de contingence avec permutation. Après l'article séminal de TEMPLETON *et al.* (1995) élaborant et utilisant la NGCA, environ 30 articles utilisant cette méthodologie ont été publiés, les deux tiers dans les seules deux dernières années (tableau 1). Ces études phylogéographiques couvrent une grande variété d'organismes et de scénarios géographiques, la plupart d'entre elles utilisent des généalogies d'un seul gène mitochondrial et concernent le niveau intraspécifique.

Points forts et limites de l'analyse cladistique emboîtée

La principale force de la NGCA est que les inférences phylogéographiques sont obtenues de manière objective en testant différentes hypothèses de processus historiques particuliers ayant pu modéliser la structure génétique des populations. La structure des tests statistiques est strictement liée à la topologie de la généalogie de gènes et les inférences sont basées sur des prédictions explicites de la théorie de la coalescence et de la dynamique des populations, enlevant toute subjectivité à l'analyse. De plus, elle offre des avantages en regard de certains problèmes spécifiques inhérents aux autres méthodes hiérarchiques et généalogiques. Elle permet une plus grande précision dans l'évaluation du flux génique en détectant des situations particulières de restriction de flux

Taxonomie	Extension géographique	Type de données	Références
Plantes	Floride + Bahamas	RFLP + analyse hétéroduplex de 3 régions ADNcp	Maskas & Cruzan, 2000
Champignons	Mondiale	séquences IGS1	James <i>et al.</i> , 2001
Protozoaires	Mondiale séquences Petites Antilles + Porto Rico	IGS, EF1 α , calmoduline, CHA, 44.1 séquences Cytochrome <i>b</i>	Carbone & Kohn, 2001 Perkins, 2001
Rotifères	Péninsule Ibérique	séquences COI	Gómez <i>et al.</i> , 2000
Poissons	Canada	séquences D-loop + ATPase 6 + ND1	Turgeon & Bematchez, 2001
	Nord Est USA	RFLP ADNmt	Wilson, 1997
	Sud-Ouest USA	séquences Cytochrome <i>b</i>	Johnson & Jordan, 2000
	Péninsule Balkanique	séquences Cytochrome <i>b</i>	Durand <i>et al.</i> , 1999
	Europe	séquences Cytochrome <i>b</i>	Durand <i>et al.</i> , 2000
	Scandinavie	séquences D-loop	Refseth <i>et al.</i> , 1998
	Europe	séquences D-loop	Nesbø <i>et al.</i> , 1999
	Amérique du Sud	séquences ATPase 6 + 8	Sivasundar <i>et al.</i> , 2001
	Europe	séquences RFLP ADNmt + D-loop	Bematchez, 2001
	Atlantique Nord	séquences Cytochrome <i>b</i>	Nesbø <i>et al.</i> , 2000
Amphibiens	centre USA	mtDNA RFLP	Templeton <i>et al.</i> , 1995
	Nord Ouest USA	séquences Cytochrome <i>b</i> + ND2	Nielson <i>et al.</i> , 2001
	Ouest USA	séquences Cytochrome <i>b</i>	Bos & Sites, 2001
Reptiles	Taiwan	séquences Cytochrome <i>b</i>	Creer <i>et al.</i> , 2001
Mammifères	Afrique de l'Est	RFLP ADNm	Templeton & Georgiadis, 1996
	Europe	séquences Cytochrome <i>b</i> + D-loop	Seddon <i>et al.</i> , 2001
	USA	RFLP ADNmt	Templeton, 1998
	Mondiale	séquences chromosome Y non-recombinantes	Karafet <i>et al.</i> , 1999
	Mondiale	SNPs chromosome Y	Hammer <i>et al.</i> , 1998
	Afrique de l'Est	RFLP ADNmt	Templeton & Georgiadis, 1996
Mollusques	Centre USA	séquences ADNr 16S	Turner <i>et al.</i> , 2000
Arthropodes	Nord-Est Australie	séquences COI	Hurwood & Hughes, 2001
	Est USA	RFLP ADH/Amy	Templeton, 1998
	Vosges	séquences D-loop	Mardulyn, 2001
	Nord Péninsule Ibérique	séquences COII	Gómez-Zurita <i>et al.</i> , 2000a

Tableau 1. Couverture taxonomique, géographique et génétique des études publiées jusqu'à fin 2001 utilisant l'outil de l'analyse cladistique species complex pour l'inférence phylogéographique.

génétique qui affectent un sous-ensemble de l'échantillon total, lequel peut ainsi être analysé avec des méthodes appropriées pour calculer Nm . De même, la connaissance de l'information contenue dans la généalogie du marqueur génétique permet d'examiner les distributions spatiale et temporelle de la variation génétique, et pas seulement sa distribution spatiale actuelle comme dans le cas des méthodes hiérarchiques. La NGCA évalue l'adéquation du protocole d'échantillonnage, identifie des déficiences d'échantillonnage en termes de taille d'échantillon ou de lacunes dans l'échantillonnage spatial qui resteraient non détectées en utilisant les autres méthodologies disponibles, particulièrement dans le cas des méthodes généalogiques. De plus, en relation avec les caractéristiques de l'échantillon, la base généalogique de la méthode permet à des haplotypes uniques ou à des populations faiblement échantillonnées d'être informatives dans l'analyse, alors qu'elles seraient statistiquement inadéquates avec d'autres approches. Finalement, à cause de la structure hiérarchique du système emboîté, où les groupes de niveau supérieur recouvrent des périodes évolutives plus longues que ceux de niveau inférieur, la méthode révèle la juxtaposition temporelle de processus historiques et de processus de structuration de population conduisant à la distribution géographique actuelle de la variabilité génétique.

Toutefois, la NGCA partage quelques inconvénients avec certaines des autres méthodes. De toute évidence, le pouvoir inférentiel de cette approche statistique est limité par la taille de l'échantillon et le protocole d'échantillonnage, mais comme nous l'avons dit elle permet d'identifier ces déficits d'échantillonnage dans certains cas. Une des limites les plus importantes de la NGCA est qu'elle repose complètement sur une généalogie de gène pour fonder toutes les inférences phylogéographiques. Si cette généalogie de gène ne reflète pas strictement l'histoire des populations étudiées, à cause d'introgressions génétiques, d'effets de sélection aléatoire de lignées ou de rétention de polymorphismes ancestraux, les inférences qui en sont issues ne seront pas exactes. La solution à ce problème est évidente, elle consiste à étudier plusieurs marqueurs moléculaires et à analy-

ser la congruence du signal phylogénétique et des inférences populationnelles et historiques entre ces marqueurs. Un troisième facteur affectant le pouvoir inférentiel de la NGCA est le niveau de résolution obtenu avec le marqueur utilisé. Un faible nombre d'haplotypes étroitement apparentés dans l'échantillon conduit à un faible nombre de comparaisons statistiques possibles, lesquelles sont fonction de la taille et du degré de ramification du réseau qui les relie. En plus, l'absence d'haplotypes intérieurs dans le réseau compromet la résolution de la NGCA, qui utilise les distances intérieur-apex [(I-T) Dc et (I-T) Dn] comme élément clé pour l'inférence des processus modelant la structure spatiale des populations.

Tous les obstacles discutés jusqu'ici sont reliés au type et à la qualité des données utilisées pour la NGCA et ils ne sont pas inhérents à la méthode elle-même. Par conséquent, le chercheur peut identifier la source des problèmes et essayer de les résoudre ou bien en accroissant la taille de l'échantillon ou bien en ajoutant des marqueurs génétiques. Toutefois, comme noté par TURNER *et al.* (2000), il y a des problèmes liés aux postulats de base de la NGCA; la violation de ces derniers invalidera complètement les inférences relatives à la structure des populations et à leur histoire. Par exemple, un de ces postulats de base est que les haplotypes dérivés (haplotypes apicaux dans les réseaux) ont une distribution géographique plus restreinte que les haplotypes ancestraux (haplotypes intérieurs) sous l'hypothèse nulle d'absence d'association géographique. Cependant, intuitivement on peut s'attendre à ce que la sélection puisse être à l'origine d'une structure géographique différente qui violerait ce postulat.

APPLICATION : LA PHYLOGÉOGRAPHIE DU COMPLEXE *TIMARCHA* *GOETTINGENSIS*

Le complexe d'espèces *Timarcha goettingensis*

Pour illustrer le potentiel de la NGCA en phylogéographie, je vais présenter une réanalyse d'un travail déjà publié dans lequel cette méthode avait été appli-

quée à des séquences partielles du gène COII de l'ADNmt d'un échantillon de coléoptères Chrysomelidae du complexe d'espèces *Timarcha goettingensis* (GÓMEZ-ZURITA *et al.*, 2000a). Dans cette réanalyse, le nombre de spécimens et de localités étudiées a été augmenté afin d'évaluer l'impact des caractéristiques de l'échantillon sur les inférences de la NGCA. De surcroît, le même jeu de données de COII est analysé en utilisant les statistiques *F* et le maximum de parcimonie pour montrer de manière comparative le type d'informations fourni par ces méthodologies dans le cas particulier de cette étude.

Pour plusieurs raisons, le genre *Timarcha*, et plus spécifiquement le complexe d'espèces *T. goettingensis*, se prête particulièrement bien à une étude phylogéographique au moyen de la NGCA. Toutes les espèces du genre sont aptères, essentiellement allopatriques, et elles ont typiquement des aires de distribution restreintes. Elles tendent également à avoir des exigences tout à fait spécifiques pour leur habitat, car il s'agit de coléoptères phytophages spécialistes, se nourrissant sur une seule plante hôte ou sur quelques proches parentes (oligophagie), et particulièrement dans le cas des taxa du complexe *T. goettingensis*, elles sont distribuées dans des aires où l'orographie et les adaptations à l'altitude peuvent être des freins à la dispersion. En résumé, c'est un groupe de coléoptères qui peut être considéré *a priori* comme ayant évolué d'une manière fortement conditionnée par la géographie et où la colonisation de nouvelles aires est un événement rare.

Le complexe *T. goettingensis* est un assemblage monophylétique de taxa étroitement apparentés (GÓMEZ-ZURITA *et al.*, 2000b), distribué en Europe de l'Ouest et Centrale avec une diversité taxonomique maximale dans la Péninsule Ibérique (BECHYNÉ, 1948). Tous les taxons du complexe sont très similaires morphologiquement mais très plastiques dans les caractères typiquement utilisés dans sa taxonomie, qui est encore largement non résolue. Une approche taxonomique basée sur des données cytogénétiques a révélé l'homogénéité cytologique des formes du complexe (PETITPIERRE, 1970, 1976), et les tentatives de taxonomie moléculaire n'ont pas

non plus permis d'identifier les détails de la structure de la lignée, bien que quelques schémas géographiques émergent (GÓMEZ-ZURITA *et al.*, 2000b). Ces études se sont principalement centrées sur les représentants ibériques du complexe, où les problèmes taxonomiques sont particulièrement importants. La compréhension de la connexion entre ces schémas géographiques et la structure génétique du complexe, c'est à dire la phylogéographie, n'est pas simple pour ce groupe étant données les difficultés à relier l'origine des clades avec des événements paléogéographiques particuliers sur une aire géographique très compliquée où les barrières à la dispersion ne sont pas évidentes. Une méthodologie à fort pouvoir de résolution comme la NGCA a clairement fait la preuve de son pouvoir explicatif pour la phylogéographie du complexe (GÓMEZ-ZURITA *et al.*, 2000a).

Échantillonnage, marqueur moléculaire, et analyses populationnelles et phylogénétiques

Cette étude porte sur 157 spécimens échantillonnés dans 36 localités de l'aire de distribution du complexe d'espèces (fig. 3, tableau 2). Chacun de ces individus a été séquencé pour un fragment de 354 pb de l'extrémité 5' du gène mitochondrial COII en utilisant les protocoles de PCR et les méthodes de séquençage détaillées dans GÓMEZ-ZURITA *et al.*, (2000a). Les séquences ont été alignées manuellement, révélant 67 positions variables qui définissent un total de 49 haplotypes.

En général, chaque localité a son propre haplotype distinctif ou deux ou plusieurs haplotypes se distinguant par une ou quelques mutations, bien qu'il y ait quelques exceptions correspondant à des localités avec une diversité d'haplotypes plus forte et quelques cas d'haplotypes partagés par plusieurs populations.

Le calcul des indices de fixations et des nombres estimés de migrants par génération a été réalisé, à partir des données de séquence, au moyen du logiciel DnaSP 3.51 (ROZAS & ROZAS, 1999) et les analyses de maximum de parcimonie ont été effectuées avec le logiciel PAUP* 4.0b8 (SWOFFORD,

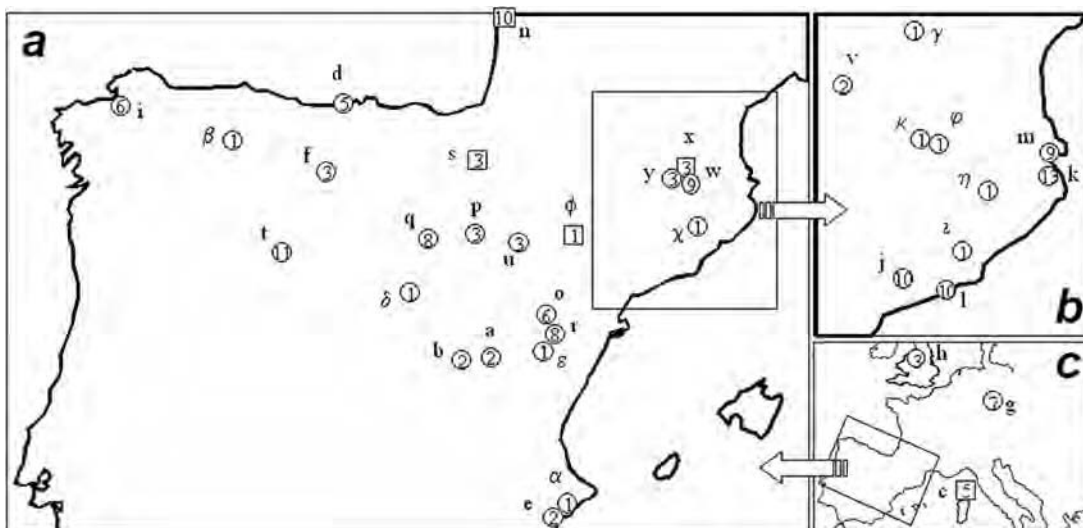


Figure 3. Plan d'échantillonnage pour l'étude phylogéographique du complexe d'espèces *Timarcha goettingensis*. Pour chaque localité, le nombre de spécimens étudié est donné (les localités représentées par un carré sont celles qui ne sont pas incluses dans l'article original de GÓMEZ-ZURITA *et al.*, 2000a). Les codes lettre sont les mêmes que dans les tableaux 2 et 3; les lettres grecques sont utilisées pour les localités où un seul individu a été trouvé et n'ont pas été utilisées pour obtenir les données présentées dans le tableau 3.

2000). Afin de réaliser la NGCA, des réseaux d'haplotypes COII furent construits dans les limites de la parcimonie avec le programme TCS 1.3 (CLEMENT *et al.*, 2000). La structure hiérarchique des tests d'association géographique des haplotypes fut définie sur l'arbre COII en appliquant les règles d'emboîtement décrites par TEMPLETON *et al.* (1987) et les règles additionnelles pour traiter les ambiguïtés proposées par TEMPLETON & SING (1993). Les tests de contingence par permutations et les analyses de distance géographique basées sur le schéma emboîté pour tester les hypothèses alternatives de distribution spatiale non-aléatoire des haplotypes ont été faits avec le programme GeoDis 2.0 (POSADA *et al.*, 2000). L'interprétation phylogéographique des déviations significatives observées au niveau des valeurs de distances ou différences de distances calculées, c'est-à-dire D_c , D_n , $(I-T) D_c$ et $(I-T) D_n$, par rapport à ce qui serait attendu sous l'hypothèse d'une distribution géographique aléatoire des haplotypes a été faite en utilisant une version révisée de

la clé d'interprétation proposée par TEMPLETON *et al.* (1995) (http://bioag.byu.edu/zoology/crandall_lab/dposada/documents/).

Analyses hiérarchique et généalogique des données du gène COII dans le cas du complexe *T. goettingensis*

Le tableau 3 montre les valeurs estimées de F_{ST} et Nm basées sur les comparaisons deux à deux des séquences COII entre les différentes localités étudiées dans tous les cas où plus d'un individu par localité était disponible. Ces mesures de flux génique ne sont pas fiables dans la plupart des cas en raison de la taille limitée de l'échantillon. Toutefois, ces estimations sont simplement une indication puisqu'il n'y a pas de connaissance *a priori* des limites de population, et que mélanger les échantillons en fonction de leur proximité géographique ou de leur ressemblance génétique ne semble pas être une solution adéquate. D'une façon générale, l'analyse révèle une forte structure géographique de l'échantillon, avec 83 % de

Nom de la localité	Coordonnées géographiques*	Taxon	Nb. individus	Haplotypes
Guadalaviar (a)	40.40, -1.72	<i>T. aurichalcea</i>	2	1, 2
Tragacete (b)	40.35, -1.85		2	3, 4
Zicavo (c)	41.90, 9.12	<i>T. cornuta</i>	5	5, 6 (3), 7
Cueto (d)	43.48, -3.80	<i>T. cyanescens</i>	5	8 (4), 9
Port de la Carrasqueta (e)	38.61, -0.48	<i>T. fallax</i>	2	11, 12
Peña Amaya (f)	42.67, -4.17	<i>T. geniculata</i>	3	13
Jena (g)	50.85, 11.52	<i>T. goettingensis</i>	7	15 (6), 16
Stinchcombe (h)	51.68, 2.38		3	17
Fiobre-Bergondo (i)	43.33, -8.23	<i>T. gougeleti</i>	6	18 (5), 19
Alió (j)	41.28, 1.30	<i>T. interstitialis</i>	10	20, 21 (9)
L'Escala (k)	42.13, 3.13		13	23 (12), 24
Garraf (l)	41.27, 1.97		10	21 (9), 25 (1)
Sant Pere de Roda (m)	42.32, 3.17		9	23 (1), 25 (5), 26, 27, 28
Dunes Le Canon (n)	44.68, -1.23	<i>T. maritima</i>	10	29
Ejulve (o)	40.78, -0.53	<i>T. perezi</i>	6	13 (5), 31
El Moncayo (p)	41.80, -1.85		3	13
Puerto de Oncala (q)	41.95, -2.32		8	34 (7), 35
Puerto de Cuarto Pelado (r)	40.55, -0.55		8	13
Tafalla (s)	42.52, -1.67		3	37 (2), 38
Villanubla (t)	41.72, -4.83		11	34
Zaragoza (u)	41.62, -0.92		3	39 (2), 40
Pla de l'Artiga (v)	42.75, 0.75	<i>T. recticollis</i>	2	41, 42
Port de la Creueta (w)	42.32, 1.98	<i>T. sinuaticollis</i>	9	43 (8), 44
Vall d'Eina (x)	42.42, 2.13		3	45 (2), 46
Vall de Núria (y)	42.40, 2.13		3	45 (2), 47
Serra d'Aitana (α)	38.65, -0.32	<i>T. fallax</i>	1	10
Puerto de Ventana (β)	43.05, -6.00	<i>T. geniculata</i>	1	14
Collformic (χ)	41.80, 2.33	<i>T. monserratisensis</i>	1	30
Layna (δ)	41.10, -2.30	<i>T. perezi</i>	1	32
Puerto de Linares (ε)	40.33, -0.53		1	33
Sariñena (φ)	41.78, -0.15		1	36
Caramany (γ)	43.53, 1.75	<i>T. interstitialis</i>	1	22
L'Esquirol (η)	42.03, 2.37		1	21
Sant Just Desvern (ι)	41.37, 2.08		1	21
Queixans (φ)	42.40, 2.07	<i>T. sinuaticollis</i>	1	48
Viliella (κ)	42.42, 1.72	1	49	

* Coordonnées géographiques en degrés décimaux.

Tableau 2. Sites d'échantillonnage, taxons, nombre de spécimens et d'haplotypes COII pour l'analyse phylogéographique du complexe *T. goettingensis*. Chaque localité est affectée du même code lettre qu'en figure 3. Les fréquences d'haplotypes sont indiquées entre parenthèses.

toutes les comparaisons par paires montrant une valeur de F_{ST} supérieure à 0,9 et très peu de migrants par génération ($< 0,08$). 17% des comparaisons montrent des valeurs de F_{ST} ne dépassant pas 0,9, ce qui implique un certain échange génétique entre les localités. Sans complètement écarter la possibilité de flux génique récurrent entre certaines de ces locali-

tés, la plupart de ces valeurs de F_{ST} relativement faibles semblent plus vraisemblablement résulter du partage d'ancêtres communs par les populations et de la rétention d'un polymorphisme ancestral plutôt que d'un réel échange génétique. Comme exemple de la non fiabilité de ces estimations, la localité corse (c) donne des valeurs de F_{ST} inférieures à 0,9 avec

Fst\Nm	a	b	c	d	e	f	g	h	i	j	k	l
a Guadalaviar		0.25	0.03	0.12	0.11	0.02	0.03	0.02	0.02	0.02	0.01	0.02
b Tragacete	0.667		0.03	0.13	0.13	0.02	0.03	0.01	0.03	0.02	0.02	0.02
c Zicavo	0.947	0.940		0.12	0.04	0.01	0.05	0.02	0.02	0.03	0.03	0.03
d Cueto	0.803	0.797	0.807		0.18	0.30	0.12	0.09	0.17	0.09	0.09	0.09
e Carrasqueta	0.824	0.800	0.920	0.739		0.05	0.04	0.03	0.05	0.03	0.03	0.03
f Amaya	0.963	0.963	0.971	0.629	0.913		0.02	0.00	0.01	0.00	0.00	0.00
g Jena	0.945	0.945	0.913	0.813	0.925	0.970		0.11	0.01	0.01	0.01	0.01
h Stinchcombe	0.970	0.971	0.955	0.845	0.949	1.000	0.700		0.00	0.00	0.00	0.00
i Fiobre	0.957	0.951	0.962	0.743	0.913	0.980	0.964	0.990		0.00	0.00	0.00
j Alio	0.969	0.966	0.947	0.852	0.944	0.994	0.954	0.992	0.984		0.04	inf
k Escala	0.972	0.969	0.944	0.849	0.948	0.996	0.952	0.993	0.986	0.850		0.04
l Garraf	0.968	0.966	0.946	0.851	0.943	0.994	0.953	0.992	0.983	0.000	0.850	
m Sant Pere	0.938	0.931	0.884	0.810	0.913	0.957	0.894	0.939	0.947	0.465	0.574	0.400
n Le Canon	0.973	0.970	0.951	0.849	0.946	1.000	0.963	1.000	0.989	0.909	0.963	0.909
o Ejulve	0.951	0.951	0.962	0.614	0.900	0.000	0.959	0.989	0.960	0.984	0.986	0.983
p Moncayo	0.963	0.963	0.971	0.629	0.913	0.000	0.970	1.000	0.980	0.994	0.996	0.994
q Oncala	0.949	0.949	0.956	0.441	0.892	0.974	0.966	0.993	0.968	0.985	0.988	0.985
r Pelado	0.963	0.963	0.971	0.629	0.913	0.000	0.970	1.000	0.980	0.994	0.996	0.994
s Tafalla	0.961	0.959	0.953	0.830	0.934	0.980	0.960	0.982	0.970	0.974	0.977	0.974
t Villanubla	0.960	0.960	0.965	0.458	0.905	1.000	0.974	1.000	0.982	0.993	0.995	0.993
u Zaragoza	0.928	0.920	0.938	0.698	0.889	0.939	0.943	0.966	0.839	0.960	0.963	0.959
v Artiga	0.933	0.923	0.921	0.807	0.912	0.968	0.938	0.968	0.957	0.938	0.945	0.936
w Creueta	0.951	0.947	0.953	0.758	0.922	0.980	0.958	0.986	0.966	0.978	0.980	0.977
x Eina	0.935	0.930	0.951	0.754	0.916	0.964	0.945	0.977	0.957	0.974	0.976	0.973
y Nuria	0.935	0.930	0.952	0.754	0.916	0.964	0.945	0.977	0.957	0.974	0.976	0.973

Tableau 3. F_{ST} (sous la diagonale) et Nm (au-dessus de la diagonale) estimés à partir des données de COII dans les localités échantillonnées du complexe d'espèces *Timarcha goettingensis* quand plus d'un spécimen était disponible par localité. On utilise les estimateurs de Hudson *et al.* (1992b). Code lettre comme sur la carte de la figure 3.

deux localités ibériques (d, m), alors que de toute évidence un flux génique récurrent entre l'île et le continent est très improbable.

Une analyse en maximum de parcimonie des 49 haplotypes COII (50 caractères informatifs) utilisant une recherche heuristique de PAUP* avec 10 répétitions par additions de séquences au hasard a donné un total de 2 400 arbres équiparcimonieux de 121 pas (C.I. : 0,612; R.I. : 0,872). La figure 4 montre le consensus strict de ces arbres et le phylogramme d'un arbre choisi au hasard pour montrer les longueurs de branches. Cet arbre montre un certain degré de structure phylogénétique qui peut être relié à la géographie. Deux clades principaux peuvent être distingués avec très peu de recouvrement dans leur distribution géographique. On iden-

tifie un de ces clades comme le «clade oriental», typiquement européen avec des échantillons du nord-est de la Péninsule Ibérique et des Pyrénées vers le nord, incluant des échantillons de Corse. L'autre clade, ou «clade occidental», correspond à un clade typiquement ibérique avec le reste des échantillons étudiés plus un groupe isolé de localités des Pyrénées orientales. Ces deux clades divergents parapatriciens pourraient être le résultat d'un événement passé d'isolement géographique suivi de diversification et de contact secondaire des lignées au Nord-Est de la Péninsule Ibérique. À l'intérieur de chaque clade la résolution et la robustesse des nœuds est considérablement réduite et seuls les groupes majeurs de séquences géographiquement associées peuvent être distingués, ce qui ne permet pas l'identification de schémas phylogéographiques détaillés.

Emboîtement (« nest »)	Emboîtement au niveau inférieur ^a	Dc	Dn	Inférence
1WC-1*	13 (i)		P	Résultat non conclusif
1EC-1 2WC-2	48 (t) 1WC-1 (i) I-T	G G	G G G	Résultat non conclusif Fragmentation passée
2WC-4*	8 (t) 1WC-3 (i) I-T	P P G	G P P	Fragmentation passée
2EC-1*	49 (t) 1EC-1 (i) 1EC-2 (t) I-T	P	P P G P	Expansion contiguë de l'aire de distribution
2EC-2*	1EC-3 (i) 1EC-4 (i)	P	P G	Résultat non conclusif
3WC-4*	18,19 (i) 2WC-3 (t) I-T	P P	G P G	Colonisation à longue distance
3WC-6*	2WC-5 (i) 45,46,47 (t) I-T	P G	G G	Flux génique restreint avec isolation par la distance
3EC-1*	2EC-2 (i) I-T	P P		Expansion contiguë de l'aire de distribution
4WC-1	3WC-1 (t)		P	Plan d'échantillonnage géographique inadéquat pour discriminer entre fragmentation et isolation par la distance
4WC-2*	3WC-3 (i) 3WC-4 (t) I-T	G P	P G P	Expansion contiguë de l'aire de distribution
4WC-3*	3WC-5 (i) 3WC-6 (t) I-T	P P G	G P G	Résolution génétique insuffisante pour discriminer entre expansion d'aire de distribution/colonisation et dispersion/flux génique restreints (flux génique/dispersion restreints mais avec de la dispersion à longue distance)
4EC-2*	3EC-3 (i) 3EC-4 (t) I-T	P P P	G P G	Échantillonnage inadéquat pour distinguer entre expansion contiguë de l'aire de distribution et colonisation à longue distance
5WC-1*	4WC-1 (i) 4WC-3 (t)	P	G	Expansion contiguë de l'aire de distribution
5EC-1*	4EC-1 (i) 4EC-2 (t) I-T	P G P	P G P	Expansion contiguë de l'aire de distribution
Clades de niveau 5* I-T P	5WC-1 (i) 5EC-1 (t)	P	G	Expansion contiguë de l'aire de distribution
Clades de niveau *	5WC-1 (t) 5EC-1 (i) I-T	P G	G G	Fragmentation passée
Cladogramme total*	5WC-1 (i) 5EC-1 (t) 5WC-2 (t) I-T	P P P	G	Expansion contiguë de l'aire de distribution
Cladogramme total*	5WC-1 (t) 5EC-1 (i) 5WC-2 (t) I-T	P P G	G G	Fragmentation passée
Cladogramme total*	5WC-1 (t) 5EC-1 (t) 5WC-2 (i) I-T	P G P P	P	Expansion contiguë de l'aire de distribution

* Distribution spatiale significativement non-aléatoire des haplotypes évaluée par des tests de contingence avec permutations au seuil 5 % (Templeton *et al.*, 1995).

^a (i) et (t) indiquent quand le niveau d'emboîtement inférieur est intérieur ou à l'extrémité du réseau (« tip »), respectivement.

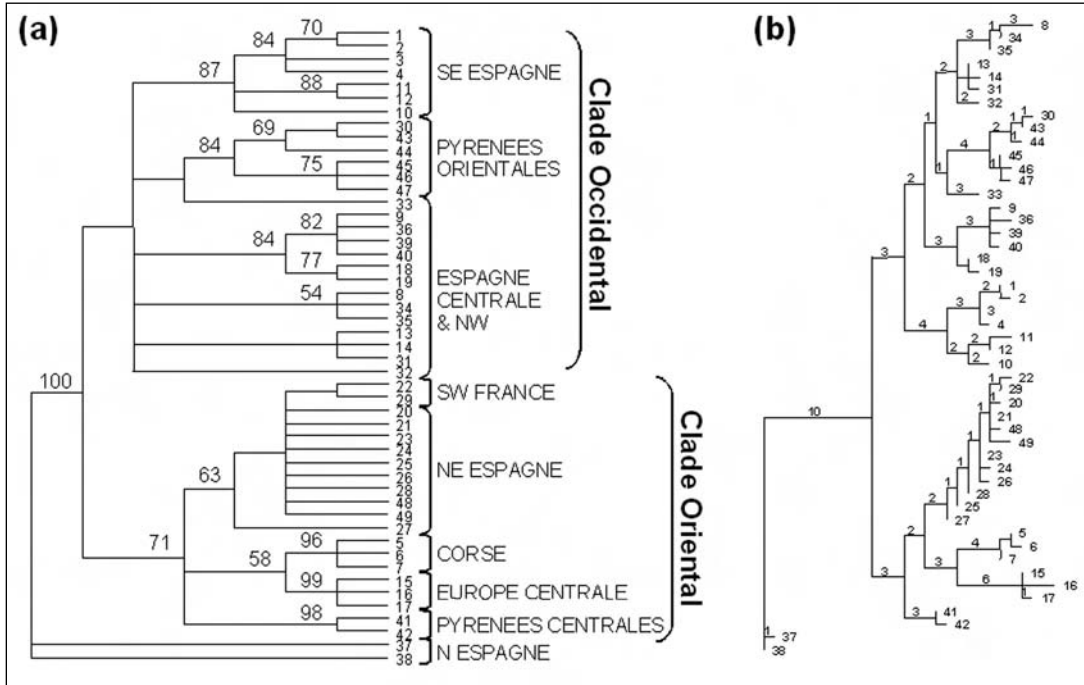


Figure 4. (a) Arbre de consensus strict des 2400 arbres équiparcimonieux (121 pas) obtenus par l'analyse en maximum de parcimonie des 49 haplotypes COII décrits à partir de 157 individus étudiés appartenant au complexe d'espèces *T. goettingensis*. Les nombres au-dessus des branches indiquent les valeurs de bootstrap supérieures à 50 % après 500 pseudo-répliques. Les parenthèses indiquent la distribution géographique globale de chaque groupe d'haplotypes, et deux clades principaux, « occidental » et « oriental » sont également identifiés (détails dans le texte). (b) Phylogramme d'un des 2400 arbres obtenus dans l'analyse de parcimonie, les longueurs de branches sont indiquées au-dessus des branches.

◀ Tableau 4. Résumé des patrons de significativité obtenus pour les statistiques NGCA Dc, Dn, (I T)Dc et (I T)Dn sur la base des données de fréquences d'haplotypes, de la localisation géographique des sites d'échantillonnage (représentés par les coordonnées géographiques) et la structure statistique définie par l'emboîtement sur le réseau COII. Pour les emboîtements des niveaux supérieurs (clades de niveau 5 et cladogramme total), tous les enrachements alternatifs possibles sont considérés. « P » et « G » désignent les valeurs statistiques significativement respectivement plus petites et plus grandes au seuil 5%, comparées à l'hypothèse nulle d'absence de structure géographique ; les résultats non significatifs ne sont pas présentés. Les tests sont faits avec GeoDis 2.0 dans les cas où il y avait au moins deux haplotypes dans plus d'une localité pour un emboîtement particulier. On donne également le scénario le plus vraisemblable expliquant chaque patron significatif particulier déduit en utilisant la clé d'inférence modifiée d'après Templeton *et al.* (1995).

Une analyse géographique cladistique emboîtée du complexe *T. goettingensis*

La combinaison des deux types d'analyses rapides présentées ci-dessus a fourni des informations intéressantes sur la phylogéographie du complexe *T. goettingensis*. Ce complexe d'espèces apparaît fortement structuré géographiquement, avec très peu ou pas de flux génique entre les populations, et les différentes lignées évolutives sont probablement apparues en une succession historique fortement dépendante de leur distribution géographique. Cependant, la taille limitée des échantillons et la faible résolution de l'analyse phylogénétique conduisent à des inférences très limitées quant à la phylogéographie du groupe. C'est particulièrement problématique

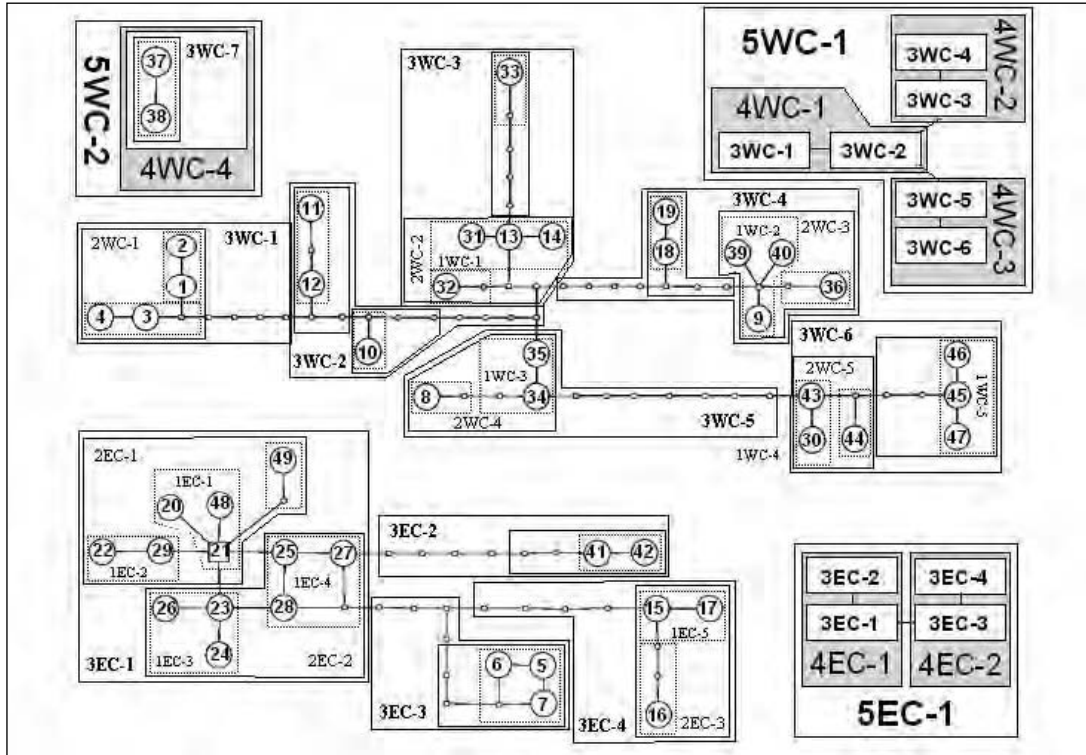


Figure 5. Réseaux d'haplotypes COII du complexe *T. goettingensis* obtenus avec l'algorithme de TEMPLETON *et al.* (1992).

Les nombres d'haplotypes sont ceux donnés au tableau 2. Les haplotypes sont groupés en suivant les règles proposées par TEMPLETON *et al.* (1987) et par TEMPLETON & SING (1993) pour réaliser l'analyse géographique cladistique emboîtée. Les groupes présentant une variation génétique et géographique sont identifiés comme NWC-*n* ou NEC-*n* pour les haplotypes des clades occidental et oriental, respectivement (*N* identifie le niveau d'emboîtement de 1 à 5 et *n* chaque groupe individuel). Les groupes de niveau 4 et 5 sont présentés séparément pour la clarté de la figure.

dans le cas présent, où le cadre fourni par l'aire géographique étudiée ne permet pas d'identifier des schémas de colonisation dans un contexte temporel (comme c'est le cas par exemple dans le cadre d'îles océaniques). Puisque les autres méthodes ne sont pas suffisamment informatives pour l'inférence phylogéographique, la NGCA semble être un outil approprié pour mieux exploiter l'information contenue dans les séquences mitochondriales de ce complexe d'espèces en ce qui concerne son évolution sur le plan géographique. La résolution accrue fournie par les réseaux d'haplotypes, le traitement statistique des données moins sensible à la taille des échantillons, l'information supplémentaire apportée

par les distances géographiques entre sites d'échantillonnage et le cadre permettant de tester des hypothèses alternatives constituent un ensemble de traits prometteurs de la NGCA pour aider à comprendre la phylogéographie du complexe *T. goettingensis*.

Les limites de la parcimonie, c'est-à-dire le nombre maximal de pas autorisés pour connecter deux haplotypes dans l'échantillon avec une probabilité inférieure à 5 % d'inclure de l'homoplasie, est de 8 pas pour les données du COII de *Timarcha*. Si on relie les haplotypes COII de la manière la plus parcimonieuse dans ces limites de parcimonie avec le programme TCS 1.13 on obtient trois réseaux indépendants comme on le voit sur la figure 5.

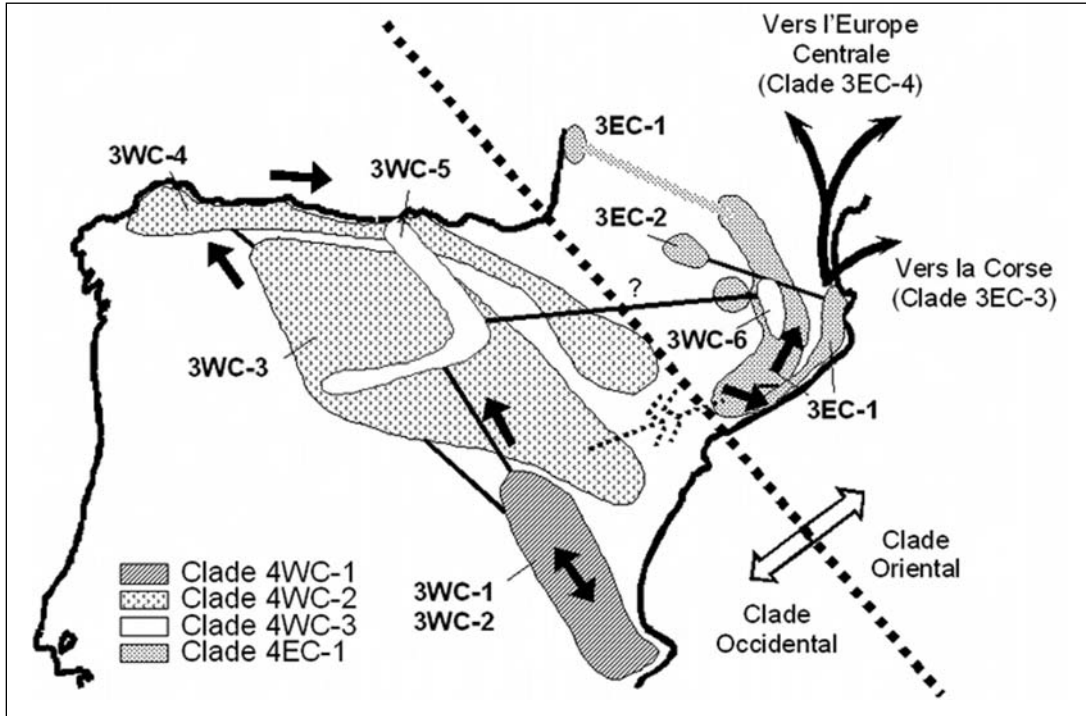


Figure 6. Affichage de la structure généalogique COII du complexe d'espèces *T. goettingensis* sur la distribution géographique des haplotypes et résumé des inférences sur la structure et l'histoire des populations, dérivées de l'analyse géographique cladistique emboîtée. La structure n'est montrée que pour les niveaux d'emboîtement les plus élevés (groupes de niveaux 3 à 5) avec différentes trames et les codes d'identification des groupes. Les flèches doubles représentent un flux génique réduit, les flèches simples indiquent les expansions de distribution spatiale de faible amplitude, et la connexion pointillée brisée entre les groupes d'haplotypes occidental et oriental indique un événement de fragmentation passé. Le point d'interrogation indique une inférence ambiguë concernant la structure observée pour le clade 4WC-3.

Chaque réseau groupe tous les haplotypes inclus dans les deux clades identifiés auparavant comme « occidental » et « oriental », respectivement (plus un réseau avec les deux haplotypes divergents 37 et 38 qui enracaient l'analyse phylogénétique). Les réseaux sont alors utilisés pour définir une série emboîtée de groupes d'haplotypes basée sur la topologie de la généalogie (fig. 5). Les groupes, et les groupes de niveau inférieur qu'ils contiennent, définissent les classes utilisées ensuite dans les tests de contingence par permutation pour trouver les associations géographiques d'haplotypes et les déviations significatives des distances géographiques entre aires de distribution d'haplotypes par rapport à une

situation de distribution spatiale aléatoire des haplotypes (TEMPLETON *et al.*, 1995). Un résumé des résultats et des inférences concernant la structure et l'histoire des populations dérivées des patrons significatifs et de la clé d'interprétation proposée par TEMPLETON sont montrés au tableau 4.

La figure 6 montre une superposition de la structure du réseau COII et des catégories emboîtées de niveau supérieur sur la distribution géographique correspondante des spécimens échantillonnés. Le résultat le plus notable, déjà montré à l'étape de reconstruction du réseau et suggéré par l'analyse phylogénétique, est l'existence très nette de deux groupes parapatricaux d'haplotypes d'ADN mito-

chondriaux. Ces deux groupes, correspondant aux lignées divergentes « orientale » (5EC-1) et « occidentale » (5WC-1) du complexe d'espèces *T. goettingensis*, semblent être apparus consécutivement à des processus historiques, très vraisemblablement un événement de fragmentation passé. Cependant, un scénario alternatif peut également être considéré selon la NGCA, et ceci illustre un problème associé à ce type d'analyse. Les inférences de processus historiques et de structure de population reposent sur la connaissance de la position de la racine qui permet d'assigner un statut intérieur ou apical aux groupes. Au niveau le plus élevé du réseau, lorsqu'on compare les deux clades frères constitués par ces deux lignées, nous disposons de très peu d'arguments suggérant lequel pourrait être le plus ancien. Dans ce cas particulier, si on examine les deux alternatives, placer la racine dans le clade « occidental » conduit à privilégier l'hypothèse d'une expansion continue de la distribution spatiale depuis l'aire ibérique du complexe vers l'Europe, et placer la racine dans le clade « oriental » implique un événement de fragmentation passé. Dans chaque lignée, la NGCA montre un haut niveau de distribution spatiale non aléatoire des haplotypes COII, ce qui peut être expliqué dans la plupart des cas par une extension de la distribution des populations de proche en proche à partir de la partie la plus méridionale de l'aire de distribution de chaque lignée vers les zones septentrionales.

Une interprétation plus détaillée de ces schémas peut être trouvée dans l'étude phylogéographique originale de ce complexe d'espèces (GÓMEZ-ZURITA *et al.*, 2000a). En dépit de l'utilisation d'un échantillonnage légèrement différent, les deux analyses sont très similaires quand aux structures géographiques révélées et aux inférences historiques expliquant ces patrons géographiques particuliers. Toutefois, accroître l'échantillon en termes de nombre de spécimens et de localités introduit de manière évidente des changements dans la topologie des réseaux dont la taille et la structure se développe, introduisant de nécessaires changements dans le schéma d'emboîtement (dépendant de cette topologie) et par conséquent dans les tests réalisés dans la NGCA qui s'ensuit.

Remerciements. — Carlos Juan et Eduard Petitpierre sont coauteurs de l'article original sur la phylogéographie du complexe *T. goettingensis* et beaucoup de leurs idées et suggestions sont également présentes dans ce travail. Je suis très reconnaissant à Pierre Deleporte et Jean François Silvain pour leur révision critique du manuscrit et pour leur travail de traduction en français. Ce travail a été supporté par une Bourse Marie Curie du programme « Improving Human Research Potential and the Socio-economic Knowledge Base » de la Communauté Européenne sous le numéro de contrat HPMF-CT-2000-00744.

RÉFÉRENCES

- AVISE J.-C., 2000. *Phylogeography. The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- AVISE J.-C., ARNOLD J., BALL R.M., Jr., BERMINGHAM E., LAMB T., NEIGEL J.E., REEB C.A. & SAUNDERS N.C., 1987. Intraspecific phylogeography : The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, 18 : 489-522.
- BECHYNÉ J., 1948. Contribution à la connaissance du genre *Timarcha* Latr. 12 : Études phylogénétiques et zoogéographiques (Col. Phytophaga, Chrysomelidae). *Sbor. Narod. Mus. Praze*, 4 : 1-62.
- BERNATCHEZ L., 2001. The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, 55 : 351-379.
- BOS D.H. & SITES J.W., Jr., 2001. Phylogeography and conservation genetics of the Columbia spotted frog (*Rana luteiventris*; Amphibia, Ranidae). *Mol. Ecol.*, 10 : 1499-1513.
- BOSSART J.L. & PROWELL, D.P., 1998. Genetic estimates of population structure and gene flow : limitations, lessons and new directions. *Trends Ecol. Evol.*, 13 : 202-206.
- BROWER A.V.Z., DESALLE R. & VOGLER A.P., 1996. Gene trees, species trees, and Systematics : A cladistic perspective. *Annu. Rev. Ecol. Syst.*, 27 : 423-450.

- CARBONE I. & KOHN L.M., 2001. A microbial population-species interface : nested cladistic and coalescent inference with multilocus data. *Mol. Ecol.*, 10 : 947-964.
- CLEMENT M., POSADA D. & CRANDALL K.A., 2000. TCS : a computer program to estimate gene genealogies. *Mol. Ecol.*, 9 : 1657-1660.
- CRANDALL K.A., 1994. Intraspecific cladogram estimation : accuracy at higher levels of divergence. *Syst. Biol.*, 43 : 222-235.
- CRANDALL K.A., 1996. Multiple interspecies transmissions of human and simian T-cell leukemia I sequences. *Mol. Biol. Evol.*, 13 : 115-131.
- CREER S., MALHOTRA A., THORPE R.S. & CHOU W.H., 2001. Multiple causation of phylogeographical pattern as revealed by nested clade analysis of the bamboo viper (*Trimeresurus stejnegeri*) within Taiwan. *Mol. Ecol.*, 10 : 1967-1981.
- DURAND J.D., TEMPLETON A.R., GUINAND B., IMSIRIDOU A. & BOUVET Y., 1999. Nested clade and phylogeographic analyses of the chub, *Leuciscus cephalus* (Teleostei, Cyprinidae), in Greece : Implications for Balkan Peninsula biogeography. *Mol. Phylogenet. Evol.*, 13 : 566-580.
- DURAND J.D., ÜNLÜ E., DOADRIO I., PIPOYAN S. & TEMPLETON A.R., 2000. Origin, radiation, dispersion and allopatric hybridization in the chub *Leuciscus cephalus*. *Proc. R. Soc. Lond. B*, 267 : 1687-1697.
- FELSENSTEIN J., 1982. How can we infer geography and history from gene frequencies ? *J. Theor. Biol.*, 96 : 9-20.
- FELSENSTEIN J., 1992. Estimating effective population size from samples of sequences : Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.*, 59 : 139-147.
- GÓMEZ A., CARVALHO G.R. & LUNT D.H., 2000. Phylogeography and regional endemism of a passively dispersing zooplankton : mitochondrial DNA variation in rotifer resting egg banks. *Proc. R. Soc. Lond. B*, 267 : 2189-2197.
- GÓMEZ-ZURITA J., PETITPIERRE E. & JUAN C., 2000a. Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, Chrysomelidae). *Mol. Ecol.*, 9 : 557-570.
- GÓMEZ-ZURITA J., JUAN C. & PETITPIERRE E., 2000b. The evolutionary history of the genus *Timarcha* (Coleoptera, Chrysomelidae) inferred from mitochondrial COII gene and partial 16S rDNA sequences. *Mol. Phylogenet. Evol.*, 14 : 304-317.
- HAMMER M.F., KARAFET T., RASANAYAGAM A., WOOD E.T., ALTHEIDE T.K., JENKINS T., GRIFFITHS R.C., TEMPLETON A.R. & ZEGURA S.L., 1998. Out of Africa and back again : Nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.*, 15 : 427-441.
- HUDSON R.R., BOOS D.D. & KAPLAN N.L., 1992a. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.*, 9 : 138-151.
- HUDSON R.R., SLATKIN M. & MADDISON W.P., 1992b. Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132 : 583-589.
- HUELSENBECK J.P. & HILLIS D., 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.*, 42 : 247-264.
- HURWOOD D.A. & HUGHES J.M., 2001. Nested clade analysis of the freshwater shrimp, *Caridina zebra* (Decapoda : Atyidae), from north-eastern Australia. *Mol. Ecol.*, 10 : 113-125.
- JAMES T.Y., MONCALVO J.-M., LI S. & VILGALYS R., 2001. Polymorphism at the ribosomal DNA spacers and its relation to breeding structure of the widespread mushroom *Schizophyllum commune*. *Genetics*, 157 : 149-161.
- JOHNSON J.B. & JORDAN S., 2000. Phylogenetic divergence in leatherside chub (*Gila copei*) inferred from mitochondrial cytochrome b sequences. *Mol. Ecol.*, 9 : 1029-1035.
- KARAFET T.M., ZEGURA S.L., POSUKH O., OSIPOVA L., BERGEN A., LONG J., GOLDMAN D., KLITZ W., HARIHARA S., DE KNIFF P., WIEBE V., GRIFFITHS R.C., TEMPLETON A.R. & HAMMER M.F., 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.*, 64 : 817-831.

- KIMURA M., 1953. «Stepping-stone» model of population. *Annu. Rep. Natl. Inst. Genet. Japan*, 3 : 62-63.
- LYNCH M. & CREASE T.J., 1990. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.*, 7 : 377-394.
- MARDULYN P., 2001. Phylogeography of the Vosges mountains populations of *Gonioctena pallida* (Coleoptera : Chrysomelidae) : a nested clade analysis of mitochondrial DNA haplotypes. *Mol. Ecol.*, 10 : 1751-1763.
- MASKAS S.D. & CRUZAN M.B., 2000. Patterns of intraspecific diversification in the *Piriqueta caroliniana* complex in Southeastern North America and the Bahamas. *Evolution*, 54 : 815-827.
- NEE S., HOLMES E.C., RAMBAUT A. & HARVEY P.H., 1996. Inferring population history from molecular phylogenies. In : P.H. Harvey, A.J. Leigh Brown, J. Maynard Smith & S. Nee (eds.), *New Uses for New Phylogenies*, p. 66-80, Oxford University Press, New York.
- NEI M., 1982. Evolution of human races at the gene level. In : B. Bonne-Tamir, T. Cohen & R.M. Goodman (eds.), *Human Genetics, part A : The Unfolding Genome*, p. 167-181, Alan R. Liss, New York.
- NESBØ C.L., FOSSHEIM T., VØLLESTAD L.A. & JAKOBSEN K.S., 1999. Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Mol. Ecol.*, 8 : 1387-1404.
- NESBØ C.L., RUENESS E.K., IVERSEN S.A., SKAGEN D.W. & JAKOBSEN K.S., 2000. Phylogeography and population history of Atlantic mackerel (*Scomber scombrus* L.) : a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proc. R. Soc. Lond. B*, 267 : 281-292.
- NIELSON M., LOHMAN K. & SULLIVAN J., 2001. Phylogeography of the tailed frog (*Ascaphus truei*) : Implications for the biogeography of the Pacific Northwest. *Evolution*, 55 : 147-160.
- PERKINS S.L., 2001. Phylogeography of Caribbean lizard malaria : tracing the history of vector-borne parasites. *J. Evol. Biol.*, 14 : 34-45.
- PETITPIERRE E., 1970. Cytotaxonomy and evolution of *Timarcha* Latr. (Coleoptera : Chrysomelidae). *Genét. Ibér.*, 22 : 67-120.
- PETITPIERRE E., 1976. Further cytotaxonomical and evolutionary studies on the genus *Timarcha* Latr. (Coleoptera : Chrysomelidae). *Genét. Ibér.*, 28 : 57-81.
- POSADA D., CRANDALL K.A. & TEMPLETON A.R., 2000. GeoDis : A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.*, 9 : 487-488.
- REFSETH U.H., NESBØ C.L., STACY J.E., VØLLESTAD L.A., FJELD E. & JAKOBSEN K.S., 1998. Genetic evidence for different migration routes of freshwater fish into Norway revealed by analysis of current perch (*Perca fluviatilis*) populations in Scandinavia. *Mol. Ecol.*, 7 : 1015-1027.
- ROFF D.A. & BENTZEN P., 1989. The statistical analysis of mitochondrial DNA polymorphisms : χ^2 and the problem of small samples. *Mol. Biol. Evol.*, 6 : 539-545.
- ROZAS J. & ROZAS R., 1999. DnaSP, Version 3 : an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15 : 174-175.
- SEDDON J.M., SANTUCCI F., REEVE N.J. & HEWITT G.M., 2001. DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor* : Pleistocene refugia, postglacial expansion and colonization routes. *Mol. Ecol.*, 10 : 2187-2198.
- SIVASUNDAR A., BERMINGHAM E. & ORTÍ G., 2001. Population structure and biogeography of migratory freshwater fishes (*Prochilodus* : Characiformes) in major South American rivers. *Mol. Ecol.*, 10 : 407-417.
- SLATKIN M., 1985. Gene flow in natural populations. *Annu. Rev. Ecol. Syst.*, 16 : 393-430.
- SLATKIN M., 1987. Gene flow and the geographic structure of natural populations. *Science*, 236 : 787-792.

- SLATKIN M. & MADDISON W.P., 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, 123 : 603-613.
- SWOFFORD D.L., 2000. *PAUP**. *Phylogenetic Analysis Using Parsimony (*and other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts.
- TEMPLETON A.R., 1998. Nested clade analyses of phylogeographic data : testing hypotheses about gene flow and population history. *Mol. Ecol.*, 7 : 381-397.
- TEMPLETON A.R. & SING C.F., 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, 134 : 659-669.
- TEMPLETON A.R. & GEORGIADIS N.J., 1996. A landscape approach to conservation genetics : conserving evolutionary processes in the African Bovidae. In : J.C.Avisé & J.L.Hamrick (eds.), *Conservation Genetics : Case Histories from Nature*, p. 398-430, Chapman & Hall, New York.
- TEMPLETON A.R., BOERWINKLE E. & SING C.F., 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, 117 : 343-351.
- TEMPLETON A.R., CRANDALL K.A. & SING C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132 : 619-633.
- TEMPLETON A.R., ROUTMAN E. & PHILLIPS C.A., 1995. Separating population structure from population history : A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, 140 : 767-782.
- TURGEON J. & BERNATCHEZ L., 2001. Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedii*) : evidence supporting extensive secondary contacts between two glacial races. *Mol. Ecol.*, 10 : 987-1001.
- TURNER T.F., TREXLER J.-C., HARRIS J.L. & HAYNES J.L., 2000. Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. *Genetics*, 154 : 777-785.
- WILSON P., 1997. *Mitochondrial DNA variation and biogeography among some etheostomid darters of the Central Highlands*. PhD thesis, Division of Biomedical and Biological Sciences. Washington University, St. Louis.
- WRIGHT S., 1931. Evolution in Mendelian populations. *Genetics*, 16 : 97-159.